VECTORES DE NEOANTÍGENO DE ALFAVIRUS

LISTADO DE SECUENCIAS

[0001] La presente solicitud contiene un Listado de Secuencias que se presentó a través del sistema de presentación electrónica EFS-Web y se incorpora en su totalidad a la presente mediante referencia. Dicha copia de ASCII, creada en junio 19, 2018, se denomina 40317WO_CRF_sequencelisting.txt, y tiene un tamaño de 470,747 bytes.

ANTECEDENTES

[0002] Las vacunas terapéuticas basadas en neoantígenos específicos del tumor son muy prometedoras como una nueva generación de inmunoterapia contra el cáncer personalizada. [1–3] Los cánceres con una carga mutacional elevada, tales como el cáncer pulmonar no microcítico (NSCLC) y melanoma, son dianas particularmente atractivas de dicha terapia dada la probabilidad relativamente mayor de generación de neoantígenos. [4,5] Las pruebas preliminares muestran que la vacunación basada en antígenos puede provocar respuestas de linfocitos T[6] y que la terapia celular dirigida a neoantígenos puede producir regresión tumoral en determinadas circunstancias en pacientes seleccionados.[7]

[0003] Una pregunta para el diseño de vacunas de neoantígenos es cuál de las tantas mutaciones codificantes presentes en tumores de sujetos pueden generar los «mejores» neoantígenos terapéuticos, p. ej., antígenos que puedan provocar inmunidad antitumoral y producir regresión del tumor.

[0004] Se han propuesto métodos iniciales que incorporan análisis basado en mutaciones mediante el uso de secuenciación de nueva generación, expresión génica de ARN y predicción de afinidad de unión a MHC de péptidos de neoantígenos candidatos[8]. Sin embargo, estos métodos propuestos pueden fallar en modelar la totalidad del proceso de generación de epítopos, que contiene muchas etapas (p. ej., transporte de TAP, escisión proteasomal y/o reconocimiento de TCR) además de la expresión génica y unión a MHC[9]. Por consiguiente, es probable que los métodos existentes tengan un reducido valor predictivo positivo (PPV) bajo. (Figura 1A)

[0005] De hecho, los análisis de péptidos presentados por células tumorales realizados por múltiples grupos han mostrado que <5 % de los péptidos que se predice que se presentarán utilizando la expresión génica y afinidad de unión a MHC se pueden encontrar en el MHC de superficie del tumor[10,11] (Figura 1B). Esta baja correlación entre predicción de unión y presentación de MHC se reforzó aún más mediante las observaciones recientes sobre la falta de mejora en la exactitud de predicción de neoantígenos con unión restringida para la respuesta inhibidora del punto de control sobre la cantidad de mutaciones solamente.[12]

[0006] Este bajo valor predictivo positivo (PPV) de los métodos existentes para predecir la presentación representa un problema para el diseño de vacunas basado en neoantígenos. Si las vacunas se diseñan mediante el uso de predicciones con un bajo PPV, la mayoría de los pacientes

tienen pocas probabilidades de recibir un neoantígeno terapéutico y aún menos tienen la probabilidad de recibir más de uno (incluso asumiendo que todos los péptidos presentados son inmunogénicos). Por lo tanto, es poco probable que la vacunación con neoantígenos con los métodos actuales tenga éxito en una cantidad sustancial de sujetos que tienen tumores. (Figura 1C)

[0007] Adicionalmente, los enfoques anteriores generaron neoantígenos candidatos usando solamente mutaciones de actuación en cis, y en gran medida descuidaron la consideración de fuentes adicionales de neo-ORF, incluidas mutaciones en factores de corte y empalme, que se producen en múltiples tipos de tumores y conducen al corte y empalme anómalo de muchos genes[13], y mutaciones que crean o eliminan los sitios de escisión de la proteasa.

[0008] Finalmente, los enfoques estándar al análisis de transcriptoma y genoma tumoral pueden perderse mutaciones somáticas que dan lugar a neoantígenos candidatos debido a las condiciones subóptimas en la construcción de bibliotecas, captura de transcriptoma y exoma, secuenciación o análisis de datos. Asimismo, los enfoques de análisis de tumor estándar pueden promover inadvertidamente errores de secuencia o polimorfismos de línea germinal como neoantígenos, lo cual conduce al uso ineficiente de la capacidad de la vacuna o riesgo de autoinmunidad, respectivamente.

[0009] Además de los desafíos de los métodos de predicción de neoantígenos actuales existen también determinados desafíos con los sistemas de vectores disponibles que se pueden usar para la administración de neoantígenos en humanos, muchos de los cuales son derivados de humanos. Por ejemplo, muchos humanos tienen inmunidad preexistente a virus humanos como resultado de la exposición natural anterior, y esta inmunidad pude ser un importante obstáculo para el uso de virus recombinantes humanos para la administración de neoantígenos para tratamiento contra el cáncer.

COMPENDIO

[0010] Se describe en la presente una composición para la administración de un sistema de expresión de neoantígenos, que comprende: el sistema de expresión de neoantígenos, donde el sistema de expresión de neoantígenos comprende uno o más vectores, donde el uno o más vectores comprende: (a) una estructura principal de alfavirus de ARN, donde la estructura principal de alfavirus de ARN comprende: (i) al menos una secuencia promotora de nucleótidos y (ii) al menos una secuencia de poliadenilación (poli(A)); y (b) un casete de neoantígeno, donde el casete de neoantígeno comprende: (i) al menos una secuencia de ácido nucleico que codifica neoantígenos derivada de un tumor presente dentro de un sujeto, que comprende: (I) al menos una secuencia de ácido nucleico que codifica neoantígenos del MHC clase I específica derivada del tumor, y que comprende: (A) una secuencia de ácidos nucleico que codifica el epítopo del MHC clase I con al menos una alteración que distingue a la secuencia de péptidos codificada de la secuencia de péptidos correspondiente codificada por una secuencia de ácidos nucleico de tipo salvaje, y (B) opcionalmente, una secuencia enlazadora 5’ y (C)

opcionalmente, una secuencia enlazadora 3'; (ii) opcionalmente, una segunda secuencia promotora de nucleótidos unida de forma operativa con la secuencia de ácido nucleico que codifica neoantígenos; y (iii) opcionalmente, al menos una secuencia de ácido nucleico que codifica el antígeno del MHC clase II; (iv) opcionalmente, al menos una secuencia de ácido nucleico que codifica una secuencia enlazadora de ácido nucleico GPGPG (SEQ ID NO:56); y (v) opcionalmente, al menos una segunda secuencia de poli(A), donde la segunda secuencia de poli(A) es una secuencia de poli(A) natural o una secuencia de poli(A) exógena con respecto al alfavirus.

[0011] También se describe en la presente una composición para la administración de un sistema de expresión de neoantígenos, que comprende: el sistema de expresión de neoantígenos, donde el sistema de expresión de neoantígenos comprende uno o más vectores, donde el uno o más vectores comprende: (a) una estructura principal de alfavirus de ARN, donde la estructura principal de alfavirus de ARN comprende la secuencia de ácido nucleico establecida en la SEQ ID NO:6, donde la secuencia de estructura principal de alfavirus de ARN comprende una secuencia promotora de nucleótidos 26S y una secuencia de poli(A), donde la secuencia promotora 26S es endógena con respecto a la estructura principal de alfavirus de ARN, y donde la secuencia de poli(A) es endógena con respecto a la estructura principal de alfavirus de ARN; y (b) un casete de neoantígeno integrado entre la secuencia promotora de nucleótidos 26S y la secuencia de poli(A), donde el casete de neoantígeno comprende: (i) al menos una secuencia de ácido nucleico que codifica neoantígenos derivada de un tumor presente dentro de un sujeto, que comprende: (I) al menos 10 secuencias de ácido nucleico que codifican el neoantígeno del MHC clase I específicas del tumor y específicas del sujeto, unidas de manera lineal entre sí, cada una de las cuales comprende: (A) una secuencia de ácidos nucleico que codifica el epítopo del MHC clase I con al menos una alteración que distingue a la secuencia de péptidos codificada de la secuencia de péptidos correspondiente codificada por una secuencia de ácidos nucleico de tipo salvaje, donde la secuencia de ácido nucleico que codifica el epítopo del MHC clase I codifica un epítopo del MHC clase I con una longitud de 7-15 aminoácidos, (B) una secuencia enlazadora 5', donde la secuencia enlazadora 5' codifica una secuencia de aminoácidos del extremo N natural del epítopo del MHC I, y donde la secuencia enlazadora 5' codifica un péptido con una longitud de al menos 3 aminoácidos, (C) una secuencia enlazadora 3', donde la secuencia enlazadora 3' codifica una secuencia de ácido del extremo N natural del epítopo del MHC I, y donde la secuencia enlazadora 3' codifica un péptido con una longitud de al menos 3 aminoácidos, y donde el casete de neoantígeno se une de forma operativa con la secuencia promotora de nucleótidos 26S, donde cada una de las secuencias de ácido nucleico que codifican el neoantígeno del MHC clase I codifica un polipéptido con una longitud de entre 13 y 25 aminoácidos, y donde cada extremo 3' de cada secuencia de ácido nucleico que codifica el neoantígeno del MHC clase I se encuentra unido al extremo 5' de la secuencia de ácido nucleico que codifica el

neoantígeno del MHC clase I con la excepción de la secuencia de ácido nucleico que codifica el neoantígeno del MHC clase I final en el casete de neoantígeno; y (ii) al menos dos secuencias de ácido nucleico que codifican el antígeno del MHC clase II que comprenden: (I) una secuencia del MHC clase II de PADRE (SEQ ID NO:48), (II) una secuencia del MHC clase II de toxoide tetánico (SEQ ID NO:46), (III) una primera secuencia de ácido nucleico que codifica una secuencia enlazadora de aminoácidos GPGPG (SEQ ID NO:56) que une la secuencia del MHC clase II de PADRE y la secuencia del MHC clase II de toxoide tetánico, (IV) una segunda secuencia de ácido nucleico que codifica una secuencia enlazadora de aminoácidos GPGPG (SEQ ID NO:56) que une el extremo 5' de las al menos dos secuencias de ácido nucleico que codifican el antígeno del MHC clase II con las al menos 20 secuencias de ácido nucleico que codifican el neoantígeno del MHC clase I específicas del tumor y específicas del sujeto, (V) opcionalmente, una tercera secuencia de ácido nucleico que codifica una secuencia enlazadora de aminoácidos GPGPG (SEQ ID NO:56) en el extremo 3' de las al menos dos secuencias de ácido nucleico que codifican el antígeno del MHC clase II.

[0012] En algunos aspectos, se describe en la fórmula una secuencia ordenada de cada elemento del casete de neoantígeno, de 5' a 3', que comprende:

Pa-(L5b-Nc-L3d)X-(G5e-Uf)Y-G3g

[0013] donde P comprende la segunda secuencia promotora de nucleótidos, donde a = 0 o 1, N comprende una de las secuencias de ácido nucleico que codifican el epítopo del MHC clase I, donde c = 1, L5 comprende la secuencia enlazadora 5', donde b = 0 o 1, L3 comprende la secuencia enlazadora 3', donde d = 0 o 1, G5 comprende una de la al menos una secuencia de ácido nucleico que codifica una secuencia enlazadora de aminoácidos (SEQ ID NO: 56) GPGPG, donde e = 0 o 1, G3 comprende una de la al menos una secuencia de ácido nucleico que codifica una secuencia enlazadora de aminoácidos (SEQ ID NO: 56) GPGPG, donde g = 0 o 1, U comprende una de la al menos una secuencia de ácido nucleico que codifica el antígeno del MHC clase II, donde f = 1, X = 1 a 400, donde para cada X el Nc correspondiente es una secuencia de ácido nucleico que codifica el epítopo e Y = 0, 1 o 2, donde para cada Y el Uf correspondiente es una secuencia de ácido nucleico que codifica el antígeno. En algunos aspectos, para cada X el Nc correspondiente es una secuencia de ácido nucleico que codifican el epítopo del MHC clase I distinta. En algunos aspectos, para cada Y el Uf correspondiente es una secuencia de ácido nucleico que codifica el antígeno del MHC clase II distinta.

[0014] En algunos aspectos, a = 0, b = 1, d = 1, e = 1, g = 1, h = 1, X = 20, Y = 2, la al menos una secuencia promotora de nucleótidos es una secuencia promotora de nucleótidos 26S individual proporcionada por la estructura principal de alfavirus de ARN, la al menos una secuencia de poliadenilación poli(A) es una secuencia de poli(A) de al menos 100 nucleótidos A consecutivos

proporcionada por la estructura principal e alfavirus de ARN, cada N codifica un epítopo del MHC clase I con una longitud de 7-15 aminoácidos. L5 es una secuencia enlazadora 5' natural que codifica una secuencia de aminoácidos del extremo N natural del epítopo del MHC I, y donde la secuencia enlazadora 5' codifica un péptido que tiene una longitud de al menos 3 aminoácidos, L3 es una secuencia enlazadora 3' natural que codifica una secuencia de ácido nucleico terminal natural del epítopo del MHC I, y donde la secuencia enlazadora 3' codifica un péptido que tiene una longitud de al menos 3 aminoácidos, U es cada una de una secuencia de PADRE clase II y una secuencia del MHC clase II de toxoide tetánico, la estructura principal de alfavirus de ARN es la secuencia establecida en la SEQ ID NO:6, y cada una de las secuencias de ácido nucleico que codifican el neoantígeno del MHC clase I codifica un polipéptido con una longitud de entre 13 y 25 aminoácidos.

[0015] En algunos aspectos, cualquiera de las composiciones anteriores comprende, además, un vehículo de administración de nanopartículas. El vehículo de administración de nanopartículas, en algunos aspectos, puede ser una nanopartícula lipídica (LNP). En algunos aspectos, la LNP comprende aminolípidos ionizables. En algunos aspectos, los aminolípidos ionizables comprenden moléculas similares a MC3 (dilinoleilmetil-4-dimetilaminobutirato). En algunos aspectos, el vehículo de administración de nanopartículas encapsula el sistema de expresión de neoantígenos.

[0016] En algunos aspectos, cualquiera de las composiciones anteriores comprende, además, una pluralidad de LNP, donde las LNP comprenden: el sistema de expresión de neoantígenos; un lípido catiónico; un lípido no catiónico; y un lípido conjugado que inhibe la agregación de LNP, donde al menos alrededor de 95 % de las LNP en la pluralidad d LNP: tienen una morfología no lamelar; o son electrodensas.

[0017] En algunos aspectos, el lípido no catiónico es una mezcla de (1) un fosfolípido y (2) colesterol o un derivado de colesterol.

[0018] En algunos aspectos, el lípido conjugado que inhibe la agregación de las LNP es un conjugado de polietilenglicol (PEG)-lípido. En algunos aspectos, el conjugado de PEG-lípido se selecciona del grupo que consiste en un conjugado de PEG-diacilglicerol (PEG-DAG), un conjugado de PEG-dialquiloxipropilo (PEG-DAA), un conjugado de PEG-fosfolípido, un conjugado de PEG-ceramida (PEG-Cer), y una mezcla de estos. En algunos aspectos, el conjugado de PEG-DAA se selecciona del grupo que consiste en un conjugado de PEG-dideciloxipropilo (C_{10}), un conjugado de PEG-dilauriloxipropilo (C_{12}), un conjugado de PEG-dimiristiloxipropilo (C_{14}), un conjugado de PEG-dipalmitiloxipropilo (C_{16}), un conjugado de PEG-diestheariloxipropilo (C_{18}) y una mezcla de estos.

[0019] En algunos aspectos, el sistema de expresión de neoantígenos está completamente encapsulado en las LNP.

[0020] En algunos aspectos, la morfología no lamelar de las LNP comprende una estructura de fase

cúbica o hexagonal inversa (H_{II}).

[0021] En algunos aspectos, el lípido catiónico comprende de alrededor de 10 % molar a alrededor de 50 % molar de los lípidos totales presentes en las LNP. En algunos aspectos, el lípido catiónico comprende de alrededor de 20 % molar a alrededor de 50 % molar de los lípidos totales presentes en las LNP. En algunos aspectos, el lípido catiónico comprende de alrededor de 20 % molar a alrededor de 40 % molar de los lípidos totales presentes en las LNP.

[0022] En algunos aspectos, el lípido no catiónico comprende de alrededor de 10 % molar a alrededor de 60 % molar de los lípidos totales presentes en las LNP. En algunos aspectos, el lípido no catiónico comprende de alrededor de 20 % molar a alrededor de 55 % molar de los lípidos totales presentes en las LNP. En algunos aspectos, el lípido no catiónico comprende de alrededor de 25 % molar a alrededor de 50 % molar de los lípidos totales presentes en las LNP.

[0023] En algunos aspectos, el lípido conjugado comprende de alrededor de 0,5 % molar a alrededor de 20 % molar de los lípidos totales presentes en las LNP. En algunos aspectos, el lípido conjugado comprende de alrededor de 2 % molar a alrededor de 20 % molar de los lípidos totales presentes en las LNP. En algunos aspectos, el lípido conjugado comprende de alrededor de 1,5 % molar a alrededor de 18 % molar de los lípidos totales presentes en las LNP.

[0024] En algunos aspectos, más del 95 % de las LNP tienen una morfología no lamelar. En algunos aspectos, más del 95 % de las LNP son electrodensas.

[0025] En algunos aspectos, cualquiera de las composiciones anteriores comprende, además, una pluralidad de LNP, donde las LNP comprenden: un lípido catiónico que comprende de 50 % molar a 65 % molar de los lípidos totales presentes en las LNP; un lípido conjugado que inhibe la agregación de LNP que comprende de 0,5 % molar a 2 % molar de los lípidos totales presente en las LNP; y un lípido no catiónico que comprende: ya sea una mezcla de un fosfolípido y colesterol o un derivado de este, donde el fosfolípido comprende de alrededor de 4 % molar a 10 % molar de los lípidos totales presentes en las LNP y el colesterol o un derivado de este comprende de alrededor 30 % molar a 40 % molar de los lípidos totales presentes en las LNP; ya sea una mezcla de un fosfolípido y colesterol o un derivado de este, donde el fosfolípido comprende de alrededor de 3 % molar a 15 % molar de los lípidos totales presentes en las LNP y el colesterol o un derivado de este comprende de alrededor 30 % molar a 40 % molar de los lípidos totales presentes en las LNP; o hasta 49,5 % molar de los lípidos totales presentes en las LNP y comprende una mezcla de un fosfolípido y colesterol o un derivado de este, donde el colesterol o derivado de este colesterol de 30 % molar a 40 % molar de los lípidos totales presentes en las LNP.

[0026] En algunos aspectos, cualquiera de las composiciones anteriores comprende, además, una pluralidad de LNP, donde las LNP comprenden: un lípido catiónico que comprende de 50 % molar a

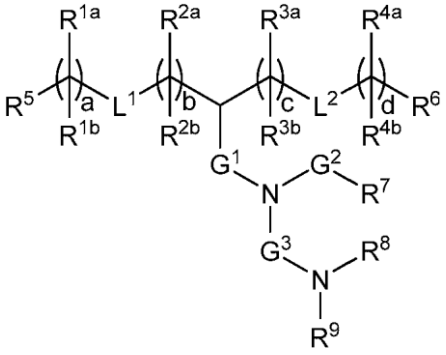
85 % molar de los lípidos totales presentes en las LNP; un lípido conjugado que inhibe la agregación de LNP que comprende de 0,5 % molar a 2 % molar de los lípidos totales presente en las LNP; y un lípido catiónico que comprende de 13 % molar a 49,5 % molar de los lípidos totales presentes en las LNP.

[0027] En algunos aspectos, el fosfolípido comprende dipalmitoilfosfatidilcolina (DPPC), distearoilfosfatidilcolina (DSPC) o una mezcla de estas.

[0028] En algunos aspectos, el lípido conjugado comprende un conjugado de polietilenglicol (PEG)-lípido. En algunos aspectos, el conjugado de PEG-lípido comprende un conjugado de PEG-diacilglicerol (PEG-DAG), un conjugado de PEG-dialquiloxipropilo (PEG-DAA), o una mezcla de estos. En algunos aspectos, el conjugado de PEG-DAA comprende un conjugado de PEG-dimistiriloxipropilo (PEG-DMA), un conjugado de PEG-diestеariloxipropilo (PEG-DSA), o una mezcla de estos. En algunos aspectos, la parte de PEG del conjugado tiene un peso molecular promedio de alrededor de 2000 Dalton.

[0029] En algunos aspectos, el lípido conjugado comprende de alrededor de 1 % molar a 2 % molar de los lípidos totales presentes en las LNP.

[0030] En algunos aspectos, la LNP comprende un compuesto que tiene una estructura de la Fórmula I que tiene una estructura de la Fórmula I:

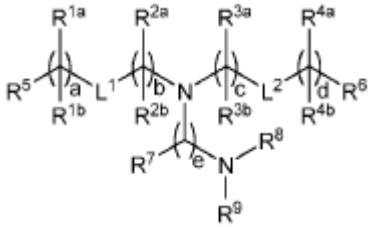


I

o una sal, tautómero, profármaco o estereoisómero farmacéuticamente aceptable de este, donde: cada uno de L^1 y L^2 son independientemente -0(C=0)-, -(C=0)0-, -C(=0)-, -0-, $-S(0)_x$-, -S-S-, -C(=0)S-, -SC(=0)-, - $R^aC(=0)$-, -C(=0) R^a-, - $R^aC(=0)$ R^a-, -OC(=0) R^a-, - $R^aC(=0)$0- o un enlace directo; G^1 es alquileno Ci-C_2, - (C=0)-, -0(C=0)-, -SC(=0)-, - $R^aC(=0)$- o un enlace directo: -C(=0)-, -(C=0)0-, -C(=0)S-, -C(=0) R^a- o un enlace directo; G es alquileno Ci-C_6; R^a es H o alquilo C1-C12; R^{la} y R^{lb} son, en cada aparición, independientemente ya sea: (a) H o alquilo C_1-C_{12}; o (b) R^{la} es H o alquilo C_1-C_{12}, y R^{lb} junto con el átomo de carbono al cual está unido tomados juntos con un R^{lb} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^{2a} y R^{2b} son, en cada aparición, independientemente ya sea: (a) H o alquilo C_1-C_{12}; o (b) R^{2a} es H o alquilo C_1-C_{12}, y R^{2b} junto con el átomo de carbono al cual está unido tomados juntos con un R^{2b} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^{3a} y R^{3b} son, en cada aparición, independientemente ya sea (a): H o

alquilo C_1-C_{12}; o (b) R^{3a} es H o alquilo C_1-C_{12}, y R^{3b} junto con el átomo de carbono al cual está unido tomados juntos con un R adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^{4a} y R^{4b} son, en cada aparición, independientemente ya sea: (a) H o alquilo C1-C12; o (b) R^{4a} es H o alquilo C1-C12, y R^{4b} junto con el átomo de carbono al cual está unido tomados juntos con un R^{4b} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^5 y R^6 son, en cada aparición, independientemente H o metilo; R^7 es alquilo C4-C20; cada uno de R^8 y R^9 es independientemente alquilo C1-C12; o R^8 y R^9, junto con el átomo de nitrógeno al que están unidos, forman un anillo heterocíclico de 5, 6 o 7 miembros; cada uno de a, b, c y d es independientemente un entero seleccionado de 1 a 24; y x es 0, 1 o 2.

[0031] En algunos aspectos, la LNP comprende un compuesto que tiene una estructura de la Fórmula II:



II

o una sal, tautómero, profármaco o estereoisómero farmacéuticamente aceptable de este, donde: cada uno de L^1 y L^2 es independientemente -0(C=0)-, -(C=0)0- o un enlace doble carbono-carbono; R^{la} y R^{lb} son, en cada aparición, independientemente ya sea (a) H o alquilo C_1-C_{12} o (b) R^{la} es H o alquilo C_1-C_{12}, y R^{lb} junto con el átomo de carbono al cual está unido tomados juntos con un R^{lb} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^{2a} y R^{2b} son, en cada aparición, independientemente ya sea (a) H o alquilo C_1-C_{12} o (b) R^{2a} es H o alquilo C_1-C_{12}, y R^{2b} junto con el átomo de carbono al cual está unido tomados juntos con un R^{2b} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^{3a} y R^{3b} son, en cada aparición, independientemente ya sea (a) H o alquilo C_1-C_{12} o (b) R^{3a} es H o alquilo C_1-C_{12}, y R^{3b} junto con el átomo de carbono al cual está unido tomados juntos con un R^{3b} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; R^{4a} y R^{4b} son, en cada aparición, independientemente ya sea (a) H o alquilo C_1-C_{12} o (b) R^{4a} es H o alquilo C_1-C_{12}, y R^{4b} junto con el átomo de carbono al cual está unido tomados juntos con un R^{4b} adyacente y el átomo de carbono al cual está unido forman un enlace doble carbono-carbono; cada uno de R^5 y R^6 es independientemente metilo o cicloalquilo; R^7 es, en cada aparición, independientemente, H o -alquilo C_1-C_{12}; cada uno de R^8 y R^9 es independientemente alquilo C1-C12 no sustituido; o R^8 y R^9, junto con el átomo de nitrógeno al que están unidos, forman un anillo heterocíclico de 5, 6 o 7 miembros que comprende un átomo de nitrógeno; cada uno de a y d es

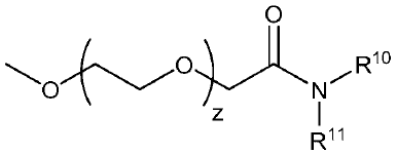
independientemente un entero de 0 a 24; cada uno de b y c es independientemente un entero seleccionado de 1 a 24; y e es 1 o 2, siempre que: al menos uno de R^{la}, R^{2a}, R^{3a} o R^{4a} sea alquilo C1-C12, o al menos uno de L^1 o L^2 sea -0(C=0)- o -(C=0)0-; y R^{la} y R^{lb} no sean isopropilo cuando a es 6 o n-butilo cuando a es 8.

[0032] En algunos aspectos, cualquiera de las composiciones comprende, además, uno o más excipientes que comprenden un lípido neutro, un esteroide y un lípido conjugado con polímero. En algunos aspectos, el lípido neutro comprende al menos uno de l,2-Distearoil-*sn*-glicero-3-fosfocolina (DSPC), l,2-Dipalmitoil-*sn*-glicero-3-fosfocolina (DPPC), l,2-Dimiristoil-*sn*-glicero-3-fosfocolina (DMPC), 1-Palmitoil-2-oleoil-*sn*-glicero-3-fosfocolina (POPC), 1,2-dioleoil-*sn*-glicero-3-fosfocolina (DOPC) y l,2-Dioleoil-*sn*-glicero-3-fosfoetanolamina (DOPE). En algunos aspectos, el lípido neutro es DSPC.

[0033] En algunos aspectos, la relación molar del compuesto con respecto al lípido neutro oscila de alrededor de 2:1 a alrededor de 8:1.

[0034] En algunos aspectos, el esteroide es colesterol. En algunos aspectos, la relación molar del compuesto con respecto al colesterol oscila de alrededor de 2:1 a 1:1.

[0035] En algunos aspectos, el lípido conjugado con polímero es un lípido pegilado. En algunos aspectos, la relación molar del compuesto con respecto al lípido pegilado oscila de alrededor de 100:1 a alrededor de 25:1. En algunos aspectos, el lípido pegilado es PEG-DAG, un PEG polietileno (PEG-PE), un PEG-succinoil-diacilglicerol (PEG-S-DAG), PEG-cer o un PEG dialquioxipropilcarbamato. En algunos aspectos, el lípido pegilado tiene la siguiente estructura III:



III

o una sal, tautómero o estereoisómero farmacéuticamente aceptable de este, donde: cada uno de R^{10} y R^{11} es independientemente una cadena de alquilo recta o ramificada, saturada o no saturada que contiene de 10 a 30 átomos de carbono, donde la cadena de alquilo se interrumpe opcionalmente con uno o más enlaces de éster; y z tiene un valor promedio que oscila de 30 a 60. En algunos aspectos, cada uno de R^{10} y R^{11} es independientemente una cadena de alquilo recta saturada que tiene de 12 a 16 átomos de carbono. En algunos aspectos, el z promedio es de alrededor de 45.

comenzar aquí

[0036] En algunos aspectos, la LNP se autoensambla para producir estructuras que no sean de bicapa cuando se mezcla con ácido nucleico polianiónico. En algunos aspectos, las estructuras que no son bicapa tienen un diámetro de entre 60 nm y 120 nm. En algunos aspectos, las estructuras que no

son bicapa tienen un diámetro de alrededor de 70 nm, alrededor de 80 nm, alrededor de 90 nm o alrededor de 100 nm. En algunos aspectos, donde el vehículo de administración de nanopartículas tiene un diámetro de alrededor de 100 nm.

[0037] En algunos aspectos, el casete de neoantígeno se integra entre la al menos una secuencia promotora de nucleótidos y la al menos una secuencia de poli(A). En algunos aspectos, la al menos una secuencia promotora de nucleótidos se une de forma operativa con la secuencia de ácido nucleico que codifica neoantígenos.

[0038] En algunos aspectos, el uno o más vectores comprende uno o más vectores de ARN de cadena +. En algunos aspectos, el uno o más vectores de ARN de cadena + comprende un casquete 5' 7-metilguanosina (m7g). En algunos aspectos, el uno o más vectores de ARN de cadena + se produce mediante transcripción *in vitro*. En algunos aspectos, el uno o más vectores son autorreplicantes dentro de una célula de mamífero.

[0039] En algunos aspectos, la estructura principal de alfavirus de ARN comprende al menos una secuencia de nucleótidos de un virus Aura, un virus Fort Morgan, un virus de encefalitis equina venezolana, un virus del río Ross, un virus del bosque Semliki, un virus Sindbis o un virus Mayaro. En algunos aspectos, la estructura principal de alfavirus de ARN comprende al menos una secuencia de nucleótidos de un virus de encefalitis equina venezolana. En algunos aspectos, la estructura principal de alfavirus de ARN comprende al menos secuencias para la amplificación mediada por proteínas no estructurales, una secuencia promotora 26S, una secuencia de poli(A), un gen de la proteína no estructural 1 (nsP1), un gen de nsP2, un gen de nsP3 y un gen de nsP4, codificados por la secuencia de nucleótidos del virus Aura, el virus Fort Morgan, el virus de la encefalitis equina venezolana, el virus del río Ross, el virus del bosque Semliki, el virus Sindbis o el virus Mayaro. En algunos aspectos, la estructura principal de alfavirus de ARN comprende al menos secuencias para la amplificación mediada por proteínas no estructurales, una secuencia promotora 26S y una secuencia de poli(A) codificadas por la secuencia de nucleótidos del virus Aura, el virus Fort Morgan, el virus de la encefalitis equina venezolana, el virus del río Ross, el virus del bosque Semliki, el virus Sindbis o el virus Mayaro. En algunos aspectos, las secuencias para la amplificación mediada por proteínas no estructurales se seleccionan del grupo que consiste en: una UTR 5' de alfavirus, un CSE 51-nt, un CSE 24-nt, una secuencia promotora subgenómica 26S, un CSE 19-nt y una UTR 3' de alfavirus, o combinaciones de estos.

[0040] En algunos aspectos, la estructura principal de alfavirus de ARN no codifica cápside de proteínas estructurales de virión, E2 y E1. En algunos aspectos, el casete de antígenos se inserta en lugar de las proteínas estructurales de virión dentro de la secuencia de nucleótidos del virus Aura, el virus Fort Morgan, el virus de la encefalitis equina venezolana, el virus del río Ross, el virus del bosque

Semliki, el virus Sindbis o el virus Mayaro.

[0041] En algunos aspectos, el virus de encefalitis equina venezolana (VEE) comprende la cepa TC-83. En algunos aspectos, el virus de encefalitis equina venezolana comprende la secuencia que se establece en las SEQ ID NO:3 o SEQ ID NO:5. En algunos aspectos, el virus de encefalitis equina venezolana comprende la secuencia de las SEQ ID NO:3 o SEQ ID NO:5 que comprenden además una deleción entre los pares de bases 7544 y 11175. En algunos aspectos, la estructura principal de alfavirus de ARN es la secuencia establecida en la SEQ ID NO:6 o la SEQ ID NO:7. En algunos aspectos, el casete de neoantígeno se inserta para reemplazar la deleción entre los pares de base 7544 y 11175 establecida en la secuencia de la SEQ ID NO:3 o la SEQ ID NO:5.

[0042] En algunos aspectos, la inserción del casete de neoantígeno proporciona la transcripción de un ARN policistrónico que comprende los genes nsP1-4 y la al menos una secuencia de ácido nucleico que codifica el antígeno, donde los genes nsP1-4 y la al menos una secuencia de ácido nucleico que codifica el antígeno están en marcos abierto de lectura separados.

[0043] En algunos aspectos, la al menos una secuencia promotora de nucleótidos es la secuencia promotora de nucleótidos 26S natural codificada por la estructura principal de alfavirus de ARN. En algunos aspectos, la al menos una secuencia promotora de nucleótidos es un promotor de ARN exógeno. En algunos aspectos, la segunda secuencia promotora de nucleótidos es una secuencia promotora de nucleótidos 26S. En algunos aspectos, la segunda secuencia promotora de nucleótidos comprende múltiples secuencias promotoras de nucleótidos 26S, donde cada secuencia promotora de nucleótidos 26S proporciona la transcripción de uno o más marcos abiertos de lectura separados.

[0044] En algunos aspectos, el uno o más vectores de expresión de neoantígenos tienen un tamaño de al menos 300 nt. En algunos aspectos, el uno o más vectores de expresión de neoantígenos tienen un tamaño de al menos 1 kb. En algunos aspectos, el uno o más vectores de expresión de neoantígenos tienen un tamaño de 2 kb. En algunos aspectos, el uno o más vectores de expresión de neoantígenos tienen un tamaño menor que 5 kb.

[0045] En algunos aspectos, al menos una de la al menos una secuencia de ácido nucleico que codifican el neoantígeno codifica una secuencia de polipéptidos o parte de esta que se presenta por el MHC clase I en la célula tumoral. En algunos aspectos, cada una de las secuencias de ácidos nucleicos que codifican el antígeno están directamente enlazadas entre sí. En algunos aspectos, al menos una de la al menos una secuencia de ácidos nucleicos que codifican el antígeno está enlazada a una secuencia de ácidos nucleicos que codifican el antígeno distinta con una secuencia de ácido nucleico que codifica un enlazador. En algunos aspectos, el enlazador enlaza dos secuencias del MHC clase I o una secuencia del MHC clase I con una secuencia del MHC clase II. En algunos aspectos, el enlazador se selecciona del grupo que consiste en: (1) residuos de glicina consecutivos, con una longitud de al menos 2, 3, 4, 5,

6, 7, 8, 9 o 10 residuos; (2) residuos de alanina consecutivos, con una longitud de al menos 2, 3, 4, 5, 6, 7, 8, 9 o 10; (3) dos residuos de arginina (RR); (4) alanina, alanina, tirosina (AAY); (5) una secuencia de consenso de una longitud de al menos 2, 3, 4, 5, 6, 7, 8, 9 o 10 residuos de aminoácidos que se procesa de forma eficiente por un proteasoma de mamífero; y (6) una o más secuencias naturales que flanquean al antígeno derivadas de la proteína cognada de origen y que tienen una longitud de al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o 2-20 residuos de aminoácidos. En algunos aspectos, el enlazador enlaza dos secuencias del MHC clase II o una secuencia del MHC clase II a una secuencia del MHC clase I. En algunos aspectos, el enlazador comprende la secuencia GPGPG (SEQ ID NO: 56).

[0046] En algunos aspectos, al menos una secuencia de la al menos una secuencia de ácido nucleico que codifica antígenos está enlazada, de forma operativa o directamente, a una secuencia separada o contigua que mejora la expresión, estabilidad, tráfico de células, procesamiento y presentación y/o la inmunogenicidad de la al menos una secuencia de ácido nucleico que codifica antígenos. En algunos aspectos, la secuencia separada o contigua comprende al menos uno de: una secuencia de ubiquitina, una secuencia de ubiquitina modificada para mejorar el direccionamiento a proteasoma (p. ej., la secuencia de ubiquitina contiene una sustitución de Gly a Ala en la posición 76), una secuencia de señal de inmunoglobulina (p. ej., IgK), una secuencia de complejo mayor de histocompatibilidad clase I, proteína de membrana asociada a lisosoma (LAMP)-1, proteína de membrana asociada a lisosoma de célula dendrítica humana y una secuencia de complejo mayor de histocompatibilidad clase II; donde opcionalmente la secuencia de ubiquitina modificada para aumentar el direccionamiento a proteasoma es A76.

[0047] En algunos aspectos, al menos una de la la al menos una secuencia de ácido nucleico que codifica el neoantígeno codifica una secuencia de polipéptidos o una parte de esta que tiene afinidad de unión aumentada a su alelo del MHC correspondiente con respecto a la secuencia de ácidos nucleicos de tipo salvaje correspondiente traducida. En algunos aspectos, al menos una de la la al menos una secuencia de ácido nucleico que codifica el neoantígeno de las múltiples secuencias codifica una secuencia de polipéptidos o una parte de esta que tiene estabilidad de unión aumentada a su alelo del MHC correspondiente con respecto a la secuencia de ácidos nucleicos de tipo salvaje correspondiente traducida. En algunos aspectos, al menos una de la al menos una secuencia de ácido nucleico que codifica el neoantígeno de las múltiples secuencias codifica una secuencia de polipéptidos o una parte de esta que tiene una probabilidad aumentada de presentación en su alelo del MHC correspondiente con respecto a la secuencia de ácidos nucleicos parental de tipo salvaje correspondiente traducida.

[0048] En algunos aspectos, al menos una alteración comprende una alteración puntual, una mutación de desplazamiento de marco, una mutación de eliminación, una variante de corte y empalme, una reorganización genómica o un antígeno sometido a corte y empalme generado por el proteasoma.

[0049] En algunos aspectos, el tumor se selecciona del grupo que consiste en: cáncer de pulmón, melanoma, cáncer de mama, cáncer de ovario, cáncer de próstata, cáncer de riñón, cáncer gástrico, cáncer de colon, cáncer testicular, cáncer de cabeza y cuello, cáncer pancreático, cáncer de vejuga, cáncer de cerebro, linfoma de linfocitos B, leucemia mielógena aguda, leucemia linfoblástica aguda del adulto, leucemia mielógena crónica, leucemia linfocítica crónica, leucemia linfocítica de linfocitos T, cáncer pulmonar no microcítico y cáncer pulmonar microcítico.

[0050] En algunos aspectos, la secuencia de ácido nucleico que codifica el neoantígeno comrpende al menos 2-10, 2, 3, 4, 5, 6, 7, 8, 9 o 10 secuencia de ácidos nucleicos. En algunos aspectos, la al menos una secuencia de ácido nucleico que codifica el neoantígeno comprende al menos 11-20, 15-20, 11-100, 11-200, 11-300, 11-100, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o hasta 400 secuencia de ácidos nucleicos.

[0051] En algunos aspectos, la al menos una secuencia de ácido nucleico que codifica el neoantígeno comprende al menos 2-400 secuencias de ácido nucleico y donde al menos dos de las secuencias de ácido nucleico que codifican el neoantígeno codifican secuencias de polipéptidos o partes de estas que se presentan por MHC clase I en la superficie de la célula tumoral. En algunos aspectos, al menos dos de las secuencias de ácido nucleico que codifican el neoantígeno codifican secuencias de polipéptidos o partes de estas que se presentan por MHC clase I en la superficie de la célula tumoral. En algunos aspectos, cuando se administran al sujeto y se traducen, al menos uno de los neoantígenos codificados por la al menos una secuencia de ácido nucleico que coodifica el neoantígeno se presenta en células presentadoras de antígenos, lo que da como resultado una respuesta inmunitaria dirigida a al menos uno de los neoantígenos de la superficie de la célula tumoral. En algunos aspectos, al menos una secuencia de ácido nucleico que codifica el neoantígeno, cuando se administran al sujeto y se traducen, al menos uno de los neoantígenos del MHC clase I o clase II se presenta en células presentadoras de antígenos lo cual da como resultado una respuesta inmunitaria dirigida a al menos uno de los neoantígenos en la superficie de la célula tumoral, y donde opcionalmente la expresión de cada una de la al menos una secuencia de ácido nucleico que codifican el neoantígeno del es impulsada por al menos un promotor.

[0052] En algunos aspectos, cada secuencia de ácidos nucleicos que codifican el neoantígeno del MHC case I codifica una secuencia de polipéptidos con una longitud de entre 8 y 35 aminoácidos, opcionalmente una longitud de 9-17, 9-25, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34 o 35 aminoácidos.

[0053] En algunos aspectos, al menos una secuencia de ácidos nucleicos que codifican el antígeno del MHC clase II está presente. En algunos aspectos, al menos una secuencia de ácidos nucleicos que codifican el antígeno del MHC clase II está presente y comprende al menos una secuencia de ácidos

nucleicos que codifican el neoantígeno del MHC clase II que comprende al menos una mutación la distingue de la secuencia peptídica correspondiente codificada por la secuencia de ácidos nucleicos de tipo salvaje original. En algunos aspectos, la o las secuencias de ácidos nucleicos que codifican el antígeno del MHC clase II son de una longitud de 12-20, 12, 13, 14, 15, 16, 17, 18, 19, 20 o 20-40 aminoácidos. En algunos aspectos, la o las secuencias de ácidos nucleicos que codifican el antígeno del MHC clase II están presentes y comprenden al menos una secuencia de ácidos nucleicos que codifican el antígeno del MHC clase II universal, donde opcionalmente la o las secuencias universales comprenden al menos uno de toxoide tetánico y PADRE.

[0054] En algunos aspectos, la al menos una secuencia promotora de nucleótidos o la segunda secuencia promotora de nucleótidos son inducibles. En algunos aspectos, la al menos una secuencia promotora de nucleótidos o la segunda secuencia promotora de nucleótidos son no inducibles

[0055] En algunos aspectos, la al menos una secuencia de poli(A) comprende una secuencia de poli(A) natural para el alfavirus. En algunos aspectos, la al menos una secuencia de poli(A) comprende una secuencia de poli(A) exógena para el alfavirus. En algunos aspectos, la al menos una secuencia de poli(A) se une de forma operativa a al menos una de la al menos una secuencia de ácido nucleico que codifica el antígeno. En algunos aspectos, la al menos una secuencia de poli(A) tiene al menos 20, al menos 30, al menos 40, al menos 50, al menos 60, al menos 70, al menos 80 o al menos 90 nucleótidos A consecutivos. En algunos aspectos, la al menos una secuencia de poli(A) tiene al menos 100 nucleótidos A consecutivos.

[0056] En algunos aspectos, el casete de neoantígeno comprende además al menos uno de: una secuencia de intrón, una secuencia de elemento regulador postranscripcional del virus de la hepatitis de la marmota (WPRE), una secuencia de entrada interna al ribosoma (IRES), una secuencia de nucleótidos que codifica una secuencia de péptidos 2A autoescindible, una secuencia de nucleótidos que codifica un sitio de escisión de furina o una secuencia en la región no codificante 5’ o 3’conocida por mejorar la eficiencia de traducción, estabilidad o exportación nuclear del ARNm que se une de forma operativa a al menos una secuencia de ácido nucleico que codifica el antígeno.

[0057] En algunos aspectos, el casete de neoantígeno comprende además un gen reportero, que incluye, de modo no taxativo, proteína fluorescente verde (GFP), una variante de GFP, fosfatasa alcalina secretada, luciferasa, una variante de luciferasa o un péptido o epítopo detectable. En algunos aspectos, el péptido o epítopo detectable se selecciona del grupo que consiste en una etiqueta HA, una etiqueta Flag, una etiqueta His o una etiqueta V5.

[0058] En algunos aspectos, el uno o más vectores comprenden, además, una o más secuencias de ácidos nucleicos que codifican al menos un modulador inmunitario. En algunos aspectos, el modulador inmunitario es un anticuerpo anti-CTLA4 o un fragmento de unión al antígeno de este, un anticuerpo

anti-PD-1 o un fragmento de unión al antígeno de este, un anticuerpo anti-PD-L1 o un fragmento de unión al antígeno de este, un anticuerpo anti-4-1BB o un fragmento de unión al antígeno de este o un anticuerpo anti-OX-40 o un fragmento de unión al antígeno de este. En algunos aspectos, el anticuerpo o fragmento de unión al antígeno de este es un fragmento Fab, un fragmento Fab', un Fv de cadena simple (scFv), un anticuerpo de dominio simple (sdAb) ya sea como simple específico o múltiples especificidades enlazados entre sí (p. ej., dominio de anticuerpo de camélido), o anticuerpo de cadena simple de longitud completa (p. ej., IgG de longitud completa con cadenas ligera y pesada enlazadas por un enlazador flexible). En algunos aspectos, las secuencias de cadena ligera y pesada del anticuerpo son una secuencia contigua separada por una secuencia de autoescisión tal como 2A o IRES; o las secuencias de cadena ligera y pesada del anticuerpo están enlazadas por un enlazador flexible tal como residuos de glicina consecutivos.

[0059] En algunos casos, el modulador inmunitario es una citocina. En algunos aspectos, la citocina es al menos una de IL-2, IL-7, IL-12, IL-15 o IL-21 o variantes de cada una de estas.

[0060] Además, en la presente se describe un vector de adenovirus que comprende un casete de neoantígeno, donde el casete de neoantígeno comprende: una pluralidad de secuencias de ácido nucleico que codifican el antígeno derivadas de un tumor presente dentro de un sujeto, que comprende: al menos dos secuencias de ácido nucleico que codifican el neoantígeno del MHC clase II, donde cada una comprende al menos una alteración que la distingue de la secuencia de ácido nucleico original de tipo salvaje correspondiente, y opcionalmente, al menos una secuencia de ácido nucleico que codifica el antígeno del MHC clase II; y al menos una secuencia promotora unida de manera operativa a al menos una secuencia de la pluralidad.

[0061] En algunos aspectos, el vector de adeovirus es un vector de adenovirus de chimpancé (ChAd), opcionalmente un vector C68. En algunos aspectos, el vector de adenovirus comprende la secuencia que se establece en las SEQ ID NO:1. En algunos aspectos, el vector de adenovirus comprende la secuencia que se establece en SEQ ID NO:1, con la excepción de que la secuencia se elimina completamente o se elimina funcionalmente en al menos un gen seleccionado del grupo que consiste en los genes E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 de adenovirus de chimpancé de la secuencia que se establece en SEQ ID NO: 1, donde opcionalmente la secuencia se elimina completamente o se elimina funcionalmente en: (1) E1A y E1B; (2) E1A, E1B y E3; o (3) E1A, E1B, E3 y E4 de la secuencia que se establece en SEQ ID NO: 1. En algunos aspectos, el vector de adenovirus comprende un gen o secuencia reguladora obtenida de la secuencia de SEQ ID NO: 1, donde opcionalmente el gen se selecciona del grupo que consiste en los genes de repetición terminal invertida (ITR), E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 del adenovirus del chimpancé de la secuencia que se establece en SEQ ID NO: 1.

[0062] En algunos aspectos, el casete de neoantígeno se inserta en el vector de adenovirus en la región E1, región E3 y/o cualquier región AdV eliminada que permita la incorporación del casete de neoantígeno.

[0063] En algunos aspectos, la al menos una secuencia promotora del vector de adenovirus es inducible. En algunos aspectos, la al menos una secuencia promotora del vector de adenovirus es no inducible. En algunos aspectos, la al menos una secuencia promotora del vector de adenovirus es una secuencia promotora de CMV, SV40, EF-1, RSV, PGK o EBV.

[0064] En algunos aspectos, el casete de neoantígeno del vector de adenovirus comprende, además, al menos una secuencia de poliA unida de forma operativa a al menos una de las secuencias de la pluralidad, donde opcionalmente la secuencia de poliA se encuentra en una ubicación 3' con respecto a la al menos una secuencia de la pluralidad.

[0065] En algunos aspectos, el vector de adenovirus se genera a partir de un vector adenoviral de primera generación, de segunda generación o dependiente de auxiliares.

[0066] En algunos aspectos, el vector de adenovirus comprende una o más eliminaciones entre el par de bases número 577 y 3407, y donde opcionalmente el vector de adenovirus comprende además una o más eliminaciones entre el bar de bases 27.141 y 32.022 o ente el par de bases 27.816 y 31.332 de la secuencia que se establece en SEQ ID NO:1. En algunos aspectos, el vector de adenovirus comprende además una o más eliminaciones entre el par de bases número 3957 y 10346, par de bases 21787 y 23370, y par de bases número 33486 y 36193 de la secuencia que se establece en SEQ ID NO:1.

[0067] En algunos aspectos, la al menos una secuencia de ácido nucleico que codifica el neoantígeno del MHC clase I se selecciona al llevar a cabo las etapas de: (a) obtener los datos de secuenciación de nucleótidos tumorales de al menos uno de exoma, transcriptoma o genoma completo del tumor, donde los datos de secuenciación de nucleótidos tumorales se usan para obtener datos que representan las secuencias de péptidos de cada uno de un conjunto de neoantígenos; (b) ingresar la secuencia de péptidos de cada neoantígeno en un modelo de presentación para generar un conjunto de probabilidades numéricas de que cada uno de los neoantígenos se presente por uno o más de los alelos del MHC en la superficie de células tumorales del tumor, donde el conjunto de probabilidades numéricas se identificó al menos en función de los datos de espectrometría de masas recibidos; y (c) seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades numéricas para generar un conjunto de neoantígenos seleccionados que se usan para generar la al menos una secuencia de ácido nucleico que codifica el neoantígeno del MHC clase I.

[0068] En algunos aspectos, cada una de la al menos una secuencia de ácido nucleico que codifica el neoantígeno del MHC clase I se selecciona al llevar a cabo las etapas de: (a) obtener los datos de

secuenciación de nucleótidos tumorales de al menos uno de exoma, transcriptoma o genoma completo del tumor, donde los datos de secuenciación de nucleótidos tumorales se usan para obtener datos que representan las secuencias de péptidos de cada uno de un conjunto de neoantígenos; (b) ingresar la secuencia de péptidos de cada neoantígeno en un modelo de presentación para generar un conjunto de probabilidades numéricas de que cada uno de los neoantígenos se presente por uno o más de los alelos del MHC en la superficie de la célula tumoral, donde el conjunto de probabilidades numéricas se identificó al menos en función de los datos de espectrometría de masas recibidos; y (c) seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades numéricas para generar un conjunto de neoantígenos seleccionados que se usan para generar la al menos una secuencia de ácido nucleico que codifica el neoantígeno del MHC clase I.

[0069] En algunos aspectos, una cantidad del conjunto de neoantígenos seleccionados es 2-20.

[0070] En algunos aspectos, el modelo de presentación representa dependencia entre; la presencia de un par de un alelo del MHC particular y un aminoácido particular en una posición particular de una secuencia de péptidos; y la probabilidad de presentación en la superficie de células tumorales, por el alelo de MHC particular del par, de tal secuencia de péptidos que comprende el aminoácido particular en la posición particular.

[0071] En algunos aspectos, seleccionar el conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con mayor probabilidad de presentarse en la superficie de células tumorales respecto de neoantígenos no seleccionados en función del modelo de presentación. En algunos aspectos, seleccionar el conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con mayor probabilidad de poder inducir una respuesta inmunitaria específica de tumor en el sujeto respecto de neoantígenos no seleccionados en función del modelo de presentación. En algunos aspectos, seleccionar el conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con mayor probabilidad de poder presentarse a linfocitos T sin tratamiento por células que presentan antígenos (APC) profesionales respecto de neoantígenos no seleccionados en función del modelo de presentación, donde opcionalmente la APC es una célula dendrítica (DC). En algunos aspectos, seleccionar el conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con mayor probabilidad de someterse a inhibición mediante tolerancia central o periférica respecto de neoantígenos no seleccionados en función del modelo de presentación. En algunos aspectos, seleccionar el conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con menor probabilidad de poder inducir una respuesta autoinmunitaria al tejido normal en el sujeto respecto de neoantígenos no seleccionados en función del modelo de presentación. En algunos aspectos, los datos de secuenciación de nucleótidos de exoma o transcriptoma se obtienen al realizar secuenciación en el tejido tumoral. En algunos aspectos, la secuenciación es secuenciación de nueva generación (NGS) o cualquier enfoque

de secuenciación masiva en paralelo.

[0072] En algunos aspectos, el casete de neoantígeno comprende secuencias de epítopo de unión formadas por secuencias adyacentes en el casete de neoantígeno. En algunos aspectos, la o cada secuencia de epítopo de unión tiene una afinidad mayor que 500 nM para MHC. En algunos aspectos, cada secuencia de epítopo de unión es no propia. En algunos aspectos, el casete de neoantígeno no codifica una secuencia de ácidos nucleicos de epítopo del MHC clase I o clase II no terapéutico que comprende una secuencia de ácidos nucleicos de tipo salvaje traducida, donde se predice que el epítopo no terapéutico aparecerá en un alelo del MHC del sujeto. En algunos aspectos, la secuencia de epítopo del MHC clase I o clase II no terapéutico predicho es una secuencia de epítopo de unión formada por secuencias adyacentes en el casete de neoantígeno. En algunos aspectos, la predicción se basa en probabilidades de presentación generadas al ingresar secuencias de los epítopos no terapéuticos en un modelo de presentación. En algunos aspectos, un orden de la al menos una secuencia de ácido nucleico que codifica el antígeno en el casete de neoantígeno se determina mediante una serie de etapas que comprenden: (a) generar un conjunto de secuencias de casete de neoantígeno correspondiente a diferentes órdenes de las múltiples secuencias de ácidos nucleicos que codifican el antígeno; (b) determinar, para cada secuencia de casete de neoantígeno candidato, una puntuación de presentación en función de la presentación de epítopos no terapéuticos en la secuencia de casete de neoantígeno candidato y (c) seleccionar una secuencia de casete candidato asociada con una puntuación de presentación por debajo de un umbral predeterminado como la secuencia de casete de neoantígeno para una vacuna de neoantígeno.

[0073] También se describe en la presente una composición farmacéutica que comprende cualquiera de las composiciones descritas en la presente (tal como un vector basado en ChAd o en un alfavirus descrito en la presente) y un portador farmacéuticamente aceptable. En algunos aspectos, la composición farmacéutica comprende adicionalmente un adyuvante. En algunos aspectos, la composición farmacéutica comprende además un modulador inmunitario. En algunos aspectos, el modulador inmunitario es un anticuerpo anti-CTLA4 o un fragmento de unión al antígeno de este, un anticuerpo anti-PD-1 o un fragmento de unión al antígeno de este, un anticuerpo anti-PD-L1 o un fragmento de unión al antígeno de este, un anticuerpo anti-4-1BB o un fragmento de unión al antígeno de este o un anticuerpo anti-OX-40 o un fragmento de unión al antígeno de este.

[0074] También se describe en la presente una secuencia de nucleótidos aislada o conjunto de secuencias de nucleótidos aisladas que comprenden el casete de neoantígenos de cualquiera de las reivindicaciones de composición anteriores y uno o más elementos obtenidos de la secuencia de la SEQ ID NO:3 o la SEQ ID NO:5, opcionalmente donde el uno o más elementos se seleccionan del grupo que consiste en la secuencias necesarias para la amplificación mediada por proteínas no estructurales, la

secuencia promotora de nucleótidos 26S, la secuencia de poli(A) y los genes nsP1-4 de la secuencia establecida en la SEQ ID NO:3 o SEQ ID NO:5; y opcionalmente donde la secuencia de nucleótidos es ADNc. En algunos aspectos, la secuencia o conjunto de secuencias de nucleótidos aisladas comprende un casete de neoantígeno descrito en la presente insertado en la posición 7544 de las secuencias establecidas en las SEQ ID NO:6 o SEQ ID NO:7. En algunos aspectos, la secuencia de nucleótidos aislada comprende, además, una secuencia promotora de nucleótidos de polimerasa de ARN T7 o SP6 5' del uno o más elementos obtenidos de la secuencia de la SEQ ID NO:3 o la SEQ ID NO:5 y, opcionalmente, uno o más sitios de restricción 3' de la secuencia de poli(A). En algunos aspectos, el casete de neoantígeno descrito en la presente se inserta en la posición 7563 de la SEQ ID NO:8 o SEQ ID NO:9. En otro aspecto, las secuencias establecidas en las SEQ ID NO:8 o SEQ ID NO:9 comprenden, además, un nucleótido de adenina adicional insertado en la posición 17.

[0075] También se describe en la presente una secuencia de nucleótidos aislada que comprende un casete de neoantígeno descrito en la presente y al menos un promotor descrito en la presente. En algunos aspectos, la secuencia de nucleótidos aislada comprende además un gen basado en ChAd En algunos aspectos, el gen basado en ChAd se obtiene de la secuencia de SEQ ID NO: 1, donde opcionalmente el gen se selecciona dentro del grupo que consiste en los genes ITR, E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 del adenovirus del chimpancé de la secuencia que se establece en la SEQ ID NO: 1, y donde opcionalmente la secuencia de nucleótidos es ADNc.

[0076] También se describe en la presente una célula aislada que comprende la secuencia de nucleótidos aislada descrita en la presente, donde opcionalmente la célula es una célula BHK-21, CHO, HEK293 o variantes de esta, 911, HeLa, A549, LP-293, PER.C6 o AE1-2a.

[0077] También se describe en la presente un vector que comprende una secuencia de nucleótidos aislada descrita en la presente.

[0078] También se describe en la presente un kit que comprende un vector o una composición descritos en la presente e instrucciones de uso.

[0079] También se describe en la presente un método para tratar a un sujeto con cáncer, donde el método comprende administrar al sujeto un vector descrito en la presente o una composición farmacéutica descrita en la presente. En algunos aspectos, la al menos una secuencia de ácido nucleico que codifica un neoantígeno del MHC clase I derivada de un tumor se deriva del tumor del sujeto con cáncer. En algunos aspectos, la al menos una secuencia de ácido nucleico que codifica un neoantígeno del MHC clase I no se deriva del tumor del sujeto con cáncer.

[0080] También se describe en la presente un método para inducir una respuesta inmunitaria en un sujeto, donde el método comprende administrarle al sujeto cualquiera de las composiciones, vectores o composiciones farmacéuticas descritas en la presente.

[0081] En algunos aspectos, el vector o la composición se administran por vía intramuscular (IM), intradérmica (ID) o subcutánea (SC), o intravenosa (IV).

[0082] En algunos aspectos, los métodos descritos en la presente, comprenden, además, l administración de uno o más moduladores inmunitarios, donde opcionalmente el modulado inmunitario se administra antes de, al mismo tiempo que, o después de la administración del vector o la composición farmacéutica. En algunos aspectos, el uno o más moduladores inmunitarios se selecciona del grupo que consiste en un anticuerpo anti-CTLA4 o un fragmento de unión al antígeno de este, un anticuerpo anti-PD-1 o un fragmento de unión al antígeno de este, un anticuerpo anti-PD-L1 o un fragmento de unión al antígeno de este, un anticuerpo anti-4-1BB o un fragmento de unión al antígeno de este o un anticuerpo anti-OX-40 o un fragmento de unión al antígeno de este. En algunos aspectos, el modulador inmunitario se administra por vía intravenosa (IV), intramuscular (IM), intradérmica (ID) o subcutánea (SC). En algunos aspectos, la administración subcutánea se encuentra cerca del sitio de administración del la composiciónfarmacéutica o en proximidad estrecha con respecto a uno o más ganglios linfáticos de drenaje del vector o de la composición.

[0083] En algunos aspectos, los métodos descritos en la presente comprenden, además, administrarle al sujeto una segunda composición de vacuna. En algunos aspectos, la segunda composición de vacuna se administra antes de la administración de la composición o la composición farmacéutica descrita anteriormente. En algunos aspectos, la segunda composición de vacuna se administra después de la administración de la composición o la composición farmacéutica descrita anteriormente. En algunos aspectos, la segunda composición de vacuna es la misma que la composición o la composición farmacéutica descrita anteriormente. En algunos aspectos, la segunda composición de vacuna es diferente la composición o la composición farmacéutica descrita anteriormente. En algunos aspectos, la segunda composición de vacuna comprende un vector de adenovirus de chimpancé que codifica al menos una secuencia de ácido nucleico que codifica el neoantígeno. En algunos aspectos, la al menos una secuencia de ácido nucleico que codifica el neoantígeno codificada por el vector de adenovirus de chimpancé es igual que la al menos una secuencia de ácido nucleico que codifica el neoantígeno de cualquiera de las composiciones o vectores anteriores.

[0084] También se describe en la presente un método para elaborar el uno o más vectores de cualquiera de las composiciones anteriores, donde el método comprende: obtener una secuencia de ADN linealizada que comprende la estructura principal de alfavirus de ARN y el casete de neoantígeno; transcribir *in vitro* la secuencia de ADN linealizada mediante la adición de la secuencia de ADN linealizada a una reacción de transcripción *in vitro* que contiene todos los componentes necesarios para transcribir la secuencia de ADN linealizada en ARN, que comprende opcionalmente de manera adicional la adición *in vitro* del casquete m7g al ARN resultante; y aislar el uno o más vectores de la reacción de

transcripción *in vitro*. En algunos aspectos, la secuencia de ADN linealizada se genera mediante la linealización de una secuencia de plásmido de ADN o mediante amplificación mediante el uso de PCR. En algunos aspectos, la secuencia de plásmido de ADN se genera mediante el uso de uno de recombinación bacteriana o síntesis de ADN de genoma completo o síntesis de ADN de genoma completo con amplificación del ADN sintetizado en células bacterianas. En algunos aspectos, el aislamiento del uno o más vectores a partir de la reacción de transcripción *in vitro* implica uno o más de extracción de fenol cloroformo, purificación basada en columna de sílice o métodos de purificación de ARN similares.

[0085] También se describe en la presente un método para elaborar cualquiera de las composiciones descritas en la presente, donde el método comprende: proporcionar componentes para el vehículo de administración de nanopartículas; proporcionar el sistema de expresión de neoantígenos; y proporcionar condiciones suficientes para que el vehículo de administración de nanopartículas y el sistema de expresión de neoantígenos produzcan la composición para administración del sistema de expresión de neoantígenos. En algunos aspectos, las condiciones se proporcionan mediante mezcla microfluídica.

[0086] También se describe en la presente un método de elaboración de un vector de adenovirus descrito en la presente, donde el método comprende: obtener una secuencia plasmídica que comprende la o las secuencias promotoras y el casete de neoantígeno; transfectar la secuencia plasmídica en una o más células hospedadoras; y aislar el vector de adenovirus de la o las células hospedadoras.

[0087] En algunos aspectos, aislar comprende: lisar la célula hospedadora para obtener un lisado celular que comprenda el vector de adenovirus; y purificar el vector de adenovirus del lisado celular.

[0088] En algunos aspectos, la secuencia plasmídica se genera mediante el uso de recombinación bacteriana o síntesis de ADN de genoma completo o síntesis de ADN de genoma completo con amplificación del ADN sintetizado en células bacterianas. En algunos aspectos, la o las células hospedadoras son células CHO, HEK293 o variantes de estas, 911, HeLa, A549, LP-293, PER.C6 y/o AE1-2a. En algunos aspectos, purificar el vector de adenovirus del lisado celular implica una o más de separación cromatográfica, centrifugación, precipitación del virus y/o filtración.

BREVE DESCRIPCIÓN DE LAS DISTINTAS VISTAS DE LAS FIGURAS

[0089] Estos y otros aspectos, características y ventajas de la presente invención se comprenderán mejor con referencia a la siguiente descripción y figuras que la acompañan donde:

[0090] **La Figura (FIG.) 1A** muestra enfoques clínicos actuales para la identificación de neoantígenos.

[0091] **La FIG. 1B** muestra que <5 % de los péptidos unidos predichos se presentan en células tumorales.

[0092] **La FIG. 1C** muestra el impacto del problema de especificidad de predicción de neoantígeno.

[0093] **La FIG. 1D** muestra que la predicción de unión no es suficiente para la identificación de neoantígenos.

[0094] **La FIG. 1E** muestra la probabilidad de presentación de MHC-I en función de la longitud de péptidos.

[0095] **La FIG. 1F** muestra un ejemplo de espectro de péptidos generado a partir del rango dinámico estándar de Promega. Esta figura muestra la SEQ ID NO: 59.

[0096] **La FIG. 1G** muestra cómo la adición de características aumenta el valor predictivo positivo del modelo.

[0097] **La FIG. 2A** es una vista general de un entorno para identificar probabilidades de presentación de péptidos en pacientes, de acuerdo con una modalidad.

[0098] **Las FIG. 2B** y **FIG. 2C** ilustran un método para obtener información de presentación, de acuerdo con una modalidad. La **FIG. 2B** describe la SEC ID NO: 62. La **FIG. 2C** divulga la SEQ ID NOS 62-67, respectivamente, en orden de aparición. **FIG. 2D** revela la SEC ID NO: 157. La **FIG. 2E** divulga la SEQ ID NOS 62-65, 68 y 67, respectivamente, en orden de aparición.

[0099] **La FIG. 3** es un diagrama de bloques de alto nivel que ilustra los componentes lógicos de computadora del sistema de identificación de presentación, de acuerdo con una modalidad.

[00100] **La FIG. 4A** ilustra un conjunto de datos de entrenamiento de ejemplo, de acuerdo con una modalidad relacionada con los alelos del MHC clase I. La **FIG. 4A** describe secuencias de péptidos como SEQ ID NOS 70-73 y secuencias de flanqueo-C como SEQ ID NOS 74, 158, 159 y 159, respectivamente, en orden de aparición. La **FIG. 4B** ilustra un conjunto de datos de entrenamiento ilustrativo, de acuerdo con una modalidad relacionada con un alelo del MHC de clase II. La **FIG. 4B** describe la SEQ ID NO: 75.

[00101] **La FIG. 5** ilustra un modelo de red de ejemplo asociado con un alelo de MHC.

[00102] **La FIG. 6A** ilustra un modelo de red de ejemplo NNH(·) compartido por alelos de MHC de acuerdo con una modalidad. La **FIG. 6B** ilustra un ejemplo de modelo de red $NN_H(\cdot)$ compartido por los alelos del MHC, de acuerdo con otra modalidad.

[00103] **La FIG. 7** ilustra la generación de una probabilidad de presentación para un péptido asociado con un alelo del MHC mediante el uso de un modelo de red de ejemplo.

[00104] **La FIG. 8** ilustra la generación de una probabilidad de presentación para un péptido asociado con un alelo del MHC mediante el uso de modelos de red de ejemplo.

[00105] **La FIG. 9** ilustra la generación de una probabilidad de presentación para un péptido asociado con alelos del MHC mediante el uso de modelos de red de ejemplo.

[00106] **La FIG. 10** ilustra la generación de una probabilidad de presentación para un péptido

asociado con alelos del MHC mediante el uso de modelos de red de ejemplo.

[00107] **La FIG. 11** ilustra la generación de una probabilidad de presentación para un péptido asociado con alelos del MHC mediante el uso de modelos de red de ejemplo.

[00108] **La FIG. 12** ilustra la generación de una probabilidad de presentación para un péptido asociado con alelos del MHC mediante el uso de modelos de red de ejemplo.

[00109] **La FIG. 13A** muestra los resultados de rendimiento para la presentación de péptidos determinados mediante espectrometría de masas al comparar varios modelos de presentación de péptidos. Se muestran los resultados para el máximo del modelo de presentación por alelos que se muestra en la ecuación (12) mediante el uso de la función de dependencia afín $g_h(\cdot)$ y la función de expit $f(\cdot)$ y se entrenaron en un subconjunto de datos de espectrometría de masas para HLA-A*02:01 y HLA-B*07:02 (“MS”). También se muestran modelos del estado de la técnica basados en predicciones de afinidad NETMHCpan “Afinidad” y basados en predicciones de estabilidad NETMHCstab “Estabilidad”. Los datos muestran el valor predictivo positivo (PPV) a una tasa de memoria de 10 % y las barras de error (como se indican con líneas continuas) muestran intervalos de confianza de 95 %.

[00110] **La FIG. 13B** muestra los resultados de rendimiento para la presentación de péptidos determinados mediante epítopos de linfocitos T al comparar varios modelos de presentación de péptidos. Se muestran los resultados para el máximo del modelo de presentación por alelos que se muestra en la ecuación (12) mediante el uso de la función de dependencia afín $g_h(\cdot)$ y la función de expit $f(\cdot)$ y se entrenaron en un subconjunto de datos de espectrometría de masas para HLA-A*02:01. También se muestran modelos del estado de la técnica basados en predicciones de afinidad NETMHCpan “Afinidad” y basados en predicciones de estabilidad NETMHCstab “Estabilidad”. Los datos muestran el valor predictivo positivo (PPV) a una tasa de memoria de 10 % y las barras de error (como se indican con líneas continuas) muestran intervalos de confianza de 95 %.

[00111] **La FIG. 13C** muestra los resultados de rendimiento para la presentación de péptidos determinados mediante espectrometría de masas para un modelo ilustrativo de función de sumas (ecuación (13)), un modelo ilustrativo de suma de funciones (ecuación (19)) y un modelo ilustrativo de segundo orden (ecuación (23)) para la predicción de la presentación de péptidos. La primera columna hace referencia al área bajo la curva (AUC) de la característica operativa del receptor (ROC) cuando se aplicó cada modelo de presentación al conjunto de prueba, la segunda columna hacer referencia al valor de pérdida de probabilidad logarítmica negativa y la tercera columna hace referencia al valor predictivo positivo (PPV) con una tasa de memoria de 10 %.

[00112] **La FIG. 13D** muestra los resultados de rendimiento para la presentación de péptidos determinados mediante espectrometría de masas para dos modelos de presentación ilustrativos que se entrenan con y sin datos de espectrometría de masas de alelo simple La primera columna hace

referencia al área bajo la curva (AUC) de la característica operativa del receptor (ROC) cuando se aplicó cada modelo de presentación al conjunto de prueba, la segunda columna hacer referencia al valor de pérdida de probabilidad logarítmica negativa y la tercera columna hace referencia al valor predictivo positivo (PPV) con una tasa de memoria de 10 %.

[00113] **La FIG. 13E** muestra los resultados de rendimiento para la presentación de péptidos determinados mediante espectrometría de masas para dos modelos de presentación ilustrativos que se entrenan con y sin datos de espectrometría de masas de alelo simple “Correlación” se refiere a la correlación entre las etiquetas reales que indican si el péptido se presentó en el alelo correspondiente en los datos de prueba y la etiqueta para predicción.

[00114] **La FIG. 13F** muestra la frecuencia de residuos de anclaje comunes en las posiciones 2 (P2) y 9 (P9) entre nonámeros previstos mediante un modelo de presentación entrenado sin datos de espectrometría de masas de alelo simple.

[00115] **La FIG. 13G** muestra resultados del rendimiento para la presentación de péptidos determinados mediante espectrometría de masas para un modelo ilustrativo que incorporaba las secuencias flanqueantes de extremo C y N como variables de interacción con alelos y un modelo de presentación ilustrativo que incorporaba secuencias flanqueantes de extremo C y N como variables de no interacción con alelos. La primera columna hace referencia al área bajo la curva (AUC) de la característica operativa del receptor (ROC) cuando se aplicó cada modelo de presentación al conjunto de prueba, la segunda columna hacer referencia al valor de pérdida de probabilidad logarítmica negativa y la tercera columna hace referencia al valor predictivo positivo (PPV) con una tasa de memoria de 10 %.

[00116] **La FIG. 13H** muestra la dependencia entre la abundancia de ARNm y la frecuencia de los péptidos presentados en una célula tumoral según se determina mediante espectrometría de masas. El eje horizontal indica la expresión de ARNm en función de cuartiles de transcripción por millón (TPM). El eje vertical indica la fracción de epítopos presentados a partir de genes en cuartiles de expresión de ARNm correspondientes. Cada línea continua es una gráfica relacionada con las dos mediciones a partir de una muestra tumoral que se relaciona con los datos de espectrometría de masas correspondientes y con las mediciones de expresión de ARNm.

[00117] **La FIG. 13I** muestra resultados de rendimiento para la presentación de péptidos determinados mediante espectrometría de masas para modelos de presentación ilustrativos “filtro de MHCflurry + ARN” que es un modelo similar al modelo actual del estado de la técnica que predice la presentación basándose en predicciones de afinidad con un filtro de expresión génica estándar que eliminó todos los péptidos de las proteínas con mediciones de cuantificación de ARNm que fueran menores que 3,2 FPKM. El modelo denominado “Modelo Ilustrativo, sin ARN” es el modelo de presentación ilustrativo de “suma de sigmoides” que se muestra en la ecuación (21). El modelo

denominado “Modelo Ilustrativo, con ARN” es el modelo de presentación de “suma de sigmoides” que se muestra en la ecuación (19) que incorpora datos de cuantificación de ARNm a través de una función logarítmica. Los datos muestran el valor predictivo positivo (PPV) a una tasa de memoria de 20 %.

[00118] **La FIG. 13J** muestra la probabilidad de presentación de péptidos para diferentes longitudes de péptido para modelos de presentación que toman en cuenta la longitud del péptido y modelos del estado de la técnica que no toman en cuenta la longitud del péptido al predecir la presentación de péptidos. La gráfica “Verdadera (datos de prueba ciega)” mostraron la proporción de péptidos presentados de acuerdo con la longitud del péptido en una muestra del conjunto de datos de prueba de muestra. La gráfica “Modelos que ignoran longitud” indicó mediciones previstas si se aplicaran modelos del estado de la técnica que ignoran la longitud del péptido al mismo conjunto de datos de prueba para la predicción de la presentación. El modelo denominado “Modelo Ilustrativo, con ARN” es el modelo de presentación de “suma de sigmoides” que se muestra en la ecuación (19) que incorpora datos de cuantificación de ARNm a través de una función logarítmica.

[00119] **La FIG. 13K** es un histograma de longitudes de péptidos eluidos de alelos del MHC clase II en células tumorales humanas y linfocitos infiltrantes de tumores (TIL) mediante el uso de espectrometría de masas.

[00120] **La FIG. 13L** ilustra la dependencia entre la cuantificación de ARNm y los péptidos presentados por residuo para dos ejemplos de conjuntos de datos.

[00121] **La FIG. 13M** compara los resultados de rendimiento, por ejemplo, modelos de presentación entrenados y evaluados mediante el uso de dos ejemplos de conjuntos de datos.

[00122] **La FIG. 13N** es un histograma que representa la cantidad de péptidos secuenciados mediante el uso de espectrometría de masas para cada muestra de un total de 39 muestras que comprende moléculas de HLA clase II.

[00123] **La FIG. 13-O** es un histograma que representa la cantidad de muestras en las que se identificó un alelo particular de molécula del MHC clase II.

[00124] **La FIG. 13P** es un histograma que representa la proporción de péptidos presentados por las moléculas del MHC clase II en las 39 muestras totales, para cada longitud de péptido de una gama de longitudes de péptidos.

[00125] **La FIG. 13Q** es una gráfica lineal que representa la relación entre la expresión génica y la prevalencia de la presentación del producto de expresión génica por parte de una molécula del MHC clase II, para los genes presentes en las 39 muestras.

[00126] **La FIG. 13R** es una gráfica lineal que compara el rendimiento de modelos idénticos con diversas entradas en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por una molécula del MHC clase II.

[00127] **La FIG. 13S** es una gráfica lineal que compara el rendimiento de cuatro modelos distintos en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por una molécula del MHC clase II.

[00128] **La FIG. 13T** es una gráfica lineal que compara el rendimiento de un modelo de primera clase de la técnica previa con dos criterios distintos y la presentación del modelo descrito en la presente con dos entradas diferentes en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por una molécula del MHC clase II.

[00129] **La FIG. 14** ilustra una computadora de ejemplo para la implementación de las entidades mostradas en las Figuras 1 y 3.

[00130] **La FIG. 15** ilustra el desarrollo de un ensayo de activación de linfocitos T *in vitro*. Esquema del ensayo en el que la administración de un casete de vacuna a células presentadoras de antígenos conduce a la expresión, procesamiento y presentación restringida de MHC de distintos antígenos de péptidos. Los linfocitos T indicadores modificados genéticamente con receptores de linfocitos T que coinciden con la combinación péptido-MHC específica se activan, dando como resultado expresión de luciferasa.

[00131] **La FIG 16A** ilustra la evaluación de secuencias enlazadoras en casetes cortos y muestra cinco epítopos restringidos de MHC clase I (epítopos 1 a 5) concatenados en la misma posición uno con respecto al otro seguidos de dos epítopos de MHC clase II universales (MHC-II). Se generaron diversas repeticiones mediante el uso de diferentes enlazadores. En algunos casos, los epítopos de linfocitos T están directamente enlazados entre sí. En otros, los epítopos de linfocitos T están flanqueados en uno o ambos lados por su secuencia natural. En otras repeticiones, los epítopos de linfocitos T están enlazados por las secuencias no naturales AAY, RR y DPP.

[00132] **La FIG 16B** ilustra la evaluación de secuencias enlazadoras en casetes cortos y muestra información de secuencia sobre los epítopos de linfocitos T integrados en los casetes cortos. La figura describe la SEQ ID NOS 132, 133, 136, 135, 134, 160 y 161, respectivamente, en orden de aparición.

[00133] **La FIG. 17** ilustra evaluación de secuencias de direccionamiento celular agregadas a casetes de vacuna modelo. Los casetes de direccionamiento extienden los diseños de casete corto con dominios de ubiquitina (Ub), péptidos de señal (SP) y/o transmembrana (TM), presentan a continuación de los cinco epítopos de linfocitos T humanos marcadores (epítopos 1 a 5) también dos epítopos de linfocitos T de ratón SIINFEKL (SII) (SEQ ID NO: 57) y SPSYAYHQF (A5) (SEQ ID NO: 58), y usan el enlazador no natural AAY- o enlazadores naturales que flanquean los epítopos de linfocitos T en ambos lados (25mer).

[00134] **La FIG. 18** ilustra evaluación *in vivo* de las secuencias enlazadoras en casetes cortos. A) Diseño experimental de la evaluación *in vivo* de casetes de vacuna mediante el uso de ratones

transgénicos HLA-A2.

[00135] **La FIG. 19A** ilustra la evaluación *in vivo* del impacto de la posición del epítopo en casetes largos 21-mer y muestra que el diseño de casetes largos implica epítopos clase I de marcador cinco (epítopos 1 a 5) contenidos en su secuencia natural 25-mer (enlazador = secuencias flanqueantes naturales), separados con epítopos de linfocitos T clase I muy conocidos adicionales (epítopos 6 a 21) contenidos en su secuencia natural 25-mer y dos epítopos clase II universales (MHC-II0), solamente con la posición relativa de los epítopos clase I variada.

[00136] **La FIG. 19B** ilustra evaluación *in vivo* del impacto de la posición del epítopo en casetes largos 21-mer y muestra la información de secuencia en los epítopos de linfocitos T usados. La figura describe la SEQ ID NOS 132, 133, 136, 135, 134, 162-164, 137 y 165-176, respectivamente, en orden de aparición.

[00137] **La FIG. 20A** ilustra el diseño de casete final para estudios preclínicos para permitir IND (nuevo fármaco en investigación) y muestra que el diseño del casete final comprende 20 epítopos del MHC I contenidos en su secuencia natural 25-mer (enlazador = secuencias naturales flanqueantes), compuesto por 6 epítopos de primates no humanos (NHP), 5 epítopos humanos, 9 epítopos murinos, así como también 2 epítopos del MHC clase II universales.

[00138] **La FIG. 20B** ilustra el diseño de casete final para estudios preclínicos para permitir IND y muestra la información de secuencia para los epítopos de linfocitos T usados que se presentan en MHC clase I de origen primates no humanos (SEQ ID NOS 177-182, respectivamente, en orden de aparición), ratón (SEQ ID NOS 57, 58 y 183-189, respectivamente, en orden de aparición) y humano (SEQ ID NOS 134-136, 132 y 133, respectivamente, en orden de aparición), así como secuencias de 2 epítopos del MHC clase II universales PADRE y toxoide tetánico (SEQ ID NOS 160 y 190, respectivamente, en orden de aparición).

[00139] **La FIG. 21A** ilustra la producción del virus ChAdV68.4WTnt.GFP después de la transfección. Se transfectaron células HEK293A con ADN de ChAdV68.4WTnt.GFP mediante el uso del protocolo de fosfato de calcio. La replicación viral se observó 10 días después de la transfección y las placas virales de ChAdV68.4WTnt.GFP se visualizaron mediante el uso de microscopia de luz (aumento de 40x).

[00140] **La FIG. 21B** ilustra la producción del virus ChAdV68.4WTnt.GFP después de la transfección. Se transfectaron células HEK293A con ADN de ChAdV68.4WTnt.GFP mediante el uso del protocolo de fosfato de calcio. La replicación viral se observó 10 días después de la transfección y las placas virales de ChAdV68.4WTnt.GFP se visualizaron mediante el uso de microscopia fluorescente a un aumento de 40x.

[00141] **La FIG. 21C** ilustra la producción del virus ChAdV68.4WTnt.GFP después de la transfección. Se transfectaron células HEK293A con ADN de ChAdV68.4WTnt.GFP mediante el uso del protocolo de

fosfato de calcio. La replicación viral se observó 10 días después de la transfección y las placas virales de ChAdV68.4WTnt.GFP se visualizaron mediante el uso de microscopia fluorescente a un aumento de 100x.

[00142] **La FIG. 22A** ilustra la producción del virus ChAdV68.5WTnt.GFP después de la transfección. Se transfectaron células HEK293A con ADN de ChAdV68.5WTnt.GFP mediante el uso del protocolo de lipofectamina. La replicación viral (placas) se observó 10 días después de la transfección. Se realizó un lisado y se usó para reinfectar un matraz T25 de células 293A. Las placas virales de ChAdV68.5WTnt.GFP se visualizaron y fotografiaron 3 días más tarde mediante el uso de microscopía de luz (aumento 40x)

[00143] **La FIG. 22B** ilustra la producción del virus ChAdV68.5WTnt.GFP después de la transfección. Se transfectaron células HEK293A con ADN de ChAdV68.5WTnt.GFP mediante el uso del protocolo de lipofectamina. La replicación viral (placas) se observó 10 días después de la transfección. Se realizó un lisado y se usó para reinfectar un matraz T25 de células 293A. Las placas virales de ChAdV68.5WTnt.GFP se visualizaron y fotografiaron 3 días más tarde mediante el uso de microscopía fluorescente con aumento 40x.

[00144] **La FIG. 22C** ilustra la producción del virus ChAdV68.5WTnt.GFP después de la transfección. Se transfectaron células HEK293A con ADN de ChAdV68.5WTnt.GFP mediante el uso del protocolo de lipofectamina. La replicación viral (placas) se observó 10 días después de la transfección. Se realizó un lisado y se usó para reinfectar un matraz T25 de células 293A. Las placas virales de ChAdV68.5WTnt.GFP se visualizaron y fotografiaron 3 días más tarde mediante el uso de microscopía fluorescente con aumento 100x.

[00145] **La FIG. 23** ilustra el esquema de producción de partícula viral.

[00146] **La FIG. 24** ilustra el vector de ARN autorreplicante (srRNA) de VEE derivado de alfavirus.

[00147] **La FIG. 25** ilustra la expresión del indicador *in vivo* después de la inoculación de ratones C57BL/6J con srRNA de VEE-Luciferasa. Se muestran imágenes representativas de señal de luciferasa después de la inoculación de ratones C57BL/6J con srRNA de VEE-Luciferasa (10 ug por ratón, inyección intramuscular bilateral, encapsulado en MC3) en diversos puntos temporales.

[00148] **La FIG. 26A** ilustra respuestas de linfocitos T medidas14 días después de la inmunización con srRNA de VEE formulado con MC3 LNP en ratones con tumores B16-OVA. Se inyectaron ratones C57BL/6J con tumores B16-OVA con 10 ug de srRNA de VEE-Luciferasa (control), srRNA de VEE-UbAAY (Vax), srRNA de VEE-Luciferasa y anti-CTLA-4 (aCTLA-4) o srRNA de VEE-UbAAY y anti-CTLA-4 (Vax + aCTLA-4). Además, todos los ratones se trataron con mAb anti-PD1 comenzando el día 7. Cada grupo consistió en 8 ratones. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 14 días después de la inmunización. Se evaluaron las respuestas de linfocitos T específicos de

SIINFEKL ("SIINFEKL" divulgado como SEQ ID NO: 57) mediante ELISPOT de IFN-gamma y se informan como células formadoras de manchas (SFC, por sus siglas en inglés) por 106 esplenocitos. Las líneas representan medianas.

[00149] **La FIG. 26B** ilustra respuestas de linfocitos T medidas14 días después de la inmunización con srRNA de VEE formulado con MC3 LNP en ratones con tumores B16-OVA. Se inyectaron ratones C57BL/6J con tumores B16-OVA con 10 ug de srRNA de VEE-Luciferasa (control), srRNA de VEE-UbAAY (Vax), srRNA de VEE-Luciferasa y anti-CTLA-4 (aCTLA-4) o srRNA de VEE-UbAAY y anti-CTLA-4 (Vax + aCTLA-4). Además, todos los ratones se trataron con mAb anti-PD1 comenzando el día 7. Cada grupo consistió en 8 ratones. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 14 días después de la inmunización. Se evaluaron las respuestas de linfocitos T específicos de SIINFEKL ("SIINFEKL" divulgado como SEQ ID NO: 57) mediante tinción con pentámeros del MHCI, informadas como células de resultado positivo para pentámeros como un porcentaje de células positivas para CD8. Las líneas representan medianas.

[00150] **La FIG. 27A** ilustra respuestas de linfocitos T específicos de antígeno después de la sensibilización/refuerzo heterólogos en ratones con tumores B16-OVA. Los ratones C57BL/6J con tumores B16-OVA se inyectaron con adenovirus que expresa GFP (Ad5-GFP) y se reforzaron con srRNA de VEE-Luciferasa formulado con MC3 LNP (Control) o Ad5-UbAAY y reforzados con srRNA de VEE-UbAAY (Vax). Tanto el grupo de control como el Vax también se trataron con un mAb de control de IgG. Un tercer grupo se trató con la sensibilización de Ad5-GFP/refuerzo de srRNA de VEE-Luciferasa en combinación con anti-CTLA-4 (aCTLA-4), mientras que el cuarto grupo se trató con la sensibilización de Ad5-UbAAY/refuerzo de VEE-UbAAY en combinación con anti-CTLA-4 (Vax + aCTLA-4). Además, todos los ratones se trataron con mAb anti-PD-1 comenzando el día 21. Las respuestas de linfocitos T se midieron mediante de ELISPOT de IFN-gamma. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 14 días después de la inmunización con adenovirus.

[00151] **La FIG. 27B** ilustra respuestas de linfocitos T específicos de antígeno después de la sensibilización/refuerzo heterólogos en ratones con tumores B16-OVA. Los ratones C57BL/6J con tumores B16-OVA se inyectaron con adenovirus que expresa GFP (Ad5-GFP) y se reforzaron con srRNA de VEE-Luciferasa formulado con MC3 LNP (Control) o Ad5-UbAAY y reforzados con srRNA de VEE-UbAAY (Vax). Tanto el grupo de control como el Vax también se trataron con un mAb de control de IgG. Un tercer grupo se trató con la sensibilización de Ad5-GFP/refuerzo de srRNA de VEE-Luciferasa en combinación con anti-CTLA-4 (aCTLA-4), mientras que el cuarto grupo se trató con la sensibilización de Ad5-UbAAY/refuerzo de VEE-UbAAY en combinación con anti-CTLA-4 (Vax + aCTLA-4). Además, todos los ratones se trataron con mAb anti-PD-1 comenzando el día 21. Las respuestas de linfocitos T se midieron mediante de ELISPOT de IFN-gamma. Los ratones se sacrificaron y se recolectaron los bazos

y ganglios linfáticos 14 días después de la inmunización con adenovirus y 14 días después del refuerzo con srRNA (día 28 después de la sensibilización).

[00152] **La FIG. 27C** ilustra respuestas de linfocitos T específicos de antígeno después de la sensibilización/refuerzo heterólogos en ratones con tumores B16-OVA. Los ratones C57BL/6J con tumores B16-OVA se inyectaron con adenovirus que expresa GFP (Ad5-GFP) y se reforzaron con srRNA de VEE-Luciferasa formulado con MC3 LNP (Control) o Ad5-UbAAY y reforzados con srRNA de VEE-UbAAY (Vax). Tanto el grupo de control como el Vax también se trataron con un mAb de control de IgG. Un tercer grupo se trató con la sensibilización de Ad5-GFP/refuerzo de srRNA de VEE-Luciferasa en combinación con anti-CTLA-4 (aCTLA-4), mientras que el cuarto grupo se trató con la sensibilización de Ad5-UbAAY/refuerzo de VEE-UbAAY en combinación con anti-CTLA-4 (Vax + aCTLA-4). Además, todos los ratones se trataron con mAb anti-PD-1 comenzando el día 21. Las respuestas de linfocitos T se midieron mediante tinción con pentámeros del MHC clase I. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 14 días después de la inmunización con adenovirus.

[00153] **La FIG. 27D** ilustra respuestas de linfocitos T específicos de antígeno después de la sensibilización/refuerzo heterólogos en ratones con tumores B16-OVA. Los ratones C57BL/6J con tumores B16-OVA se inyectaron con adenovirus que expresa GFP (Ad5-GFP) y se reforzaron con srRNA de VEE-Luciferasa formulado con MC3 LNP (Control) o Ad5-UbAAY y reforzados con srRNA de VEE-UbAAY (Vax). Tanto el grupo de control como el Vax también se trataron con un mAb de control de IgG. Un tercer grupo se trató con la sensibilización de Ad5-GFP/refuerzo de srRNA de VEE-Luciferasa en combinación con anti-CTLA-4 (aCTLA-4), mientras que el cuarto grupo se trató con la sensibilización de Ad5-UbAAY/refuerzo de VEE-UbAAY en combinación con anti-CTLA-4 (Vax + aCTLA-4). Además, todos los ratones se trataron con mAb anti-PD-1 comenzando el día 21. Las respuestas de linfocitos T se midieron mediante tinción con pentámeros del MHC clase I. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 14 días después de la inmunización con adenovirus y 14 días después del refuerzo con srRNA (día 28 después de la sensibilización).

[00154] **La FIG. 28A** ilustra respuestas de linfocitos T específicos de antígeno después de la sensibilización/refuerzo heterólogos en ratones con tumores CT26 (BALB/c). Los ratones se inmunizaron con Ad5-GFP y se reforzaron 15 días después de la sensibilización con adenovirus con srRNA de VEE-Luciferasa formulado con MC3 LNP (Control) o sensibilizados con Ad5-UbAAY y reforzados con srRNA de VEE-UbAAY (Vax). Tanto el grupo de control como el Vax también se trataron con un mAb de control de IgG. A un grupo separado se le administró la sensibilización/refuerzo de Ad5-GFP/srRNA de VEE-Luciferasa en combinación con anti-PD-1 (aPD1), mientras que un cuarto grupo recibió la sensibilización/refuerzo de Ad5-UbAAY/srRNA de VEE-UbAAY en combinación con un mAb anti-PD-1 (Vax + aPD1). Las respuestas de linfocitos T al péptido AH1 se midieron mediante el uso de ELISPOT

de IFN-gamma. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 12 días después de la inmunización con adenovirus.

[00155] **La FIG. 28B** ilustra respuestas de linfocitos T específicos de antígeno después de la sensibilización/refuerzo heterólogos en ratones con tumores CT26 (BALB/c). Los ratones se inmunizaron con Ad5-GFP y se reforzaron 15 días después de la sensibilización con adenovirus con srRNA de VEE-Luciferasa formulado con MC3 LNP (Control) o sensibilizados con Ad5-UbAAY y reforzados con srRNA de VEE-UbAAY (Vax). Tanto el grupo de control como el Vax también se trataron con un mAb de control de IgG. A un grupo separado se le administró la sensibilización/refuerzo de Ad5-GFP/srRNA de VEE-Luciferasa en combinación con anti-PD-1 (aPD1), mientras que un cuarto grupo recibió la sensibilización/refuerzo de Ad5-UbAAY/srRNA de VEE-UbAAY en combinación con mAb anti-PD-1 (Vax + aPD1). Las respuestas de linfocitos T al péptido AH1 se midieron mediante el uso de ELISPOT de IFN-gamma. Los ratones se sacrificaron y se recolectaron los bazos y ganglios linfáticos 12 días después de la inmunización con adenovirus y 6 días después del refuerzo con srRNA (día 21 después de la sensibilización).

[00156] **La FIG. 29** ilustra ChAdV68 que provoca respuestas de linfocitos T a antígenos tumorales de ratón en ratones. Los ratones se inmunizaron con ChAdV68.5WTnt.MAG25mer, y las respuestas de linfocitos T al epítopo del MHC clase I SIINFEKL (OVA) (SEQ ID NO: 57) se midieron en ratones hembra C57BL/6J y al epítopo del MHC clase I AH1-A5 se midieron en ratones Balb/c. Media presentada de células formadoras de manchas (SFC) por 10^6 esplenocitos medidas en ensayos ELISpot. Las barras de error representan la desviación estándar.

[00157] **La FIG. 30** ilustra respuestas inmunitarias celulares en un modelo tumoral CT26 después de una única inmunización con ChAdV6, ChAdV + anti-PD-1, srRNA, srRNA + anti-PD-1 o anti-PD-1 solo. La producción de IFN-gamma específico del antígeno se midió en esplenocitos para 6 ratones de cada grupo mediante el uso de ELISpot. Los resultados se presentan como células formadoras de manchas (SFC) por10^6 esplenocitos. Mediana para cada grupo indicada mediante línea horizontal. Valores P determinados mediante el uso de prueba de comparación múltiple de Dunnett; *** P<0,0001, **P<0,001, *P<0,05. ChAdV= ChAdV68.5WTnt.MAG25mer; srRNA= srRNA de VEE-MAG25mer.

[00158] **La FIG. 31** ilustra respuestas de linfocitos T CD8 en un modelo tumoral CT26 después de una única inmunización con ChAdV6, ChAdV + anti-PD-1, srRNA, srRNA + anti-PD-1 o anti-PD-1 solo. Producción de IFN-gamma específico de antígeno en linfocitos T CD8 medida mediante el uso de ICS y resultados presentados como linfocitos T CD8 específicos de antígeno como un porcentaje de los linfocitos T CD8 totales. Mediana para cada grupo indicada mediante línea horizontal. Valores P determinados mediante el uso de prueba de comparación múltiple de Dunnett; *** P<0,0001, **P<0,001, *P<0,05. ChAdV= ChAdV68.5WTnt.MAG25mer; srRNA= srRNA de VEE-MAG25mer.

[00159] **La FIG. 32** ilustra crecimiento tumoral en un modelo tumoral CT26 después de la inmunización con sensibilización/refuerzo heterólogos ChAdV/srRNA, sensibilización/refuerzo heterólogos srRNA/ChAdV o sensibilización/refuerzo homólogos srRNA/srRNA. También se ilustra en una comparación de las inmunizaciones de sensibilización/refuerzo con o sin la administración de anti-PD1 durante la sensibilización y refuerzo. Volúmenes tumorales medidos dos veces por semana y media de volúmenes tumorales presentados para los primeros 21 días del estudio. 22-28 ratones por grupo al inicio del estudio. Las barras de error representan el error estándar de la media (SEM). Valores P determinados mediante el uso de prueba de Dunnett; *** P<0,0001, **P<0,001, *P<0,05. ChAdV= ChAdV68.5WTnt.MAG25mer; srRNA= srRNA de VEE-MAG25mer.

[00160] **La FIG. 33** ilustra la supervivencia en un modelo tumoral CT26 después de la inmunización con sensibilización/refuerzo heterólogos ChAdV/srRNA, sensibilización/refuerzo heterólogos srRNA/ChAdV o sensibilización/refuerzo homólogos srRNA/srRNA. También se ilustra en una comparación de las inmunizaciones de sensibilización/refuerzo con o sin la administración de anti-PD1 durante la sensibilización y refuerzo. Valores P determinados mediante el uso de prueba de rango logarítmico; *** P<0,0001, **P<0,001, *P<0,01. ChAdV= ChAdV68.5WTnt.MAG25mer; srRNA= srRNA de VEE-MAG25mer.

[00161] **La FIG. 34** ilustra las respuestas inmunitarias celulares específicas del antígeno medidas mediante el uso de ELISpot. Se midió la producción de IFN-gamma específico del antígeno para seis epítopos restringidos mamu A01 diferentes en PBMC para el grupo de sensibilización/refuerzo homólogos srRNA de VEE-MAG25mer-LNP1(30 µg) (FIG. 34A), srRNA de VEE-MAG25mer-LNP1(100 µg) (FIG. 34B) o srRNA de VEE-MAG25mer-LNP2(100 µg) (FIG. 34C) o sensibilización/refuerzo heterólogos ChAdV68.5WTnt.MAG25mer /srRNA de VEE-MAG25mer (FIG. 34D) mediante el uso de ELISpot 1, 2, 3, 4, 5, 6, 8, 9 o 10 semanas después de la primera inmunización de refuerzo (6 macacos rhesus por grupo). Los resultados se presentan como media de células formadoras de manchas (SFC) por 10^6 PBMC para cada epítopo en un formato de gráfico de barras apiladas. Los valores para cada animal se normalizaron a los nivels antes de la extracción de sangre (semana 0).

[00162] **La FIG. 35** muestra la respuesta inmunitaria celular específica del antígeno medida mediante el uso de ELISpot. La producción de IFN-gamma específico del antígeno para seis epítopos restringidos mamu A01 diferentes se midió en PBMC después de la inmunización con un régimen de sensibilización/refuerzo homólogos ChAdV68.5WTnt.MAG25mer /VEE-srRNA MAG25mer mediante el uso de ELISpot antes de la inmunización y 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 o 24 semanas después de la inmunización inicial. Los resultados se presentan como media de células formadoras de manchas (SFC) por 10^6 PBMC para cada epítopo (6 macacos rhesus por grupo) en un formato de gráfico de barras apiladas.

[00163] **FIG. 36** muestra la respuesta inmunitaria celular específica del antígeno medida mediante el uso de ELISpot. La producción de IFN-gamma específico del antígeno para seis epítopos restringidos mamu A01 diferentes se midió en PBMC después de la inmunización con un régimen de sensibilización/refuerzo homólogos LNP2 srRNA de VEE-MAG25mer mediante el uso de ELISpot antes de la inmunización y 4, 5, 6, 7, 8, 10, 11, 12, 13, 14 o 15 semanas después de la inmunización inicial. Los resultados se presentan como media de células formadoras de manchas (SFC) por 10^6 PBMC para cada epítopo (6 macacos rhesus por grupo) en un formato de gráfico de barras apiladas.

[00164] **FIG. 37** muestra la respuesta inmunitaria celular específica del antígeno medida mediante el uso de ELISpot. La producción de IFN-gamma específico del antígeno para seis epítopos restringidos mamu A01 diferentes se midió en PBMC después de la inmunización con un régimen de cebado/refuerzo homólogo LNP1 srRNA de VEE-MAG25mer mediante el uso de ELISpot antes de la inmunización y 4, 5, 6, 7, 8, 10, 11, 12, 13, 14 o 15 semanas después de la inmunización inicial. Los resultados se presentan como media de células formadoras de manchas (SFC) por 10^6 PBMC para cada epítopo (6 macacos rhesus por grupo) en un formato de gráfico de barras apiladas.

[00165] **FIG. 38** ilustra la determinación de métricas de distancia para dos secuencias de casete ilustrativa. La figura describe la SEQ ID NOS 191 y 192, respectivamente, en orden de aparición.

DESCRIPCIÓN DETALLADA

I. Definiciones

[00166] En general, se pretende que los términos usados en las reivindicaciones y la memoria descriptiva se interpreten con su significado común comprendido por un experto en la técnica. A continuación, se definen determinados términos para proporcionar mayor claridad. En caso de contradicción entre el significado común y las definiciones proporcionadas, se usarán las definiciones proporcionadas.

[00167] Como se usa en la presente, el término «antígeno» es una sustancia que induce una respuesta inmunitaria.

[00168] Como se usa en la presente, el término «neoantígeno» es un antígeno que tiene al menos una alteración que lo distingue del antígeno de tipo salvaje correspondiente, p. ej., mediante la mutación en una célula tumoral o modificación postraduccional específica de una célula tumoral. Un neoantígeno puede incluir una secuencia de polipéptidos o una secuencia de nucleótidos. Una mutación puede incluir un indel de desplazamiento de marco o sin desplazamiento de marco, una sustitución de cambio de sentido o sin sentido, una alteración en el sitio de corte y empalme, una reorganización genómica o fusión genética, o cualquier alteración genómica o de expresión que conduzca a un neoORF. Una mutación también puede incluir una variante de corte y empalme. Las modificaciones postraduccionales específicas de una célula tumoral pueden incluir fosforilación anómala. Las modificaciones

postraduccionales específicas de una célula tumoral pueden incluir también un antígeno sometido a corte y empalme generado por el proteasoma. Remitirse a Liepe et ál., Una gran proporción de ligandos HLA clase I son péptidos sometidos a corte y empalme generado por el proteasoma; Science. 2016 Oct 21;354(6310):354-358.

[00169] Como se usa en la presente, la expresión «neoantígeno del tumor» es un neoantígeno presente en un tejido o célula tumoral de un sujeto, pero no en el tejido o célula normal correspondiente del sujeto.

[00170] Como se usa en la presente, la expresión «vacuna basada en neoantígenos» es una construcción de vacuna basada en uno o más neoantígenos, p. ej., múltiples neoantígenos.

[00171] Como se usa en la presente, la expresión «neoantígeno candidato» es una mutación u otra anomalía que da lugar a una nueva secuencia que puede representar un neoantígeno.

[00172] Como se usa en la presente, la expresión «región codificante» es la parte o partes de un gen que codifican proteína.

[00173] Como se usa en la presente, la expresión «mutación codificante» es una mutación que se produce en una región codificante.

[00174] Como se usa en la presente, el término «ORF» significa marco de lectura abierto.

[00175] Como se usa en la presente, el término «NEO-ORF» es un ORF específico de tumor que surge de una mutación u otra anomalía tal como el corte y empalme.

[00176] Como se usa en la presente, la expresión «mutación de cambio de sentido» es una mutación que produce una sustitución de un aminoácido a otro.

[00177] Como se usa en la presente, la expresión «mutación sin sentido» es una mutación que produce una sustitución de un aminoácido a un codón de terminación o que provoca la eliminación de un codón de inicio canónico.

[00178] Como se usa en la presente, la expresión «mutación de desplazamiento de marco» es una mutación que produce un cambio en el marco de la proteína.

[00179] Como se usa en la presente, el término «indel» es una inserción o eliminación de uno o más un ácido nucleico.

[00180] Como se usa en la presente, el término «identidad» en el contexto de dos o más secuencias de ácidos nucleicos o polipéptidos, hace referencia a dos o más secuencias o subsecuencias que tienen un porcentaje especificado de nucleótidos o residuos de aminoácidos que son los mismos, cuando se comparan y se alinean para una máxima correspondencia, según se mide usando uno de los algoritmos de comparación de secuencias descritos a continuación (p. ej., BLASTP y BLASTN u otros algoritmos disponibles para los expertos) o por inspección visual. Dependiendo de la aplicación, el porcentaje

«identidad» puede existir sobre una región de la secuencia que se compara, p. ej., sobre un dominio funcional, o, de manera alternativa, puede existir sobre la longitud completa de las dos secuencias a comparar.

[00181] Para la comparación de secuencias, típicamente una secuencia actúa como una secuencia de referencia con la cual se comparan las secuencias de prueba. Cuando se usa un algoritmo de comparación de secuencias, se ingresan las secuencias de prueba y de referencia en una computadora, se designan las coordenadas de las subsecuencias, si fuera necesario, y se designan los parámetros del programa de algoritmo de secuencias. El algoritmo de comparación de secuencias después calcula el porcentaje de identidad de secuencia para la secuencia o secuencias de prueba con respecto a la secuencia de referencia, en función de los parámetros de programa designados. De manera alternativa, la similitud o disimilitud se puede establecer por la presencia o ausencia combinadas de nucleótidos particulares, o, para secuencias traducidas, aminoácidos en posiciones de secuencia seleccionadas (p. ej., motivos de secuencia).

[00182] La alineación óptima de secuencias para su comparación se puede llevar a cabo, p. ej., mediante el algoritmo de homología local de Smith & Waterman, Adv. Appl. Math. 2:482 (1981), mediante el algoritmo de alineación de homología de Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), mediante el método de búsqueda de similitud de Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), mediante implementaciones computarizadas de estos algoritmos (GAP, BESTFIT, FASTA y TFASTA en el paquete de software de Wisconsin Genetics, Genetics Computer Group, 575 Science Dr., Madison, WI) o mediante inspección visual (remitirse, en líneas generales, a Ausubel et ál., infra).

[00183] Un ejemplo de un algoritmo que es adecuado para determinar el porcentaje de identidad de secuencia y similitud de secuencia es el algoritmo BLAST, que se describe en Altschul et ál., J. Mol. Biol. 215:403-410 (1990). El software para la realización del análisis con BLAST se encuentra disponible públicamente a través del Centro Nacional para la Información Biotecnológica.

[00184] Como se usa en la presente, el término «*non-stop* o translectura» es una mutación que produce la eliminación del codón de terminación natural.

[00185] Como se usa en la presente, el término «epítopo» es la parte específica de un antígeno a la que típicamente se une un anticuerpo o receptor de linfocitos T.

[00186] Como se usa en la presente, el término «inmunogénico» es la capacidad de provocar una respuesta inmunitaria, p. ej., mediante linfocitos T, linfocitos B o ambos.

[00187] Como se usa en la presente, la expresión «afinidad de unión a HLA», «afinidad de unión a MHC» significa afinidad de unión entre un antígeno específico y un alelo del MHC específico.

[00188] Como se usa en la presente, el término «cebo» es una sonda de ácido nucleico que se usa para enriquecer una secuencia específica de ADN o ARN de una muestra.

[00189] Como se usa en la presente, el término «variante» es una diferencia entre los ácidos nucleicos de un sujeto y el genoma humano de referencia usado como control.

[00190] Como se usa en la presente, la expresión «tipificación de variante» es una determinación algorítmica de la presencia de una variante, típicamente a partir de la secuenciación.

[00191] Como se usa en la presente, el término «polimorfismo» es una variante de línea germinal, es decir, una variante que se encuentra en todas las células portadoras de ADN de un individuo.

[00192] Como se usa en la presente, el término «variante somática» es una variante que surge en células que no son de línea germinal de un individuo.

[00193] Como se usa en la presente, el término «alelo» es una versión de un gen o una versión de una secuencia genética o una versión de una proteína.

[00194] Como se usa en la presente, la expresión «tipo HLA» es el complemento de los alelos del gen HLA.

[00195] Como se usa en la presente, «degradación mediada por mutaciones terminadoras» o «NMD (por sus siglas en inglés)» es una degradación de un ARNm por una célula debido a un codón de terminación prematuro.

[00196] Como se usa en la presente, la expresión «mutación troncal» es una mutación que se origina temprano en el desarrollo de un tumor y presente en una parte sustancial de las células del tumor.

[00197] Como se usa en la presente, la expresión «mutación subclonal» es una mutación que se origina más tarde en el desarrollo de un tumor y presente únicamente en un subconjunto de las células del tumor.

[00198] Como se usa en la presente, el término «exoma» es un subconjunto del genoma que codifica proteínas. Un exoma puede ser el conjunto de exones de un genoma.

[00199] Como se usa en la presente, la expresión «regresión logística» es un modelo de regresión para datos binarios de estadísticas en el que el logit de la probabilidad de que la variable dependiente sea igual a uno se modela como una función lineal de las variables dependientes.

[00200] Como se usa en la presente, la expresión «red neuronal» es un modelo de aprendizaje automático para clasificación o regresión que consiste en múltiples capas de transformaciones lineales seguidas de no linealidades en cuanto a elementos típicamente con entrenamiento mediante descenso de gradiente estocástico y retropropagación.

[00201] Como se usa en la presente el término «proteoma» es el conjunto de todas las proteínas expresadas y/o traducidas por una célula, grupo de células o individuo.

[00202] Como se usa en la presente, el término «peptidoma» es el conjunto de todos los péptidos presentados por MHC-I o MHC-II en la superficie de la célula. El peptidoma puede hacer referencia a

una propiedad de una célula o un conjunto de células (p. ej., el peptidoma del tumor, es decir la unión de los peptidomas de todas las células que constituyen el tumor).

[00203] Como se usa en la presente, el término «ELISPOT» significa ensayo de inmunospot ligado a enzimas, que es un método común para monitorear respuestas inmunitarias en humanos y animales.

[00204] Como se usa en la presente, el término «dextrámeros» son multímeros MHC-péptido basados en dextrano usados para tinción de linfocitos T específicos de antígeno en citometría de flujo.

[00205] Como se usa en la presente, la expresión «tolerancia o tolerancia inmunitaria» es un estado de ausencia de respuesta inmunitaria a uno o más antígenos, p. ej., autoantígenos.

[00206] Como se usa en la presente, la expresión «tolerancia central» es una tolerancia afectada en el timo, ya sea por eliminación de clones de linfocitos T autorreactivos o por promoción de clones de linfocitos T autorreactivos para que se diferencien en linfocitos T reguladores inmunosupresores (Tregs).

[00207] Como se usa en la presente, la expresión «tolerancia periférica» es una tolerancia afectada en la periferia mediante la regulación por disminución o anergización de linfocitos T autorreactivos que sobreviven a la tolerancia central o promoción de estos linfocitos T para que se diferencien en Tregs.

[00208] El término «muestra» puede incluir una única célula o múltiples células o fragmentos de células o una alícuota de fluido corporal, tomados de un sujeto, por medios que incluyen venopunción, excreción, eyaculación, masaje, biopsia, aspiración con aguja, muestras por lavado, raspado, incisión quirúrgica o intervención u otros medios conocidos en la técnica.

[00209] El término «sujeto» comprende una célula, tejido u organismo, humano o no humano, ya sea in vivo, ex vivo o in vitro, masculino o femenino. El término «sujeto» incluye mamíferos, que incluyen humanos.

[00210] El término «mamífero» abarca tanto humanos como no humanos e incluye, de modo no taxativo, humanos, primates no humanos, caninos, felinos, murinos, bovinos, equinos, y porcinos.

[00211] La expresión «factor clínico» hace referencia a una medida de una afección de un sujeto, p. ej., actividad o gravedad de la enfermedad. «Factor clínico» comprende todos los marcadores del estado de salud de un sujeto, que incluye marcadores que no son de la muestra y/u otras características de un sujeto, tales como, de modo no taxativo, edad y sexo. Un factor clínico puede ser una puntuación, un valor o un conjunto de valores que pueden obtenerse a partir de la evaluación de una muestra (o población de muestras) de un sujeto o un sujeto con una afección determinada. Un factor clínico también puede predecirse mediante marcadores y/u otros parámetros tales como sustitutos de expresión génica. Los factores clínicos pueden incluir tipo de tumor, subtipo de tumor e historial de tabaquismo.

[00212] La expresión «secuencias de ácidos nucleicos que codifican antígenos derivadas de un tumor» hace referencia a secuencias de ácidos nucleicos directamente extraídas del tumor, p. ej., mediante RT-PCR; o datos de secuencia obtenidos mediante la secuenciación del tumor y la posterior

síntesis de las secuencias de ácidos nucleicos mediante el uso de datos de secuenciación, p. ej., mediante diversos métodos sintéticos o basados en PCR conocidos en la técnica.

[00213] El término «alfavirus» hace referencia a miembros de la familia *Togaviridae*, y son virus de ARN monocatenario de sentido positivo. Los alfavirus se clasifican típicamente como del viejo mundo, tales como virus Sindbis, río Ross, Mayaro, Chikungunya y bosque Semliki, o como del nuevo mundo, tales como encefalitis equina del este, Aura, Fort Morgan, encefalitis equina venezolana y su cepa derivada TC-83. Los alfavirus son típicamente virus de ARN autorreplicante.

[00214] La expresión «estructura principal de alfavirus» hace referencia a secuencia/s mínima/s de un alfavirus que permiten la autorreplicación del genoma viral. Las secuencias mínimas pueden incluir secuencias conservadas para amplificación mediada por proteína no estructural, un gen de proteína no estructural 1 (nsP1), un gen nsP2, un gen nsP3, un gen nsP4 y una secuencia de poliA, así como secuencias para expresión de ARN viral subgenómico que incluye un elemento promotor 26S.

[00215] La expresión «secuencias para amplificación mediada por proteína no estructural» incluye elementos de secuencia conservados (CSE, por sus siglas en inglés) de alfavirus que son conocidos por los expertos en la técnica. Los CSE incluyen, de modo no taxativo, una UTR 5’ de alfavirus, un CSE 51-nt, un CSE 24-nt u otra secuencia promotora subgenómica 26S, un CSE 19-nt y una UTR 3’ de alfavirus.

[00216] La expresión «ARN polimerasa» incluye polimerasas que catalizan la producción de polinucleótidos de ARN a partir de una plantilla de ADN. Las ARN polimerasas incluyen, de modo no taxativo, polimerasas derivadas de bacteriófagos que incluyen T3, T7 y SP6.

[00217] El término «lípido» incluye moléculas hidrofóbicas y/o anfifílicas. Los lípidos pueden ser catiónicos, aniónicos o neutros. Los lípidos pueden ser de origen natural o sintético, y en algunos casos biodegradables. Los lípidos pueden incluir colesterol, fosfolípidos, conjugados lipídicos que incluyen, de modo no taxativo, conjugados de polietilenglicol (PEG) (lípidos PEGilados), ceras, aceites, glicéridos, grasas y vitaminas solubles en grasas. Los lípidos pueden incluir también dilinoleilmetil-4-dimetilaminobutirato (MC3) y moléculas similares a MC3.

[00218] La expresión «nanopartícula lipídica» o «LNP» incluye estructuras tipo vesícula formadas mediante una membrana que contiene lípidos que rodea un interior acuoso, a las que también se hace referencia como liposomas. Las nanopartículas lipídicas incluyen composiciones a base de lípidos con un núcleo lipídico sólido estabilizado mediante un tensioactivo. Los núcleos lipídicos pueden ser ácidos grasos, acilgliceroles, ceras y mezcla de estos tensioactivos. Los lípidos de la membrana biológica tales como fosfolípidos, esfingomielina, sales biliares (taurocolato de sodio), y esteroles (colesterol) se pueden usar como estabilizadores. Las nanopartículas lipídicas se pueden formar mediante el uso de relaciones definidas de diferentes moléculas lipídicas que incluyen, de modo no taxativo, relaciones definidas de uno o más lípidos catiónicos, aniónicos o neutros. Las nanopartículas lipídicas pueden

encapsular moléculas dentro de una cubierta de membrana exterior y posteriormente pueden ponerse en contacto con células diana para suministrar las moléculas encapsuladas al citosol de la célula hospedadora. Las nanopartículas lipídicas pueden modificarse o funcionalizarse con moléculas no lipídicas, incluso en su superficie. Las nanopartículas lipídicas pueden ser de capa única (unilamelares) o de múltiples capas (multilamelares). Las nanopartículas lipídicas pueden formar complejos con ácido nucleico. Las nanopartículas lipídicas unilamelares pueden formar complejos con ácido nucleico, donde el ácido nucleico está en el interior acuoso. Las nanopartículas lipídicas multilamelares pueden formar complejos con ácido nucleico, donde el ácido nucleico está en el interior acuoso, o forma o está intercalado.

[00219] Abreviaturas: MHC: complejo mayor de histocompatibilidad; HLA: antígeno leucocitario humano o locus de gen MHC humano; NGS: secuenciación de nueva generación; PPV: valor predictivo positivo; TSNA: neoantígenos específicos de tumor; FFPE: fijado con formalina, embebido en parafina; NMD: degradación mediada por mutaciones terminadoras; NSCLC: cáncer pulmonar no microcítico; DC: célula dendrítica.

[00220] Cabe destacar que, tal como se usan en la memoria descriptiva y las reivindicaciones anexas, las formas singulares «un», «una», «el» y «la» incluyen las referencias al plural a menos que el contexto indique claramente lo contrario.

[00221] A menos que se indique específicamente o sea obvio a partir del contexto, tal como se usa en la presente, se entiende que la expresión “alrededor de” está dentro de un intervalo de tolerancia normal en la técnica, por ejemplo, dentro de 2 desviaciones estándar del promedio. Alrededor de puede entenderse como dentro de 10 %, 9 %, 8 %, 7 %, 6 %, 5 %, 4 %, 3 %, 2 %, 1 %, 0,5 %, 0,1 %, 0,05 % o 0,01 % del valor indicado. A menos que quede claro a partir del contexto, todos los valores numéricos proporcionados en la presente están modificados por la expresión “alrededor de”.

[00222] Se entenderá que cualesquiera términos no definidos directamente en la presente tendrá los significados que habitualmente se asocian con estos como se entiende en la técnica de la invención. Ciertos términos se describen en la presente para brindarle una guía adicional al practicante al describir las composiciones, dispositivos, métodos y similares de aspectos de la invención y cómo fabricarlos o usarlos. Se comprenderá que es posible decir lo mismo de más de una forma. Por consiguiente, puede usarse lenguaje alternativo y sinónimos para cualesquiera uno o más de los términos descritos en la presente. No debe darse importancia al hecho de que un término se detalle o describa en la presente o no. Se proporcionan algunos sinónimos o métodos, materiales y similares que se pueden sustituir. El hecho de que se mencione uno o algunos sinónimos o equivalentes no excluye el uso de otros sinónimos o equivalentes, a menos que se establezca explícitamente. El uso de ejemplos, que incluyen ejemplos de los términos, es a efectos únicamente ilustrativos y no se pretende que limite el alcance y el

significado de los aspectos de la presente invención.

[00223] Todas las referencias, patentes emitidas y solicitudes de patente citadas dentro del cuerpo de la memoria descriptiva se incorporan en la presente en su totalidad mediante esta referencia a todos los efectos.

II. Métodos para indentificar neoantígenos

[00224] En la presente se describen métodos para identificar neoantígenos de un tumor de un sujeto que probablemente se presenten en la superficie celular del tumor o células inmunitarias que incluyen células que presentan antígenos profesionales, tales como células dendríticas, y/o probablemente sean inmunógenos. A modo de ejemplo, uno de estos métodos puede comprender las etapas de: obtener los datos de secuenciación de nucleótidos tumorales de al menos uno de exoma, transcriptoma o genoma entero de la célula tumoral del sujeto, donde los datos de secuenciación de nucleótidos tumorales se usan para obtener datos que representan las secuencias de péptidos de cada uno de un conjunto de neoantígenos y donde la secuencia de péptidos de cada neoantígeno comprende al menos una alteración que la diferencia de la secuencia de péptidos de tipo salvaje correspondiente; ingresar la secuencia de péptidos de cada neoantígeno en uno o más modelos de presentación para generar un conjunto de probabilidades numéricas de que cada uno de los neoantígenos se presente por uno o más alelos de MHC en la superficie de células tumorales de la célula tumoral del sujeto o células presentes en el tumor, el conjunto de probabilidades numéricas se identificó al menos en función de los datos de espectrometría de masas recibidos y seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades numéricas de generar un conjunto de neoantígenos seleccionados.

[00225] El modelo de presentación puede comprender un modelo de regresión estadística o aprendizaje automático (p. ej., aprendizaje profundo) entrenado en un conjunto de datos de referencia (también denominado conjunto de datos de entrenamiento) que comprende un conjunto de etiquetas correspondientes, donde el conjunto de datos de referencia se obtiene de cada uno de múltiples sujetos distintos donde opcionalmente algunos sujetos pueden tener un tumor y donde el conjunto de datos de referencia comprende al menos uno de: datos que representan secuencias de nucleótidos de exoma de tejido tumoral, datos que representan secuencias de nucleótidos de exoma de tejido normal, datos que representan secuencias de nucleótidos de transcriptoma de tejido tumoral, datos que representan secuencias de proteoma de tejido tumoral y datos que representan secuencias de peptidoma de MHC de tejido tumoral y datos que representan secuencias de peptidoma de MHC de tejido normal. Los datos de referencia también pueden comprender datos de espectrometría de masas, datos de secuenciación, datos de secuenciación de ARN y datos de proteómica para líneas celulares de alelos simples modificadas genéticamente para expresar un alelo de MHC predeterminado que posteriormente se

exponen a proteína sintética, líneas celulares humanas normales y tumorales y muestras primarias frescas y congeladas y ensayos de linfocitos T (p. ej., ELISPOT). En determinados aspectos, el conjunto de datos de referencia incluye cada forma de datos de referencia.

[00226] El modelo de presentación puede comprender un conjunto de rasgos que proviene al menos en parte del conjunto de datos de referencia y donde el conjunto de rasgos comprende al menos uno de rasgos dependientes del alelo y rasgos independientes del alelo. En determinados aspectos, se incluye cada rasgo.

[00227] También se describen en la presente métodos para generar una salida para la construcción de una vacuna contra el cáncer personalizada mediante la identificación de uno o más neoantígenos de una o más células tumorales de un sujeto que es probable que se presenten en una superficie de las células tumorales. A modo de ejemplo, uno de estos métodos puede comprender las etapas de: obtener datos de secuenciación de nucleótidos de al menos uno de exoma, transcriptoma o genoma entero de las células tumorales o y células normales del sujeto, donde los datos de secuenciación de nucleótidos se usan para obtener datos que representan secuencias de péptidos de cada uno de un conjunto de neoantígenos identificados mediante la comparación de los datos de secuenciación de nucleótidos de las células tumorales y los datos de secuenciación de nucleótidos de las células normales, y donde la secuencia de péptidos de cada neoantígeno comprende al menos una alteración que la diferencia de la secuencia de péptidos de tipo salvaje correspondiente identificada de las células normales del sujeto; codificar las secuencias de péptidos de cada uno de los neoantígenos en un vector numérico correspondiente, donde cada vector numérico incluye información con respecto a múltiples aminoácidos que componen la secuencia de péptidos y un conjunto de posiciones de los aminoácidos de la secuencia de péptidos; ingresar los vectores numéricos, mediante el uso de un procesador informático, en un modelo de presentación de aprendizaje profundo, para generar un conjunto de probabilidades de presentación para el conjunto de neoantígenos, donde cada probabilidad de presentación del conjunto representa la probabilidad de que un neoantígeno correspondiente sea presentado por uno o más alelos del MHC clase II en la superficie de las células tumorales del sujeto, el modelo de presentación de aprendizaje profundo; seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades de presentación para generar un conjunto de neoantígenos seleccionados; y generar la salida para construir la vacuna contra el cáncer personalizada en función del conjunto de neoantígenos seleccionados.

[00228] En algunas modalidades, el modelo de presentación comprende múltiples parámetros identificados al menos en función de un conjunto de datos de entrenamiento y una función que representa una relación entre el vector numérico recibido como entrada y la probabilidad de presentación generada como salida en función del vector numérico y los parámetros. En determinadas modalidades,

el conjunto de datos de entrenamiento comprende etiquetas obtenidas mediante espectrometría de masas que miden la presencia de péptidos unidos a al menos un alelo del MHC clase II identificado como presente en al menos una de múltiples muestras, entrenar secuencias de péptidos codificadas como vectores numéricos que incluyen información con respecto a múltiples aminoácidos que conforman la secuencia de péptidos y un conjunto de posiciones de los aminoácidos de la secuencia de péptidos, y al menos un alelo HLA asociado con las secuencias de péptidos de entrenamiento.

[00229] La presentación de células dendríticas a rasgos de linfocitos T sin tratamiento puede comprender al menos uno de: Un rasgo descrito anteriormente. La dosis y el tipo de antígeno en la vacuna. (p. ej., péptido, ARNm, virus, etc.): (1) La vía por la cual las células dendríticas (DC) absorben el tipo de antígeno (p. ej., endocitosis, micropinocitosis) y/o (2) La eficacia con la cual las DC absorben el antígeno. La dosis y el tipo de adyuvante en la vacuna. La longitud de la secuencia de antígeno de la vacuna. La cantidad y los lugares de administración de la vacuna. Funcionamiento inmunitario basal del paciente (p. ej., como se mide por el historial de infecciones recientes, conteos sanguíneos, etc.) Para vacunas de ARN: (1) el índice de renovación del producto de proteína de ARNm en la célula dendrítica; (2) la velocidad de traducción del ARNm luego de la absorción por las células dendríticas como se mide en experimentos in vitro o in vivo y/o (3) la cantidad o rondas de traducción del ARNm después de la absorción por células dendríticas como se mide en experimentos in vitro o in vivo. La presencia de motivos de escisión de proteasa en el péptido, que opcionalmente brinda peso adicional a las proteasas típicamente expresadas en células dendríticas (como se mide por secuenciación de ARN o espectrometría de masas). El nivel de expresión de la proteasoma e inmunoproteasoma en células dendríticas activadas típicas (que pueden medirse por secuenciación de ARN, espectrometría de masas, inmunohistoquímica u otras técnicas estándar). Los niveles de expresión del alelo de MHC particular en el individuo en cuestión (p. ej., como se mide por secuenciación de ARN o espectrometría de masas), de manera opcional medidos específicamente en células dendríticas activadas u otras células inmunitarias. La probabilidad de la presentación de péptidos por el alelo de MHC particular en otros individuos que expresan el alelo de MHC particular, de manera opcional medida específicamente en células dendríticas activadas u otras células inmunitarias. La probabilidad de la presentación de péptidos por alelos de MHC en la misma familia de moléculas (p. ej., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) en otros individuos, de manera opcional medida específicamente en células dendríticas activadas u otras células inmunitarias.

[00230] Los rasgos de escape de tolerancia inmunitaria pueden comprender al menos uno de: Medición directa del autopeptidoma mediante espectrometría de masas de proteína realizada en uno o varios tipos celulares. Estimación del autopeptidoma tomando la unión de todas las subcadenas de k-mer (p. ej., 5-25) de autoproteínas. Estimación del autopeptidoma usando un modelo de presentación

similar al modelo de presentación descrito anteriormente aplicado a todas las autoproteínas que no son de mutación, que opcionalmente representan las variantes de línea germinal.

[00231] La clasificación se puede realizar usando múltiples neoantígenos proporcionados por al menos un modelo en función al menos en parte de las probabilidades numéricas. Siguiendo la clasificación, puede realizarse una selección para seleccionar un subconjunto de los neoantígenos clasificados de acuerdo con un criterio de selección. Luego de la selección, un subconjunto de los péptidos clasificados puede proporcionarse como salida.

[00232] Una cantidad del conjunto de neoantígenos seleccionados puede ser 20.

[00233] El modelo de presentación puede representar dependencia entre la presencia de un par de un alelo de MHC particular y un aminoácido particular en una posición particular de una secuencia de péptidos y la probabilidad de presentación en la superficie de células tumorales, por el alelo de MHC particular del par, de tal secuencia de péptidos que comprende el aminoácido particular en la posición particular.

[00234] Un método descrito en la presente también puede incluir aplicar uno o más modelos de presentación a la secuencia de péptidos del neoantígeno correspondiente para generar una puntuación de dependencia para cada uno de uno o más alelos de MHC que indican si el alelo de MHC presentará el neoantígeno correspondiente en función de al menos posiciones de los aminoácidos de la secuencia de péptidos del neoantígeno correspondiente.

[00235] Un método descrito en la presente también puede incluir transformar las puntuaciones de dependencia para generar una probabilidad por alelo correspondiente para cada alelo de MHC que indica una probabilidad de que el alelo de MHC correspondiente presente el neoantígeno correspondiente y combinar las probabilidades por alelo para generar la probabilidad numérica.

[00236] El paso de transformar las puntuaciones de dependencia puede modelar la presentación de la secuencia de péptidos del neoantígeno correspondiente como mutuamente excluyentes.

[00237] Un método descrito en la presente también puede incluir transformar una combinación de las puntuaciones de dependencia para generar la probabilidad numérica.

[00238] El paso de transformar la combinación de las puntuaciones de dependencia puede modelar la presentación de la secuencia de péptidos del neoantígeno correspondiente como interferencia entre los alelos de MHC.

[00239] El conjunto de probabilidades numéricas también puede identificarse por al menos un alelo un rasgo de no interacción con alelos y un método descrito en la presente también puede incluir aplicar uno de no interacción con alelos de uno o más de los modelos de presentación a los rasgos de no interacción con alelos para generar una puntuación de dependencia para los rasgos de no interacción con alelos que indica si la secuencia de péptidos del neoantígeno correspondiente se presentará en

función de los rasgos de no interacción con alelos.

[00240] Un método descrito en la presente también puede incluir combinar la puntuación de dependencia para cada alelo de MHC en uno o más de los alelos de MHC con la puntuación de dependencia del rasgo de no interacción con alelos; transformar las puntuaciones de dependencia combinadas para cada alelo de MHC para generar una probabilidad por alelo correspondiente para el alelo de MHC que indica una probabilidad de que el alelo de MHC correspondiente presente el neoantígeno correspondiente y combinar las probabilidades por alelo para generar la probabilidad numérica.

[00241] Un método descrito en la presente también puede incluir transformar una combinación de las puntuaciones de dependencia para cada uno de los alelos de MHC y la puntuación de dependencia de los rasgos de no interacción con alelos para generar la probabilidad numérica.

[00242] Un conjunto de parámetros numéricos para el modelo de presentación puede entrenarse en función de un conjunto de datos de entrenamiento que incluye al menos un conjunto de secuencias de péptidos de entrenamiento identificadas como presentes en múltiples muestras y uno o más alelos de MHC asociados con cada secuencia de péptidos de entrenamiento, donde las secuencias de péptidos de entrenamiento se identifican mediante espectrometría de masas en péptidos aislados eluidos de alelos de MHC derivados de las múltiples muestras.

[00243] Las muestras también pueden incluir líneas celulares modificadas genéticamente para expresar un único alelo de clase I o clase II de MHC.

[00244] Las muestras también pueden incluir líneas celulares modificadas genéticamente para expresar múltiples alelos de clase I o clase II de MHC.

[00245] Las muestras también pueden incluir líneas celulares humanas obtenidas o derivadas de múltiples pacientes.

[00246] Las muestras también pueden incluir muestras de tumores frescas o congeladas obtenidas de múltiples pacientes.

[00247] Las muestras también pueden incluir muestras de tejido frescas o congeladas obtenidas de múltiples pacientes.

[00248] Las muestras también pueden incluir péptidos identificados usando ensayos de linfocitos T.

[00249] El conjunto de datos de entrenamiento también puede incluir datos asociados con: abundancia de péptidos del conjunto de péptidos de entrenamiento presente en las muestras; longitud de péptidos del conjunto de péptidos de entrenamiento en las muestras.

[00250] El conjunto de datos de entrenamiento puede generarse al comparar el conjunto de secuencias de péptidos de entrenamiento mediante la alineación a una base de datos que comprende

un conjunto de secuencias de proteínas conocidas, donde el conjunto de secuencias de proteínas de entrenamiento es más largo que las secuencias de péptidos de entrenamiento y las incluye.

[00251] El conjunto de datos de entrenamiento puede generarse en función de realizar o haber realizado una secuenciación de nucleótidos en una línea celular para obtener datos de secuenciación de al menos un exoma, transcriptoma o genoma entero de la línea celular, donde los datos de secuenciación incluyen al menos una secuencia de nucleótidos que incluye una alteración.

[00252] El conjunto de datos de entrenamiento puede generarse en función de obtener datos de secuenciación de nucleótidos normales de al menos un exoma, transcriptoma o genoma entero, de muestras de tejido normal.

[00253] El conjunto de datos de entrenamiento también puede incluir datos asociados con secuencias de proteoma asociadas con las muestras,

[00254] El conjunto de datos de entrenamiento también puede incluir datos asociados con secuencias de peptidoma de MHC asociadas con las muestras,

[00255] El conjunto de datos de entrenamiento también puede incluir datos asociados con las mediciones de afinidad de unión a MHC-péptido para al menos uno de los péptidos aislados.

[00256] El conjunto de datos de entrenamiento también puede incluir datos asociados con las mediciones de estabilidad de unión a MHC-péptido para al menos uno de los péptidos aislados.

[00257] El conjunto de datos de entrenamiento también puede incluir datos asociados con transcriptomas asociados con las muestras,

[00258] El conjunto de datos de entrenamiento también puede incluir datos asociados con genomas asociados con las muestras,

[00259] Las secuencias de péptidos de entrenamiento pueden tener longitudes dentro de un intervalo de k-mers donde k es entre 8-15, inclusive para clase I de MHC o 9-30 inclusive para clase II de MHC.

[00260] Un método descrito en la presente también puede incluir codificar la secuencia de péptidos usando un esquema de codificación *one-hot*.

[00261] Un método descrito en la presente también puede incluir codificar las secuencias de péptidos de entrenamiento usando un esquema de codificación *one-hot* con relleno a la izquierda.

[00262] Un método para tratar a un sujeto con un tumor, que comprende realizar los pasos de cualquiera de los métodos de identificación descritos en la presente y que además comprende obtener una vacuna tumoral que comprende el conjunto de neoantígenos seleccionados y administrarle la vacuna tumoral al sujeto.

[00263] Un método descrito en la presente puede incluir también identificar uno o más linfocitos T que son específicos para antígenos para al menos uno de los neoantígenos del subconjunto. En algunas

modalidades, la identificación comprende cocultivar el uno o más linfocitos T con uno o más de los neoantígenos del subconjunto en condiciones que expanden el uno o más linfocitos T específicos para el antígeno. En modalidades adicionales, la identificación comprende poner en contacto el uno o más linfocitos T con un tetrámero que comprende uno o más de los neoantígenos del subconjunto en condiciones que permiten la unión entre el linfocito T y el tetrámero. En otras modalidades adicionales, el método descrito en la presente también puede incluir identificar uno o más receptores de linfocitos T (TCR) del uno o más linfocitos T identificados. En determinadas modalidades, la identificación del uno o más receptores de linfocitos T comprende secuenciar las secuencias de los receptores de linfocitos T del uno o más linfocitos T identificados. El método descrito en la presente puede comprender además modificar genéticamente múltiples linfocitos T para que expresen al menos uno o más de los receptores de linfocitos T identificados; cultivar los múltiples linfocitos T en condiciones que expanden los múltiples linfocitos T; e infundir los linfocitos T expandidos en el sujeto. En algunas modalidades, la modificación genética de los múltiples linfocitos T para que expresen al menos uno o más de los receptores de linfocitos T identificados comprende clonar las secuencias de los receptores de linfocitos T del uno o más linfocitos T identificados en un vector de expresión; y transfectar cada uno de los múltiples linfocitos T con el vector de expresión. En algunas modalidades, el método descrito en la presente comprende además cultivar el uno o más linfocitos T cultivados en condiciones que expandan el uno o más linfocitos T identificados; e infundir los linfocitos T expandidos en el sujeto.

[00264] También se describe en la presente un linfocito T aislado que es específico para al menos un neoantígeno seleccionado del subconjunto.

[00265] En la presente también se describen métodos para elaborar una vacuna tumoral, que comprenden las etapas de: obtener los datos de secuenciación de nucleótidos tumorales de al menos uno de exoma, transcriptoma o genoma entero de la célula tumoral del sujeto, donde los datos de secuenciación de nucleótidos tumorales se usan para obtener datos que representan las secuencias de péptidos de cada uno de un conjunto de neoantígenos y donde la secuencia de péptidos de cada neoantígeno comprende al menos una alteración que la diferencia de la secuencia de péptidos de tipo salvaje correspondiente; ingresar la secuencia de péptidos de cada neoantígeno en uno o más modelos de presentación para generar un conjunto de probabilidades numéricas de que cada uno de los neoantígenos se presente por uno o más alelos de MHC en la superficie de células tumorales de la célula tumoral del sujeto, el conjunto de probabilidades numéricas se identificó al menos en función de los datos de espectrometría de masas recibidos y seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades numéricas de generar un conjunto de neoantígenos seleccionados y producir o haber producido una vacuna tumoral que comprende el conjunto de neoantígenos seleccionados.

[00266] En la presente también se describe una vacuna tumoral que incluye un conjunto de neoantígenos seleccionados que se seleccionan al llevar a cabo el método que comprende las etapas de: obtener los datos de secuenciación de nucleótidos tumorales de al menos uno de exoma, transcriptoma o genoma entero de la célula tumoral del sujeto, donde los datos de secuenciación de nucleótidos tumorales se usan para obtener datos que representan las secuencias de péptidos de cada uno de un conjunto de neoantígenos y donde la secuencia de péptidos de cada neoantígeno comprende al menos una alteración que la diferencia de la secuencia de péptidos de tipo salvaje correspondiente; ingresar la secuencia de péptidos de cada neoantígeno en uno o más modelos de presentación para generar un conjunto de probabilidades numéricas de que cada uno de los neoantígenos se presente por uno o más alelos de MHC en la superficie de células tumorales de la célula tumoral del sujeto, el conjunto de probabilidades numéricas se identificó al menos en función de los datos de espectrometría de masas recibidos y seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades numéricas de generar un conjunto de neoantígenos seleccionados y producir o haber producido una vacuna tumoral que comprende el conjunto de neoantígenos seleccionados.

[00267] La vacuna tumoral puede incluir una o más de una secuencia de nucleótidos, una secuencia de polipéptidos, ARN, ADN, una célula, un plásmido o un vector.

[00268] La vacuna tumoral puede incluir uno o más neoantígenos presentados en la superficie de células tumorales.

[00269] La vacuna tumoral puede incluir uno o más neoantígenos que son inmunógenos en el sujeto.

[00270] La vacuna tumoral no puede incluir uno o más neoantígenos que inducen una respuesta autoinmunitaria contra el tejido normal en el sujeto.

[00271] La vacuna tumoral puede incluir un adyuvante.

[00272] La vacuna tumoral puede incluir un excipiente.

[00273] Un método descrito en la presente también puede incluir seleccionar neoantígenos con mayor probabilidad de presentarse en la superficie de células tumorales respecto de neoantígenos no seleccionados en función del modelo de presentación.

[00274] Un método descrito en la presente también puede incluir seleccionar neoantígenos con mayor probabilidad de poder inducir una respuesta inmunitaria específica de un tumor en el sujeto respecto de neoantígenos no seleccionados en función del modelo de presentación.

[00275] Un método descrito en la presente también puede incluir seleccionar neoantígenos con mayor probabilidad de poder presentarse a linfocitos T sin tratamiento por células que presentan antígenos (APC, por sus siglas en inglés) profesionales respecto de neoantígenos no seleccionados en función del modelo de presentación, opcionalmente donde la APC es una célula dendrítica (DC, por sus siglas en inglés).

[00276] Un método descrito en la presente también puede incluir seleccionar neoantígenos con menor probabilidad de someterse a inhibición mediante tolerancia central o periférica respecto de neoantígenos no seleccionados en función del modelo de presentación.

[00277] Un método descrito en la presente también puede incluir seleccionar neoantígenos con menor probabilidad de poder inducir una respuesta autoinmunitaria al tejido normal en el sujeto respecto de neoantígenos no seleccionados en función del modelo de presentación.

[00278] Los datos de secuenciación de nucleótidos de exoma o transcriptoma pueden obtenerse al realizar secuenciación en el tejido tumoral.

[00279] La secuenciación puede ser secuenciación de nueva generación (NGS, por sus siglas en inglés) o cualquier enfoque de secuenciación masiva en paralelo.

[00280] El conjunto de probabilidades numéricas también puede identificarse por al menos rasgos de interacción con alelos de MHC que comprenden al menos uno de: la afinidad predicha con que se unen el alelo de MHC y el péptido codificado por neoantígeno; la estabilidad predicha del complejo de péptido codificado por neoantígeno y MHC; la secuencia y longitud del péptido codificado por neoantígeno; la probabilidad de presentación de péptidos codificados por neoantígeno con secuencia similar en células de otros individuos que expresan el alelo de MHC particular según se evalúa por proteómica de espectrometría de masas u otros medios; los niveles de expresión del alelo de MHC particular en el sujeto en cuestión (p. ej., como se mide por secuenciación de ARN o espectrometría de masas); la probabilidad de presentación global por el alelo de MHC particular, independiente de las secuencias de péptidos codificados por neoantígeno en otros sujetos distintos que expresan el alelo de MHC particular; la probabilidad de presentación global por alelos de MHC independiente de las secuencias de péptidos codificados por neoantígeno en la misma familia de moléculas (p. ej., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) en otros sujetos distintos.

[00281] El conjunto de probabilidades numéricas se identifican adicionalmente por al menos rasgos de no interacción con alelos de MHC que comprenden al menos uno de: las secuencias de extremo C y N que flanquean el péptido codificado por neoantígeno en su secuencia de proteínas fuente; la presencia de motivos de escisión de proteasa en el péptido codificado por neoantígeno, opcionalmente ponderada de acuerdo con la expresión de proteasas correspondientes en las células tumorales (como se mide por secuenciación de ARN o espectrometría de masas); el índice de renovación de la proteína fuente como se mide en el tipo celular apropiado; la longitud de la proteína fuente, opcionalmente considerando las variantes de empalme específicas («isoformas») expresadas mayormente en las células tumorales como se mide por secuenciación de ARN o espectrometría de masas de proteoma, o como se predice de la anotación de mutaciones de empalme somáticas o de línea germinal detectadas en datos de secuencia de ADN o ARN; el nivel de expresión del proteasoma, inmunoproteasoma, timoproteasoma u otras

proteasas en las células tumorales (que pueden medirse por secuenciación de ARN, espectrometría de masas de proteoma o inmunohistoquímica); la expresión del gen fuente del péptido codificado por neoantígeno (p. ej., según medición a través de secuenciación de ARN o espectrometría de masas), la expresión específica del tejido típica del gen fuente del péptido codificado por neoantígeno durante varias etapas del ciclo celular; un catálogo integral de rasgos de la proteína fuente y/o sus dominios como puede encontrarse, p. ej., en uniProt o PDB http://www.rcsb.org/pdb/home/home.do; rasgos que describen propiedades del dominio de la proteína fuente que contiene el péptido, por ejemplo: estructura secundaria o terciaria (p. ej., hélice alfa en contraste con lámina beta); empalme alternativo; la probabilidad de presentación de péptidos de la proteína fuente del péptido codificado por neoantígeno en cuestión en otros sujetos distintos; la probabilidad de que el péptido no sea detectado ni representado en exceso por espectrometría de masas debido a sesgos técnicos; la expresión de varios módulos/vías de genes como se mide por secuenciación de ARN (que no es necesario que contenga la proteína fuente del péptido) que informan sobre el estado de las células tumorales, estroma, o linfocitos que infiltran el tumor (TIL, por sus siglas en inglés); la cantidad de copia del gen fuente del péptido codificado por neoantígeno en las células tumorales; la probabilidad de que el péptido se una al TAP o la afinidad de unión medida o predicha del péptido al TAP; el nivel de expresión de TAP en las células tumorales (que puede medirse por secuenciación de ARN, espectrometría de masas de proteoma, inmunohistoquímica); presencia o ausencia de mutaciones tumorales que incluyen, de modo no taxativo: mutaciones impulsoras en genes impulsores de cáncer conocidos como EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3, y en genes que codifican las proteínas implicadas en la maquinaria de presentación de antígenos (p. ej., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 o cualquiera de los genes que codifican componentes del proteasoma o inmunoproteasoma). Los péptidos cuya presentación depende de un componente de la maquinaría de presentación de antígenos que está sujeto a una mutación de pérdida de función en el tumor tienen menor probabilidad de presentación; la presencia o ausencia de polimorfismos de línea germinal funcional que incluyen, de modo no taxativo: en genes que codifican las proteínas implicadas en la maquinaria de presentación de antígenos (p. ej., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 o cualquiera de los genes que codifican componentes de la proteasoma o inmunoproteasoma); tipo tumoral (p. ej., NSCLC, melanoma); subtipo de tumor clínico (p. ej., cáncer pulmonar escamoso en contraste con no escamoso); historial de

tabaquismo; la expresión típica del gen fuente del péptido en el tipo tumoral relevante o subtipo clínico, opcionalmente estratificada por una mutación impulsora.

[00282] Al menos una alteración puede ser un indel de desplazamiento de marco o que no es de desplazamiento de marco, sustitución de cambio de sentido o sin sentido, alteración de sitio de empalme, reorganización genómica o fusión genética o cualquier alteración genómica o de expresión que conduzca a un neoORF.

[00283] La célula tumoral se puede seleccionar del grupo que consiste en: cáncer de pulmón, melanoma, cáncer de mama, cáncer de ovario, cáncer de próstata, cáncer de riñón, cáncer gástrico, cáncer de colon, cáncer testicular, cáncer de cabeza y cuello, cáncer pancreático, cáncer de cerebro, linfoma de linfocitos B, leucemia mielógena aguda, leucemia mielógena crónica, leucemia linfocítica crónica y leucemia linfocítica de linfocitos T, cáncer pulmonar no microcítico y cáncer pulmonar microcítico.

[00284] Un método descrito en la presente también puede incluir obtener una vacuna tumoral que comprende el conjunto de neoantígenos seleccionados o un subconjunto de estos, que opcionalmente comprende además administrarle la vacuna tumoral al sujeto.

[00285] Al menos uno de los neoantígenos en el conjunto de neoantígenos seleccionados, cuando está en la forma de un polipéptido, puede incluir al menos uno de: una afinidad de unión con MHC con un valor de IC50 menor que 1000 nM, para polipéptidos de Clase I de MHC una longitud de 8-15, 8, 9, 10, 11, 12, 13, 14 o 15 aminoácidos, para polipéptidos de Clase II de MHC una longitud de 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 aminoácidos presencia de motivos de secuencia dentro o cerca del polipéptido en la secuencia de proteínas original que promueve la escisión del proteosoma y presencia de secuencia de motivos de secuencia que promueven el transporte de TAP. Para los MHC clase II, la presencia de motivos de secuencia dentro o cerca del péptido que promueve la escisión mediante proteasas lisosomales o extracelulares (por ejemplo, catepsinas) o unión a HLA catalizada por HLA-DM.

[00286] En la presente también se describe un método para generar un modelo para identificar uno o más neoantígenos que probablemente se presenten en una superficie de células tumorales de una célula tumoral, que comprende las etapas de: recibir datos de espectrometría de masas que comprenden datos asociados con múltiples péptidos aislados eluidos de un complejo mayor de histocompatibilidad (MHC, por sus siglas en inglés) obtenido de múltiples muestras; obtener un conjunto de datos de entrenamiento por al menos la identificación de un conjunto de secuencias de péptidos de entrenamiento presentes en las muestras y uno o más MHC asociados con cada secuencia de péptidos de entrenamiento; entrenar un conjunto de parámetros numéricos de un modelo de presentación usando el conjunto de datos de entrenamiento que comprende las secuencias de péptidos de entrenamiento,

donde el modelo de presentación provee múltiples probabilidades numéricas de que las secuencias de péptidos de la célula tumoral sean presentadas por uno o más alelos de MHC en la superficie de células tumorales.

[00287] El modelo de presentación puede representar dependencia entre: la presencia de un aminoácido particular en una posición particular de una secuencia de péptidos y la probabilidad de presentación, por uno de los alelos de MHC en la célula tumoral, de la secuencia de péptidos que contiene el aminoácido particular en la posición particular.

[00288] Las muestras también pueden incluir líneas celulares modificadas genéticamente para expresar un único alelo de clase I o clase II de MHC.

[00289] Las muestras también pueden incluir líneas celulares modificadas genéticamente para expresar múltiples alelos de clase I o clase II de MHC.

[00290] Las muestras también pueden incluir líneas celulares humanas obtenidas o derivadas de múltiples pacientes.

[00291] Las muestras también pueden incluir muestras de tumores frescas o congeladas obtenidas de múltiples pacientes.

[00292] Las muestras también pueden incluir péptidos identificados usando ensayos de linfocitos T.

[00293] El conjunto de datos de entrenamiento también puede incluir datos asociados con: abundancia de péptidos del conjunto de péptidos de entrenamiento presente en las muestras; longitud de péptidos del conjunto de péptidos de entrenamiento en las muestras.

[00294] Un método descrito en la presente también puede incluir obtener un conjunto de secuencias de proteínas de entrenamiento en función de las secuencias de péptidos de entrenamiento mediante alineación a una base de datos que comprende un conjunto de secuencias de proteínas conocidas, donde el conjunto de secuencias de proteínas de entrenamiento es más largo que las secuencias de péptidos de entrenamiento y las incluye.

[00295] Un método descrito en la presente también puede incluir realizar o haber realizado espectrometría de masas en una línea celular para obtener datos de secuenciación de nucleótidos de al menos un exoma, transcriptoma o genoma entero de la línea celular, donde los datos de secuenciación de nucleótidos incluyen al menos una secuencia de proteínas que incluye una mutación.

[00296] Un método descrito en la presente también puede incluir: codificar la secuencia de péptidos de capacitación usando un esquema de codificación *one-hot*.

[00297] Un método descrito en la presente también puede incluir obtener datos de secuenciación de nucleótidos de al menos un exoma, transcriptoma o genoma entero de muestras de tejido normal y entrenar el conjunto de parámetros del modelo de presentación usando los datos de secuenciación de

nucleótidos normales.

[00298] El conjunto de datos de entrenamiento también puede incluir datos asociados con secuencias de proteoma asociadas con las muestras,

[00299] El conjunto de datos de entrenamiento también puede incluir datos asociados con secuencias de peptidoma de MHC asociadas con las muestras,

[00300] El conjunto de datos de entrenamiento también puede incluir datos asociados con las mediciones de afinidad de unión a MHC-péptido para al menos uno de los péptidos aislados.

[00301] El conjunto de datos de entrenamiento también puede incluir datos asociados con las mediciones de estabilidad de unión a MHC-péptido para al menos uno de los péptidos aislados.

[00302] El conjunto de datos de entrenamiento también puede incluir datos asociados con transcriptomas asociados con las muestras,

[00303] El conjunto de datos de entrenamiento también puede incluir datos asociados con genomas asociados con las muestras,

[00304] Un método descrito en la presente también puede incluir la regresión logística del conjunto de parámetros.

[00305] Las secuencias de péptidos de entrenamiento pueden tener longitudes dentro de un intervalo de k-mers donde k es entre 8-15, inclusive para clase I de MHC o 6-30, inclusive para clase II de MHC.

[00306] Un método descrito en la presente también puede incluir codificar las secuencias de péptidos de entrenamiento usando un esquema de codificación *one-hot* con relleno a la izquierda.

[00307] Un método descrito en la presente también puede incluir determinar valores para el conjunto de parámetros usando un algoritmo de aprendizaje profundo.

[00308] En la presente se describen métodos para identificar uno o más neoantígenos que probablemente se presenten en una superficie de células tumorales de una célula tumoral, que comprenden ejecutar las etapas de: recibir datos de espectrometría de masas que comprenden datos asociados con múltiples péptidos aislados eluidos de un complejo mayor de histocompatibilidad (MHC) obtenido de múltiples muestras de tumores frescas o congeladas; obtener un conjunto de datos de entrenamiento por al menos la identificación de un conjunto de secuencias de péptidos de entrenamiento presentes en las muestras tumorales y presentes en uno o más alelos de MHC asociados con cada secuencia de péptidos de entrenamiento; obtener un conjunto de secuencias de proteínas de entrenamiento en función de las secuencias de péptidos de entrenamiento y entrenar un conjunto de parámetros numéricos de un modelo de presentación usando las secuencias de proteínas de entrenamiento y las secuencias de péptidos de entrenamiento, donde el modelo de presentación provee múltiples probabilidades numéricas de que las secuencias de péptidos de la célula tumoral sean

presentadas por uno o más alelos de MHC en la superficie de células tumorales.

[00309] El modelo de presentación puede representar dependencia entre: la presencia de un par de un alelo de MHC particular y un aminoácido particular en una posición particular de una secuencia de péptidos y la probabilidad de presentación en la superficie de células tumorales, por el alelo de MHC particular del par, de tal secuencia de péptidos que comprende el aminoácido particular en la posición particular.

[00310] Un método descrito en la presente también puede incluir seleccionar un subconjunto de neoantígenos, donde el subconjunto de neoantígenos se selecciona debido a que cada uno tiene mayor probabilidad de presentarse en la superficie celular del tumor respecto de uno o más neoantígenos tumorales diferentes.

[00311] Un método descrito en la presente también puede incluir seleccionar un subconjunto de neoantígenos, donde el subconjunto de neoantígenos se selecciona debido a que cada uno tiene mayor probabilidad de poder inducir una respuesta inmunitaria específica del tumor en el sujeto respecto de uno o más neoantígenos tumorales diferentes.

[00312] Un método descrito en la presente también puede incluir seleccionar un subconjunto de neoantígenos, donde el subconjunto de neoantígenos se selecciona debido a que cada uno tiene mayor probabilidad de poder presentarse a linfocitos T sin tratamiento por células que presentan antígenos (APC) profesionales respecto de uno o más neoantígenos tumorales diferentes, opcionalmente donde la APC es una célula dendrítica (DC).

[00313] Un método descrito en la presente también puede incluir seleccionar un subconjunto de neoantígenos, donde el subconjunto de neoantígenos se selecciona debido a que cada uno tiene menor probabilidad de estar sujeto a inhibición mediante tolerancia central o periférica respecto de uno o más neoantígenos tumorales diferentes.

[00314] Un método descrito en la presente también puede incluir seleccionar un subconjunto de neoantígenos, donde el subconjunto de neoantígenos se selecciona debido a que cada uno tiene menor probabilidad de poder inducir una respuesta autoinmunitaria al tejido normal en el sujeto respecto de uno o más neoantígenos tumorales diferentes.

[00315] Un método descrito en la presente también puede incluir seleccionar un subconjunto de neoantígenos, donde el subconjunto de neoantígenos se selecciona debido a que cada uno tiene menor probabilidad de modificarse postraduccionalmente de manera diferencial en células tumorales en contraste con APC, opcionalmente donde la APC es una célula dendrítica (DC).

[00316] La práctica de los métodos de la presente empleará, a menos que se indique lo contrario, métodos convencionales de química de proteínas, bioquímica, técnicas de ADN recombinante y farmacología, dentro de la experiencia en la técnica. Estas técnicas se explican de forma exhaustiva en

la bibliografía. Véase, por ejemplo, T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., adición actual); Sambrook, et ál., *Molecular Cloning: A Laboratory Manual* (2ª edición, 1989); *Methods In Enzymology* (S. Colowick y N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18ª edición (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey y Sundberg *Advanced Organic Chemistry 3ª Ed.* (Plenum Press) Vol. A y B(1992).

III. Identificación de mutaciones específicas de tumores en neoantígenos

[00317] En la presente también se describen métodos para la identificación de ciertas mutaciones (p. ej., las variantes o alelos presentes en las células cancerosas). En particular, estas mutaciones pueden estar presentes en el genoma, transcriptoma, proteoma o exoma de células cancerosas de un sujeto que padece cáncer, pero no en el tejido normal del sujeto.

[00318] Las mutaciones genéticas en tumores pueden considerarse útiles para el direccionamiento inmunológico de tumores si llevan a cambios en la secuencia de aminoácidos de una proteína exclusivamente en el tumor. Las mutaciones útiles incluyen: (1) mutaciones no sinónimas que conducen a diferentes aminoácidos en la proteína; (2) mutaciones de translectura en las que se modifica o elimina un codón de terminación, lo que provoca la traducción de una proteína más larga con una secuencia específica de tumores novedosa en el extremo C; (3) mutaciones del sitio de empalme que llevan a la inclusión de un intrón en el ARNm maduro y, por lo tanto, una secuencia de proteínas específica de tumores única; (4) reorganizaciones cromosómicas que dan lugar a una proteína quimérica con secuencias específicas de tumores en la unión de 2 proteínas (es decir, fusión genética); (5) mutaciones o eliminaciones de desplazamiento de marco que conducen a un nuevo marco de lectura abierto con una secuencia de proteínas específica de tumores novedosa. Las mutaciones también pueden incluir una o más de indeles que no son de desplazamiento de marco, sustitución de cambio de sentido o sin sentido, alteración de sitio de empalme, reorganización genómica o fusión genética o cualquier alteración genómica o de expresión que conduzca a un neoORF.

[00319] Los péptidos con mutaciones o polipéptidos mutados que surgen, por ejemplo, de mutaciones del sitio de empalme, de desplazamiento de marco, de translectura o de fusión genética en las células tumorales pueden identificarse mediante secuenciación del ADN, ARN o la proteína en células tumorales en contraste con células normales.

[00320] Las mutaciones también pueden incluir mutaciones específicas de tumores identificadas previamente. Las mutaciones tumorales conocidas pueden encontrarse en la base de datos del Catálogo de mutaciones somáticas en el cáncer (COSMIC, por sus siglas en inglés).

[00321] Se encuentra disponible una variedad de métodos para detectar la presencia de una mutación o alelo particular en el ADN o ARN de un individuo. Los avances en este campo han proporcionado

genotipado de SNP a gran escala preciso, fácil y económico. Por ejemplo, se han descrito varias técnicas que incluyen la hibridación dinámica específica de alelos (DASH, por sus siglas en inglés), electroforesis en gel diagonal de matriz de microplacas (MADGE, por sus siglas en inglés), pirosecuenciación, ligación específica de oligonucleótidos, el sistema TaqMan, así como también diversas tecnologías de «chip» de ADN como los chips de SNP Affymetrix. Estos métodos utilizan la amplificación de la región genética objetivo, típicamente mediante PCR. Aun otros métodos, basados en la generación de moléculas de señales pequeñas mediante escisión invasiva seguida de espectrometría de masas o sondas de cadena inmovilizadas y amplificación por círculo rodante. Más adelante se resumen varios de los métodos conocidos en la técnica para detectar mutaciones específicas.

[00322] Los medios de detección basados en PCR pueden incluir amplificación de multiplex de múltiples marcadores de manera simultánea. Por ejemplo, se conoce en la técnica la selección de cebadores de PCR para generar productos de PCR que no se superponen en tamaño y pueden analizarse de manera simultánea. De manera alternativa, es posible amplificar diferentes marcadores con cebadores que están etiquetados de manera diferencial y, por lo tanto, cada uno puede detectarse de manera diferencial. Por supuesto, los medios de detección basados en hibridación permiten la detección diferencial de múltiples productos de PCR en una muestra. En la técnica se conocen otras técnicas para permitir análisis de multiplex de múltiples marcadores.

[00323] Se han desarrollado varios métodos para facilitar el análisis de polimorfismos de nucleótidos simples en ADN genómico o ARN celular. Por ejemplo, un polimorfismo de base simple puede detectarse usando un nucleótido especializado resistente a exonucleasa, como se describe, p. ej., en Mundy, C. R. (patente estadounidense n.° 4.656.127). De acuerdo con el método, un cebador complementario a la secuencia alélica inmediatamente 3' con respecto al sitio polimórfico puede hibridar con una molécula objetivo obtenida de un animal o humano particular. Si el sitio polimórfico en la molécula objetivo contiene un nucleótido que es complementario al derivado de nucleótido resistente a exonucleasa particular presente, entonces ese derivado se incorporará al extremo del cebador hibridado. Dicha incorporación vuelve al cebador resistente a la exonucleasa y, de esta forma, permite su detección. Dado que se conoce la identidad del derivado resistente a exonucleasa de la muestra, descubrir que el cebador se ha vuelto resistente a las exonucleasas revela que el o los nucleótidos presentes en el sitio polimórfico de la molécula objetivo son complementarios al del derivado de nucleótido utilizado en la reacción. Este método tiene la ventaja de que no requiere la determinación de grandes cantidades de datos externos de secuencias.

[00324] Un método basado en solución se puede usar para determinar la identidad de un nucleótido de un sitio polimórfico. Cohen, D. et ál. (Patente francesa 2.650.840; solicitud PCT n.° WO91/02087). Al igual que en el método de Mundy de la patente estadounidense n.° 4.656.127, se emplea un cebador

que es complementario a las secuencias alélicas inmediatamente 3' con respecto a un sitio polimórfico. El método determina la identidad del nucleótido de ese sitio usando derivados de didesoxinucleótidos etiquetados que, si son complementarios al nucleótido del sitio polimórfico, se incorporarán en el extremo del cebador.

[00325] Un método alternativo, conocido como Genetic Bit Analysis o GBA se describe por Goelet, P. et ál. (Solicitud PCT n.º 92/15712). El método de Goelet, P. et ál. utiliza mezclas de terminadores etiquetados y un cebador que es complementario a la secuencia 3' con respecto a un sitio polimórfico. Por lo tanto, el terminador etiquetado que se incorpora es determinado por, y complementario a, el nucleótido presente en el sitio polimórfico de la molécula objetivo que se evalúa. A diferencia del método de Cohen et ál. (Patente francesa 2.650.840; solicitud PCT n.° WO91/02087) el método de Goelet, P. et ál. Puede ser un ensayo de fase heterogénea, en el cual el cebador o la molécula objetivo se inmoviliza en una fase sólida.

[00326] Se han descrito varios procedimientos de incorporación de nucleótidos guiados por cebadores para el ensayo de sitios polimórficos en el ADN (Komher, J. S. et ál., Nucl Acids. Res. 17:7779-7784 (1989); Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990); Syvanen, A.-C., et ál., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et ál., Proc. Natl. Acad. Sci. (EUA) 88:1143-1147 (1991); Prezant, T. R. et ál., Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et ál., GATA 9:107-112 (1992); Nyren, P. et ál., Anal. Biochem. 208:171-175 (1993)). Estos métodos difieren de GBA porque estos utilizan la incorporación de desoxinucleótidos etiquetados para distinguir las bases en un sitio polimórfico. En dicho formato, dado que la señal es proporcional a la cantidad de desoxinucleótidos incorporados, los polimorfismos que se producen en ejecuciones del mismo nucleótido pueden dar como resultado señales que son proporcionales a la longitud de la ejecución (Syvanen, A.-C, et ál., Amer. J. Hum. Genet. 52:46-59 (1993)).

[00327] Una cantidad de iniciativas obtienen información de secuencias directamente de millones de moléculas individuales de ADN o ARN en paralelo. Las tecnologías de secuenciación por síntesis de molécula simple en tiempo real dependen de la detección de nucleótidos fluorescentes a medida que se incorporan en una cadena naciente de ADN que es complementaria a la plantilla que se somete a secuenciación. En un método, los oligonucleótidos con 30-50 bases de longitud se anclan covalentemente en el extremo 5’ de cubreobjetos de vidrio. Estas hebras ancladas desempeñan dos funciones. En primer lugar, actúan como sitios de captura de las hebras de plantilla objetivo si las plantillas están configuradas con colas de captura complementarias a los oligonucleótidos unidos a la superficie. También actúan como cebadores para la extensión de cebador dirigida a la plantilla que forma la base de la lectura de secuencia. Los cebadores de captura funcionan como un sitio de posición fija para la determinación de secuencias usando múltiples ciclos de síntesis, detección y escisión química

del enlazador tinte para retirar el tinte. Cada ciclo consiste en agregar la mezcla de polimerasa/nucleótidos etiquetados, enjuagar, tomar imágenes y escindir el tinte. En un método alternativo, la polimerasa se modifica con una molécula donante fluorescente y se inmoviliza en un portaobjetos de vidrio, mientras que cada nucleótido está codificado por color con un resto fluorescente aceptor unido a gamma-fosfato. El sistema detecta la interacción entre una polimerasa etiquetada de forma fluorescente y un nucleótido modificado de forma fluorescente ya que el nucleótido se incorpora en la cadena de novo. También existen otras tecnologías de secuenciación por síntesis.

[00328] Se puede utilizar cualquier plataforma de secuenciación por síntesis adecuada para identificar las mutaciones. Como se describe anteriormente, en la actualidad se encuentran disponibles cuatro plataformas principales de secuenciación por síntesis: Genome Sequencers de Roche/454 Life Sciences, 1G Analyzer de Illumina/Solexa, el sistema SOLiD de Applied BioSystems y el sistema Heliscope de Helicos Biosciences. Las plataformas de secuenciación por síntesis también han sido descritas por Pacific Biosciences y VisiGen Biotechnologies. En algunas modalidades, múltiples moléculas de ácido nucleico que se secuencian están unidas a un soporte (por ejemplo, soporte sólido). Para inmovilizar el ácido nucleico en un soporte, se puede agregar una secuencia de captura/sitio de cebado universal en el extremo 3' y/o 5' de la plantilla. Los ácidos nucleicos pueden estar unidos al soporte al hibridar la secuencia de captura con una secuencia complementaria unida covalentemente al soporte. La secuencia de captura (también denominada secuencia de captura universal) es una secuencia de ácido nucleico complementaria a una secuencia unida a un soporte que puede servir doblemente como cebador universal.

[00329] Como alternativa a una secuencia de captura, un miembro de un par de acoplamiento (como, por ejemplo, anticuerpo/antígeno, receptor/ligando o el par de avidina-biotina como se describe en, por ejemplo, la solicitud de patente estadounidense n.° 2006/0252077) puede unirse a cada fragmento que se capturará en una superficie recubierta con un segundo miembro respectivo de ese par de acoplamiento.

[00330] Después de la captura, la secuencia se puede analizar, por ejemplo, mediante detección/secuenciación de molécula simple, p. ej., como se describe en los Ejemplos y en la patente estadounidense n.° 7.283.337, que incluye secuenciación mediante síntesis dependiente de plantillas. En la secuenciación mediante síntesis, la molécula unida a la superficie se expone a múltiples trifosfatos de nucleótidos etiquetados en presencia de polimerasa. La secuencia de la plantilla se determina mediante el orden de los nucleótidos etiquetados incorporados al extremo 3’ de la cadena en crecimiento. Esto puede realizarse en tiempo real o en un modo de etapa y repetición. Para el análisis en tiempo real, se pueden incorporar diferentes etiquetas ópticas a cada nucleótido y se pueden utilizar múltiples láseres para la estimulación de los nucleótidos incorporados.

[00331] La secuenciación también puede incluir otras plataformas y técnicas de secuenciación masiva en paralelo o secuenciación de nueva generación (NGS, por sus siglas en inglés). Ejemplos adicionales de plataformas y técnicas de secuenciación masiva en paralelo son Illumina HiSeq o MiSeq, Thermo PGM o Proton, Pac Bio RS II o Sequel, Qiagen's Gene Reader y Oxford Nanopore MinION. Se pueden usar otras tecnologías de secuenciación masiva en paralelo similares actuales, así como futuras generaciones de estas tecnologías.

[00332] Se puede utilizar cualquier tipo de célula o tejido para obtener muestras de ácido nucleico para utilizar en los métodos descritos en la presente. Por ejemplo, se puede obtener una muestra de ADN o ARN de un tumor o un fluido corporal, por ejemplo, sangre, obtenida mediante técnicas conocidas (por ejemplo, venopunción) o saliva. De manera alternativa, las pruebas de ácido nucleico se pueden realizar en muestras secas (por ejemplo, pelo o piel). Además, una muestra se puede obtener para secuenciación de un tumor y otra muestra se puede obtener de tejido normal para secuenciación donde el tejido normal es del mismo tipo de tejido que el tumor. Una muestra se puede obtener para secuenciación de un tumor y otra muestra se puede obtener de tejido normal para secuenciación donde el tejido normal es de un tipo de tejido diferente que el del tumor.

[00333] Los tumores pueden incluir uno o más de cáncer de pulmón, melanoma, cáncer de mama, cáncer de ovario, cáncer de próstata, cáncer de riñón, cáncer gástrico, cáncer de colon, cáncer testicular, cáncer de cabeza y cuello, cáncer pancreático, cáncer de cerebro, linfoma de linfocitos B, leucemia mielógena aguda, leucemia mielógena crónica, leucemia linfocítica crónica y leucemia linfocítica de linfocitos T, cáncer pulmonar no microcítico y cáncer pulmonar microcítico.

[00334] De manera alternativa, se puede usar espectrometría de masas de proteína para identificar o validar la presencia de péptidos mutados unidos a proteínas de MHC en células tumorales. Los péptidos se pueden eluir con ácido de células tumorales o de moléculas de HLA que se inmunoprecipitan del tumor y luego se identifican usando espectrometría de masas.

IV. Neoantígenos

[00335] Los neoantígenos pueden incluir nucleótidos o polipéptidos. Por ejemplo, un neoantígeno puede ser una secuencia de ARN que codifica una secuencia de polipéptidos. Por lo tanto los neoantígenos útiles en vacunas pueden incluir secuencias de nucleótidos o secuencias de polipéptidos.

[00336] En la presente se describen péptidos que comprenden mutaciones específicas de tumores identificadas por los métodos descritos en la presente, péptidos que comprenden mutaciones específicas de tumores conocidas y polipéptidos mutantes o fragmentos de estos identificados por los métodos descritos en la presente. Los péptidos de neoantígenos pueden describirse en el contexto de su secuencia codificante donde un neoantígeno incluye la secuencia de nucleótidos (p. ej., ADN o ARN) que codifica la secuencia de polipéptidos relacionada.

[00337] Uno o más polipéptidos codificados por una secuencia de nucleótidos de neoantígeno puede comprender al menos uno de: una afinidad de unión con MHC con un valor de IC50 menor que 1000 nM, para péptidos de Clase I de MHC una longitud de 8-15, 8, 9, 10, 11, 12, 13, 14 o 15 aminoácidos, presencia de motivos de secuencia dentro o cerca del péptido que promueve la escisión del proteosoma y presencia de secuencia de motivos de secuencia que promueven el transporte de TAP. Para péptidos del MHC clase II, una longitud de 6-30, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 aminoácidos, presencia de motivos de secuencia dentro o cerca del péptido que promueve la escisión mediante proteasas lisosomales o extracelulares (por ejemplo, catepsinas) o unión a HLA catalizada por HLA-DM.

[00338] Uno o más neoantígenos pueden presentarse en la superficie de un tumor.

[00339] Uno o más neoantígenos pueden ser inmunógenos en un sujeto con un tumor, p. ej., pueden provocar una respuesta de linfocitos T o una respuesta de linfocitos B en el sujeto.

[00340] Uno o más neoantígenos que inducen una respuesta autoinmunitaria en un sujeto pueden excluirse de consideración en el contexto de la generación de vacunas para un sujeto que tiene un tumor.

[00341] El tamaño de al menos una molécula de péptido neoantigénico puede comprender, de modo no taxativo, alrededor de 5, alrededor de 6, alrededor de 7, alrededor de 8, alrededor de 9, alrededor de 10, alrededor de 11, alrededor de 12, alrededor de 13, alrededor de 14, alrededor de 15, alrededor de 16, alrededor de 17, alrededor de 18, alrededor de 19, alrededor de 20, alrededor de 21, alrededor de 22, alrededor de 23, alrededor de 24, alrededor de 25, alrededor de 26, alrededor de 27, alrededor de 28, alrededor de 29, alrededor de 30, alrededor de 31, alrededor de 32, alrededor de 33, alrededor de 34, alrededor de 35, alrededor de 36, alrededor de 37, alrededor de 38, alrededor de 39, alrededor de 40, alrededor de 41, alrededor de 42, alrededor de 43, alrededor de 44, alrededor de 45, alrededor de 46, alrededor de 47, alrededor de 48, alrededor de 49, alrededor de 50, alrededor de 60, alrededor de 70, alrededor de 80, alrededor de 90, alrededor de 100, alrededor de 110, alrededor de 120 o más residuos de moléculas amino, y cualquier intervalo derivado de estos. En modalidades específicas, las moléculas de péptidos neoantigénicos son iguales o menores a 50 aminoácidos.

[00342] Los polipéptidos y péptidos neoantigénicos pueden ser: para Clase I de MHC 15 residuos o menos de longitud y habitualmente consisten en de alrededor de 8 a alrededor de 11 residuos, particularmente 9 o 10 residuos; para Clase II de MHC, 6-30 residuos, inclusive.

[00343] Si se desea, un péptido más largo puede diseñarse de varias formas. En un caso, cuando las probabilidades de presentación de péptidos en alelos de HLA se predicen o conocen, un péptido más largo podría consistir en: (1) péptidos presentados individualmente con una extensión de 2-5 aminoácidos hacia el extremo N y C de cada producto génico correspondiente; (2) una concatenación de algunos o todos los péptidos presentados con secuencias extendidas para cada uno. En otro caso,

cuando la secuenciación revela una secuencia de neoepítopos larga (>10 residuos) presente en el tumor (p. ej., debido a un desplazamiento de marco, translectura o inclusión de intrones que lleva a una secuencia de péptidos novedosa), un péptido más largo consistiría en: (3) toda la extensión de aminoácidos específicos del tumor novedosos, evitando de este modo la necesidad de selección basada en pruebas in vitro o informática del péptido más corto presentado por el HLA más resistente. En ambos casos, el uso de un péptido más largo permite el procesamiento endógeno por células del paciente y puede llevar a una presentación del antígeno e inducción de respuestas de linfocitos T más eficaces.

[00344] Los polipéptidos y péptidos neoantigénicos pueden presentarse en una proteína de HLA. En algunos aspectos, los polipéptidos y péptidos neoantigénicos se presentan en una proteína de HLA con mayor afinidad que un péptido de tipo salvaje. En algunos aspectos, un polipéptido o péptido neoantigénico puede tener una IC50 al menos menor que 5000 nM, al menos menor que 1000 nM, al menos menor que 500 nM, al menos menor que 250 nM, al menos menor que 200 nM, al menos menor que 150 nM, al menos menor que 100 nM, al menos menor que 50 nM o menos.

[00345] En algunos aspectos, los polipéptidos y péptidos neoantigénicos no inducen una respuesta autoinmunitaria y/o invocan tolerancia inmunológica cuando se administran a un sujeto.

[00346] También se proporcionan composiciones que comprenden al menos dos o más péptidos neoantigénicos. En algunas modalidades, la composición contiene al menos dos péptidos distintos. Al menos dos péptidos distintos pueden derivar del mismo polipéptido. Polipéptidos distintos significa que el péptido varía en longitud, secuencia de aminoácidos o ambos. Los péptidos derivan de cualquier polipéptido conocido o que se encontró que contiene una mutación específica de tumor. Los polipéptidos adecuados de los que pueden derivar los péptidos neoantigénicos pueden encontrarse, por ejemplo, en la base de datos COSMIC. COSMIC cura la información integral de mutaciones somáticas en el cáncer humano. El péptido contiene la mutación específica del tumor. En algunos aspectos, la mutación específica del tumor es una mutación impulsora de un tipo particular de cáncer.

[00347] Los polipéptidos y péptidos neoantigénicos que tienen una actividad o propiedad deseada pueden modificarse para proporcionar determinados atributos deseados, p. ej., características farmacológicas mejoradas, y al mismo tiempo aumentar o al menos retener sustancialmente toda la actividad biológica del péptido no modificado para unir la molécula de MHC deseada y activar el linfocito T adecuado. Por ejemplo, los polipéptidos y péptidos neoantigénicos pueden someterse a diversos cambios, como sustituciones, ya sea conservadoras o no conservadoras, donde dichos cambios pueden proporcionar determinadas ventajas en su uso, como mejor unión a MHC, estabilidad o presentación. Sustituciones conservadoras se refiere a remplazar un residuo de aminoácidos con otro que es biológicamente y/o químicamente similar, p. ej., un residuo hidrofóbico por otro o un residuo polar por otro. Las sustituciones incluyen combinaciones como Gly, Ala; Val, Ile, Leu, Met; Asp, Glu; Asn, Gln;

Ser, Thr; Lys, Arg; y Phe, Tyr. El efecto de las sustituciones de aminoácidos simples también pueden analizarse usando aminoácidos D. Dichas modificaciones pueden realizarse usando procedimientos conocidos de síntesis de péptidos, como se describe, p. ej., en Merrifield, Science 232:341-347 (1986), Barany & Merrifield, The Peptides, Gross & Meienhofer, eds. (N.Y., Academic Press), pp. 1-284 (1979); y Stewart & Young, Solid Phase Peptide Synthesis, (Rockford, Ill., Pierce), 2ª Ed. (1984).

[00348] Las modificaciones de péptidos y polipéptidos con varios miméticos de aminoácidos o aminoácidos no naturales pueden ser particularmente útiles para aumentar la estabilidad del péptido y polipéptido in vivo. La estabilidad puede someterse a ensayo de varias formas. Por ejemplo, se han usado peptidasas y varios medios biológicos, como plasma humano y suero, para evaluar la estabilidad. Véase, p. ej., Verhoef et ál., Eur. J. Drug Metab Pharmacokin. 11:291-302 (1986). La semivida de los péptidos puede determinarse de manera conveniente usando un ensayo de suero humano al 25 % (v/v). El protocolo es generalmente el siguiente. Suero humano agrupado (Tipo AB, no inactivado por calor) se deslipida por centrifugación antes del uso. Luego el suero se diluye hasta 25 % con medio de cultivo de tejido RPMI y se usa para analizar la estabilidad del péptido. En intervalos de tiempo predeterminados, una pequeña cantidad de solución de reacción se retira y agrega a etanol o ácido tricloroacético acuoso al 6 %. La muestra de reacción turbia se enfría (4 grados C) durante 15 minutos y luego se centrifuga para sedimentar las proteínas séricas precipitadas. Luego se determina la presencia de los péptidos por HPLC de fase inversa usando condiciones de cromatografía específicas de la estabilidad.

[00349] Los péptidos y polipéptidos pueden modificarse para proporcionar atributos deseados que no sean semivida en suero. Por ejemplo, la capacidad de los péptidos de inducir la actividad de CTL puede potenciarse por enlace a una secuencia que contiene al menos un epítopo que es capaz de inducir una respuesta de linfocitos T auxiliares. Los conjugados de péptidos inmunógenos/linfocitos T auxiliares pueden enlazarse por una molécula espaciadora. El espaciador típicamente comprende moléculas neutras relativamente pequeñas, como aminoácidos o miméticos de aminoácidos, que están sustancialmente no cargados en condiciones fisiológicas. Los espaciadores se seleccionan típicamente de, p. ej., Ala, Gly u otros espaciadores neutros de aminoácidos no polares o aminoácidos polares neutros. Se comprenderá que el espaciador opcionalmente presente no debe comprender los mismos residuos y por lo tanto puede ser un hetero- u homo-oligómero. Cuando está presente, el espaciador por lo general tendrá al menos uno o dos residuos, más generalmente tres a seis residuos. De manera alternativa, el péptido puede estar enlazado al péptido T auxiliar sin un espaciador.

[00350] Un péptido neoantigénico puede estar enlazado al péptido T auxiliar ya sea directamente o mediante un espaciador en el extremo amino o carboxi del péptido. El extremo amino del péptido neoantigénico o el péptido T auxiliar puede estar acilado. Péptidos T auxiliares ilustrativos incluyen

toxoide tetánico 830-843, gripe 307-319, malaria circumsporozoite 382-398 y 378-389.

[00351] Las proteínas o péptidos pueden elaborarse mediante cualquier técnica conocida por los expertos en la técnica, lo cual incluye la expresión de proteínas, polipéptidos o péptidos a través de técnicas de biología molecular estándares, el aislamiento de proteínas o péptidos de fuentes naturales o la síntesis química de proteínas o péptidos. Anteriormente, se han descrito secuencias de nucleótidos y proteínas, polipéptidos y péptidos correspondientes a diversos genes, que pueden encontrarse en bases de datos computarizadas conocidas por los expertos en la técnica. Una de dichas bases de datos es la base de datos Genbank y GenPept del National Center for Biotechnology Information [Centro Nacional para la Información Biotecnológica] ubicada en la página web de National Institutes of Health [Institutos Nacionales de Salud]. Las regiones codificantes para genes conocidos pueden amplificarse y/o expresarse mediante el uso de las técnicas descritas en la presente o tal como sabrán los expertos en la técnica. De manera alternativa, los expertos en la técnica conocen diversas preparaciones comerciales de proteínas, polipéptidos y péptidos.

[00352] En un aspecto adicional, un neoantígeno incluye un ácido nucleico (p. ej., polinucleótido) que codifica un péptido neoantigénico o parte de este. El polinucleótido puede ser, p. ej., ADN, ADNc, ANP, ANC, ARN (p. ej., ARNm), ya sea de cadena simple y/o doble, o formas naturales o estabilizadas de polinucleótidos, como, p. ej., polinucleótidos con una estructura principal de fosforotioato, o combinaciones de estos y puede contener o no intrones. Un aspecto aún adicional proporciona un vector de expresión que puede expresar un polipéptido o parte de este. Los vectores de expresión para diferentes tipos de células son conocidos en la técnica y pueden seleccionarse sin experimentación innecesaria. Generalmente, el ADN se inserta en un vector de expresión, como un plásmido, en orientación adecuada y marco de lectura correcto para la expresión. Si es necesario, el ADN puede unirse a las secuencias de nucleótidos de control reguladoras de traducción y de transcripción adecuadas reconocidas por el hospedador deseado, aunque dichos controles están generalmente disponibles en el vector de expresión. Luego el vector se introduce en el hospedador mediante técnicas estándar. Se puede encontrar una guía p. ej., en Sambrook et ál. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

V. Composiciones de vacuna

[00353] En la presente también se describe una composición inmunógena, p. ej., una composición de vacuna capaz de provocar una respuesta inmunitaria específica, p. ej., una respuesta inmunitaria específica de tumores. Las composiciones de vacuna típicamente comprenden múltiples neoantígenos, p. ej., que se seleccionan usando un método descrito en la presente. Las composiciones de vacuna también pueden denominarse vacunas.

[00354] Una vacuna puede contener entre 1 y 30 péptidos, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15,

16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 péptidos diferentes, 6, 7, 8, 9, 10 11, 12, 13 o 14 péptidos diferentes o 12, 13 o 14 péptidos diferentes. Los péptidos pueden incluir modificaciones postraduccionales. Una vacuna puede contener entre 1 y 100 o más secuencias de nucleótidos, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 o más secuencias de nucleótidos diferentes, 6, 7, 8, 9, 10 11, 12, 13 o 14 secuencias de nucleótidos diferentes o 12, 13 o 14 secuencias de nucleótidos diferentes. Una vacuna puede contener entre 1 y 30 o más secuencias de neoantígenos, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100 o más secuencias de neoantígenos diferentes, 6, 7, 8, 9, 10 11, 12, 13 o 14 secuencias de neoantígenos diferentes o 12, 13 o 14 secuencias de neoantígenos diferentes.

[00355] En una modalidad, se seleccionan diferentes péptidos y/o polipéptidos o secuencias de nucleótidos que los codifican de manera que los péptidos y/o polipéptidos pueden asociarse con diferentes moléculas de MHC, como moléculas del MHC clase I y/o clase II diferente. En algunos aspectos, una composición de vacuna comprende una secuencia codificante de péptidos y/o polipéptidos que pueden asociarse con las moléculas del MHC clase I y/o clase II que se producen más frecuentemente. Por lo tanto, las composiciones de vacuna pueden comprender diferentes fragmentos que pueden asociarse con al menos 2 moléculas del MHC clase I y/clase II diferentes preferidas, al menos 3 preferidas o al menos 4 preferidas.

[00356] La composición de vacuna puede provocar una respuesta de linfocitos T citotóxicos específica y/o una respuesta de linfocitos T auxiliares específica.

[00357] Una composición de vacuna también puede comprender un adyuvante y/o un portador. Ejemplos de adyuvantes y portadores útiles se proporcionan a continuación en la presente. Una composición puede asociarse con un portador como, p. ej., una proteína o célula presentadora de antígeno como, p. ej., una célula dendrítica (DC) que puede presentar el péptido a un linfocito T.

[00358] Los adyuvantes son cualquier sustancia cuya mezcla en una composición de vacuna aumenta o modifica de otro modo la respuesta inmunitaria a un neoantígeno. Los portadores pueden ser estructuras de andamiaje, por ejemplo, un polipéptido o un polisacárido, a las cuales puede asociarse un neoantígeno. Opcionalmente, los adyuvantes están conjugados de manera covalente o no covalente.

[00359] La capacidad de un adyuvante de aumentar una respuesta inmunitaria a un antígeno se

manifiesta típicamente mediante un aumento significativo o sustancial en una reacción mediada por el sistema inmunitario o una reducción de los síntomas de la enfermedad. Por ejemplo, un aumento de la inmunidad humoral se manifiesta típicamente mediante un aumento significativo en la titulación de anticuerpos presentados al antígeno, y un aumento de la actividad de linfocitos T se manifiesta típicamente mediante mayor proliferación celular, o citotoxicidad celular o secreción de citocinas. Un adyuvante también puede alterar una respuesta inmunitaria, por ejemplo, al cambiar una respuesta principalmente humoral o de Th a una respuesta principalmente celular o de Th.

[00360] Los adyuvantes adecuados incluyen, de modo no taxativo, 1018 ISS, alumbre, sales de aluminio, Amplivax, AS15, BCG, CP-870,893, CpG7909, CyaA, dSLIM, GM-CSF, IC30, IC31, Imiquimod, ImuFact IMP321, IS Patch, ISS, ISCOMATRIX, Juvlmmune, LipoVac, MF59, monofosforil lípido A, Montanide IMS 1312, Montanide ISA 206, Montanide ISA 50V, Montanide ISA-51, OK-432, OM-174, OM-197-MP-EC, ONTAK, sistema de vectores PepTel, micropartículas de PLG, resiquimod, SRL172, Virosomas y otras partículas tipo Virus, YF-17D, trampa de VEGF, R848, beta-glucano, Pam3Cys, QS21 stimulon de Aquila (Aquila Biotech, Worcester, Mass., EUA) que deriva de saponina, extractos micobacterianos y miméticos sintéticos de pared celular bacteriana y otros adyuvantes patentados como Detox de Ribi. Quil o Superfos de Ribi. Adyuvantes como el de Freund incompleto o GM-CSF son útiles. Previamente se han descrito varios adyuvantes inmunológicos (por ejemplo, MF59) específicos para las células dendríticas y su preparación (Dupuis M, et ál., Cell Immunol. 1998; 186(1):18-27; Allison A C; Dev Biol Stand. 1998; 92:3-11). También se pueden utilizar citocinas. Varias citocinas se han relacionado directamente con la influencia de la migración de células dendríticas hacia tejidos linfoides (p. ej., TNF-alfa), acelerando la maduración de células dendríticas en células que presentan antígenos eficaces para linfocitos T (por ejemplo, GM- CSF, IL-1 e IL-4) (patente estadounidense n.° 5.849.589, la cual se incorpora específicamente a la presente en su totalidad mediante esta referencia) y actuando como inmunoadyuvantes (p. ej., IL-12) (Gabrilovich D I, et ál., J Immunother Emphasis Tumor Immunol. 1996 (6):414-418).

[00361] También se ha informado que los oligonucleótidos inmunoestimulantes de CpG mejoran los efectos de adyuvantes en un entorno de vacuna. También se pueden utilizar otras moléculas de unión de TLR como TLR 7, TLR 8 y/o TLR 9 que se unen a ARN.

[00362] Otros ejemplos de adyuvantes útiles incluyen, de modo no taxativo, CpG modificados químicamente (p. ej., CpR, Idera), Poly(I:C)(p. ej., polyi:CI2U), ADN o ARN bacteriano diferente a CpG así como también anticuerpos y moléculas pequeñas inmunoactivas como ciclofosfamida, sunitinib, bevacizumab, celebrex, NCX-4016, sildenafil, tadalafil, vardenafil, sorafinib, XL-999, CP-547632, pazopanib, ZD2171, AZD2171, ipilimumab, tremelimumab y SC58175, que pueden actuar terapéuticamente y/o como un adyuvante. El experto en la técnica puede determinar fácilmente las

cantidades y concentraciones de adyuvantes y aditivos útiles sin experimentación innecesaria. Los adyuvantes adicionales incluyen factores estimulantes de colonias, como el factor estimulante de colonias de granulocitos y macrófagos (GM-CSF, sargramostim).

[00363] Una composición de vacuna puede comprender más de un adyuvante diferente. Además, una composición terapéutica puede comprender cualquier sustancia adyuvante que incluye cualquiera de las antes mencionadas o combinaciones de estas. También se contempla que una vacuna y un adyuvante pueden administrarse juntos o separados en cualquier secuencia apropiada.

[00364] Un portador (o excipiente) puede estar presente independientemente de un adyuvante. La función de un portador puede ser, por ejemplo, aumentar el peso molecular de un mutante particular para aumentar la actividad o inmunogenicidad, para conferir estabilidad, para aumentar la actividad biológica o para aumentar la semivida en suero. Además, un portador puede ayudar a presentar péptidos a linfocitos T. Un portador puede ser cualquier portador adecuado conocido por el experto en la técnica, por ejemplo una proteína o una célula presentadora de antígenos. Una proteína portadora podría ser pero no se limita a hemocianina de lapa californiana, proteínas séricas como transferrina, albúmina sérica bovina, albúmina sérica humana, tiroglobulina u ovalbúmina, inmunoglobulinas u hormonas, como insulina o ácido palmítico. Para la inmunización de seres humanos, el portador es generalmente un portador aceptable desde el punto de vista fisiológico para los seres humanos y seguro. Sin embargo, el toxoide tetánico y/o toxoide diftérico son portadores adecuados. De manera alternativa, el portador puede ser dextranos por ejemplo sefarosa.

[00365] Los linfocitos T citotóxicos (CTL, por sus siglas en inglés) reconocen un antígeno en la forma de un péptido unido a una molécula de MHC más que el antígeno foráneo intacto en sí mismo. La molécula de MHC en sí misma está ubicada en la superficie celular de una célula que presenta antígenos. Por lo tanto, una activación de CTL es posible si hay un complejo trimérico de antígeno de péptidos, molécula de MHC y APC presente. Por consiguiente, puede potenciar la respuesta inmunitaria si no solo el péptido se usa para la activación de CTL pero si además se agregan APC con la molécula de MHC correspondiente. Por lo tanto, en algunas modalidades, una composición de vacuna contiene además al menos una célula presentadora de antígenos.

[00366] Los neoantígenos también se pueden incluir en plataformas de vacuna basadas en vectores virales, como vaccinia, viruela aviar, alfavirus de autorreplicación, marabavirus, adenovirus (remitirse, p. ej., a Tatsis et ál., Adenoviruses, *Molecular Therapy* (2004) 10, 616—629) o lentivirus, incluidos, de modo no taxativo, lentivirus de segunda o tercera generación o híbridos de segunda y tercera generación y lentivirus recombinantes de cualquier generación diseñados para dirigirse a receptores o tipos de células específicos (remitirse, p. ej., a Hu et ál., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, *Immunol Rev.* (2011) 239(1): 45-61, Sakuma et ál., Lentiviral vectors: basic to

translational, *Biochem J.* (2012) 443(3):603-18, Cooper et ál., Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, *Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey et ál., Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, *J. Virol.* (1998) 72 (12): 9873-9880). Dependiendo de la capacidad de envasado de las plataformas de vacunas basadas en vectores virales antedichas, este enfoque puede suministrar una o más secuencias de nucleótidos que codifiquen uno o más péptidos neoantigénicos. Las secuencias pueden estar flanqueadas por secuencias sin mutación, pueden estar separadas por enlazadores o pueden estar precedidas por una o más secuencias dirigidas al compartimiento subcelular (remitirse, p. ej., a Gros et ál., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen et ál., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu et ál., Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(13):3401-10). Tras la introducción en un hospedador, las células infectadas expresan los neoantígenos y, de esa forma, provocan una respuesta inmunitaria del hospedador (p. ej., CTL) contra el o los péptidos. Los vectores de vaccinia y métodos útiles para los protocolos de inmunización se encuentran descritos, p. ej., en la patente estadounidense n.º 4.722.848. Otro vector es BCG (Bacillus Calmette-Guerin). Los vectores de BCG se encuentran descritos en Stover et ál. (Nature 351:456-460 (1991)). Los expertos en la técnica podrán observar una amplia variedad de vectores de vacunas adicionales, útiles para la inmunización o administración terapéutica de los neoantígenos, p. ej., vectores de Salmonella typhi y similares, a partir de la descripción en la presente.

V.A. Casete del neoantígeno

[00367] Los métodos utilizados para la selección de uno o más neoantígenos, la clonación y construcción de un “casete” y su inserción en un vector viral son parte del conocimiento del experto dadas las instrucciones proporcionadas en la presente. «Casete del neoantígeno» se refiere a la combinación de un neoantígeno seleccionado o múltiples neoantígenos y los otros elementos reguladores necesarios para transcribir el o los neoantígenos y expresar el producto transcrito. Un neoantígeno o múltiples neoantígenos pueden enlazarse operativamente a componentes reguladores en una forma que permita la transcripción. Estos componentes incluyen elementos reguladores convencionales que pueden impulsar la expresión del o los neoantígenos en una célula transfectada con el vector viral. Por lo tanto, el casete del neoantígeno también puede contener un promotor seleccionado que está enlazado al o los neoantígenos y ubicado con otros elementos reguladores opcionales, dentro de las secuencias virales seleccionadas del vector recombinante.

[00368] Los promotores útiles pueden ser promotores constitutivos o promotores regulados (inducibles), que permitirán el control de la cantidad del o los neoantígenos que se deben expresar. Por

ejemplo, un promotor deseable es el del promotor/potenciador temprano inmediato del citomegalovirus [remitirse, p. ej., a Boshart et ál, Cell, 41:521-530 (1985)]. Otro promotor deseable incluye el promotor/potenciador de LTR del virus del sarcoma de Rous. Incluso otra secuencia promotora/potenciadora es el promotor de beta-actina citoplásmica de la gallina [T. A. Kost et ál, Nucl. Acids Res., 11(23):8287 (1983)]. Un experto en la técnica puede seleccionar otros promotores deseables o adecuados.

[00369] El casete del neoantígeno también puede incluir secuencias de ácido nucleico heterólogas a las secuencias de vectores virales que incluyen secuencias que proporcionan señales para la poliadenilación eficaz de la transcripción (poli(A), poli-A o pA) e intrones con sitios aceptores y donantes de empalme funcionales. Una secuencia poly-A común que se emplea en los vectores ilustrativos de la presente invención es la que deriva del papovavirus SV-40. La secuencia poly-A generalmente puede insertarse en el casete después de las secuencias basadas en neoantígeno y antes de las secuencias de vectores virales. Una secuencia de intrones común también puede derivarse de SV-40 y se denomina secuencia de intrones T SV-40. Un casete del neoantígeno también puede contener tal intrón, ubicado entre la secuencia promotora/potenciadora y uno o más neoantígenos. La selección de estos y otros elementos de vectores comunes es convencional [remitirse, p. ej., a Sambrook et ál., “Molecular Cloning: A Laboratory Manual.", 2a edición, Cold Spring Harbor Laboratory, Nueva York (1989) y referencias citadas allí] y muchas de estas secuencias están disponibles de fuentes comerciales e industriales así como Genbank.

[00370] Un casete del neoantígeno puede tener uno o más neoantígenos. Por ejemplo, un casete dado puede incluir 1-10, 1-20, 1-30, 10-20, 15-25, 15-20, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más neoantígenos. Los neoantígenos pueden enlazarse directamente unos con otros. Los neoantígenos también pueden enlazarse unos con otros con enlazadores. Los neoantígenos pueden estar en cualquier orientación unos respecto de otros que incluye N a C o C a N.

[00371] Como se establece anteriormente, el casete del neoantígeno puede estar ubicado en el sitio de cualquier eliminación seleccionada en el vector viral, como el sitio de la eliminación de la región del gen E1 o eliminación de la región del gen E3, entre otros que pueden seleccionarse.

[00372] El casete de neoantígeno puede describirse mediante el uso de la siguiente fórmula para describir la secuencia ordenada de cada elemento, de 5’ a 3’:

$$(P_a\text{-}(L5_b\text{-}N_c\text{-}L3_d)_X)_Z\text{-}(P2_h\text{-}(G5_e\text{-}U_f)_Y)_W\text{-}G3_g$$

[00373] donde P y P2 comprenden las secuencias promotora de nucleótidos, N comprende una secuencia de ácido nucleico que codifica el epítopo del MHC clase I, L5 comprende una secuencia enlazadora 5’, L3 comprende una secuencia enlazadora 3’, G5 comprende una secuencia de ácido nucleico que codifica una secuencia enlazadora de aminoácidos, G3 comprende una secuencia de ácido

nucleico que codifica una secuencia enlazadora de aminoácidos, U comprende una secuencia de ácido nucleico que codifica el antígeno del MHC clase II, donde para cada X el Nc correspondiente es una secuencia de ácido nucleico que codifica el epítopo, donde para cada Y el Uf correspondiente es una secuencia de ácido nucleico que codifica el antígeno. La composición y secuencia ordenada puede definirse adicionalmente mediante la selección de elementos presentes, por ejemplo, donde a = 0 o 1, donde b = 0 o 1, donde c = 1, donde d = 0 o 1, donde e = 0 o 1, donde f = 1, donde g = 0 o 1, donde h = 0 o 1, X = 1 a 400, Y = 0, 1, 2, 3, 4 o 5, Z = 1 a 400, y W = 0, 1, 2, 3, 4 o 5.

[00374] En un ejemplo, los elementos presentados incluyen a = 0, b = 1, d = 1, e = 1, g = 1, h = 0, X = 10, Y = 2, Z = 1, y W = 1, que describen que no hay promotor adicional presente (es decir, solamente la secuencia promotora de nucleótidos proporcionada por la estructura principal de alfavirus de ARN está presente), están presentes 20 epítopos del MHC clase I, está presente un enlazador 5' para cada N, está presente un enlazador 3' para cada N, están presentes 2 epítopos del MHC clase II, está presente un enlazador que une los dos epítopos del MHC clase II, está presente un enlazador que une el extremo 5' de los dos epítopos de MHC clase II con el enlazador 3' del epítopo de MHC clase I final, y está presente un enlazador que une el extremo 3' de los dos epítopos del MHC clase II con la estructura principal de alfavirus de ARN. Los ejemplos de unión del extremo 3' del casete de neoantígeno con la estructura principal de alfavirus de ARN incluye la unión directa a los elementos de UTR 3' proporcionados por la estructura principal de alfavirus de ARN, tal como CSE 19-nt 3'. Los ejemplos de unión del extremo 5' del casete de neoantígeno con la estructura principal de alfavirus de ARN incluye la unión directa a una secuencia promotora 26S, una UTR 5' de alfavirus, un CSE 51-nt o un CSE 24-nt.

[00375] Otros ejemplos incluyen: donde a = 1 describe donde está presente un promotor que no es la secuencia promotora de nucleótidos proporcionada por la estructura principal de alfavirus de ARN; donde a = 1 y Z es mayor que 1 donde están presentes múltiples promotores que no son la secuencia promotora de nucleótidos proporcionada por la estructura principal de alfavirus de ARN, donde cada uno impulsa la expresión de 1 o más secuencias de ácido nucleico que codifican el epítopo del MHC clase I distintas; donde h = 1 describe donde está presente un promotor separado para impulsar la expresión de las secuencias de ácido nucleico que codifican el antígeno del MHC clase II; y donde g = 0 describe que la secuencia de ácido nucleico que codifica el antígeno del MHC clase II, si está presente, se une directamente con la estructura principal de alfavirus de ARN.

[00376] Otros ejemplos incluyen donde cada epítopo del MHC clase I que está presente puede tener un enlazador 5', un enlazador 3', ninguno o ambos. En ejemplos donde hay presente más de un epítopo del MHC clase I en el mismo casete de neoantígeno, algunos epítopos del MHC clase I pueden tener tanto un enlazador 5' como un enlazador 3', mientras que otros epítopos del MHC clase I tienen ya sea un enlazador 5', un enlazador 3' o ninguno. En otros ejemplos donde hay presente más de un epítopo

del MHC clase I en el mismo casete de neoantígeno, algunos epítopos del MHC clase I pueden tener ya sea un enlazador 5' o un enlazador 3', mientras que otros epítopos del MHC clase I tienen ya sea un enlazador 5', un enlazador 3' o ninguno.

[00377] En ejemplos donde hay presente más de un epítopo del MHC clase II en el mismo casete de neoantígeno, algunos epítopos del MHC clase II pueden tener tanto un enlazador 5' como un enlazador 3', mientras que otros epítopos del MHC clase II tienen ya sea un enlazador 5', un enlazador 3' o ninguno. En otros ejemplos donde hay presente más de un epítopo del MHC clase II en el mismo casete de neoantígeno, algunos epítopos del MHC clase II pueden tener ya sea un enlazador 5' o un enlazador 3', mientras que otros epítopos del MHC clase II tienen ya sea un enlazador 5', un enlazador 3' o ninguno.

[00378] Las secuencias promotoras de nucleótidos P y/o P2 pueden ser la misma que una secuencia promotora de nucleótidos proporcionada por la estructura principal de alfavirus de ARN. Por ejemplo, cada una de la secuencia promotora proporcionada por la estructura principal de alfavirus de ARN, Pn y P2, puede comprender un promotor subgenómico 26S. Las secuencias promotoras de nucleótidos P y/o P2 pueden ser diferentes de la secuencia de nucleótidos proporcionada por la estructura principal de alfavirus de ARN, así como pueden ser diferentes entre sí.

[00379] El enlazador 5' L5 puede ser una secuencia natural o una secuencia no natural. La secuencia no natural incluye, de modo no taxativo, AAY, RR y DPP. El enlazador 3' L3 también puede ser una secuencia natural o una secuencia no natural. De manera adicional, L5 y L3 pueden ser ambos secuencias naturales, ambos secuencias no naturales o uno puede ser natural y el otro no natural. Para cada X, los enlazadores de aminoácidos pueden tener una longitud de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94,95, 96, 97, 98, 99, 100 o más aminoácidos. Para cada X, los enlazadores de aminoácidos también pueden tener una longitud de al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29 o al menos 30 aminoácidos.

[00380] Los enlazadores de aminoácido G5, para cada Y, pueden tener una longitud de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94,95, 96, 97, 98, 99, 100 o más aminoácidos. Para cada Y, los enlazadores de aminoácidos también pueden tener una longitud de al menos 3, al menos 4, al menos 5, al menos 6, al

menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29 o al menos 30 aminoácidos.

[00381] Los enlazadores de aminoácido G3 pueden tener una longitud de 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94,95, 96, 97, 98, 99, 100 o más aminoácidos. G3 también puede tener una longitud de al menos 3, al menos 4, al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29 o al menos 30 aminoácidos.

[00382] Para cada X, cada N puede codificar un epítopo del MHC clase I con una longitud de 7-15 aminoácidos. Para cada X, cada N también codifica un epítopo del MHC clase I que tiene 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 o 30 aminoácidos de longitud. Para cada X, cada N también codifica un epítopo del MHC clase I que tiene al menos 5, al menos 6, al menos 7, al menos 8, al menos 9, al menos 10, al menos 11, al menos 12, al menos 13, al menos 14, al menos 15, al menos 16, al menos 17, al menos 18, al menos 19, al menos 20, al menos 21, al menos 22, al menos 23, al menos 24, al menos 25, al menos 26, al menos 27, al menos 28, al menos 29 o al menos 30 aminoácidos.

V.B. Puntos de control inmunitarios

[00383] Los vectores descritos en la presente, como vectores C68 descritos en la presente o vectores de alfavirus descritos en la presente, pueden comprender un ácido nucleico que codifica al menos un neoantígeno y el mismo vector o uno diferente puede comprender un ácido nucleico que codifica al menos un inmunomodulador (p. ej., un anticuerpo como scFv) que se une a una molécula de punto de control inmunitario y bloquea su actividad. Los vectores pueden comprender un casete del neoantígeno y una o más moléculas de ácido nucleico que codifican un inhibidor de punto de control.

[00384] Las moléculas de punto de control inmunitario ilustrativas que pueden fijarse como objetivo para el bloqueo o la inhibición incluyen, de modo no taxativo,CTLA-4, 4-1BB (CD137), 4-1BBL (CD137L), PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4 (pertenece a la familia CD2 de moléculas y se expresa en todos los NK, $\gamma\delta$,y linfocitos T CD8+ ($\alpha\beta$)de memoria, CD160 (también denominado BY55) y CGEN-15049. Los inhibidores de puntos de control inmunitarios incluyen anticuerpos o fragmentos de unión a antígenos de estos u otras proteínas

de unión que se unen a uno o más de CTLA-4, PDL1, PDL2, PD1, B7-H3, B7-H4, BTLA, HVEM, TIM3, GAL9, LAG3, TIM3, B7H3, B7H4, VISTA, KIR, 2B4, CD160 y CGEN-15049 y bloquean o inhiben su actividad. Los inhibidores de puntos de control inmunitarios ilustrativos incluyen Tremelimumab (anticuerpo de bloqueo CTLA-4), anti-OX40, anticuerpo monoclonal PD-L1 (Anti-B7-H1; MEDI4736), ipilimumab, MK-3475 (bloqueador PD-1), Nivolumab (anticuerpo anti-PD1), CT-011 (anticuerpo anti-PD1), anticuerpo monoclonal BY55, AMP224 (anticuerpo anti-PDL1), BMS-936559 (anticuerpo anti-PDL1), MPLDL3280A (anticuerpo anti-PDL1), MSB0010718C (anticuerpo anti-PDL1) y Yervoy/ipilimumab (inhibidor de puntos de control anti-CTLA-4). Las secuencias que codifican anticuerpos pueden modificarse genéticamente en vectores como C68 usando la experiencia normal en la técnica. Un método ilustrativo se describe en Fang et ál., Stable antibody expression at therapeutic levels using the 2A peptide. *Nat Biotechnol.* mayo de 2005; 23(5):584-90. Publicación electrónica, 17 de abril de 2005; que se incorpora en la presente mediante esta referencia a todos los efectos.

V.C. Consideraciones adicionales para el diseño y la elaboración de vacunas

V.C.1. Determinación de un conjunto de péptidos que cubren todos los subclones tumorales

[00385] Los péptidos truncos, es decir los presentados por todos o la mayoría de los subclones tumorales, pueden priorizarse para inclurise en la vacuna.[53] Opcionalmente, si no hay péptidos truncos que se predice que se presenten y sean inmunógenos con alta probabilidad, o si la cantidad de péptidos truncos que se predice que se presentará y será inmunógena con alta proabilidad es lo suficientemente pequeña que se pueden incluir péptidos no truncos adicionales en la vacuna, entonces los péptidos adicionales pueden priorizarse al estimar la cantidad e identidad de subclones tumorales y eligiendo péptidos como para maximizar la cantidad de subclones tumorales cubiertos por la vacuna.[54]

V.C.2. Prioridad de neoantígenos

[00386] Luego de aplicar todos los filtros de neoantígenos que anteceden, más neoantígenos candidatos pueden estar disponibles para incluirlos en la vacuna de los que la tecnología de vacunas puede soportar. Además, puede haber incertidumbre sobre varios aspectos del análisis de neoantígenos y pueden existir soluciones intermedias entre diferentes propiedades de los neoantígenos de vacuna candidatos. Por lo tanto, en vez de filtros predeterminados en cada etapa del proceso de selección, un modelo multidimensional integrado puede tenerse en cuenta que coloca a los neoantígenos candidatos en un espacio con al menos los siguientes ejes y optimiza la selección usando un enfoque integrador.

1. Riesgo de autoinmunidad o tolerancia (riesgo de línea germinal) (típicamente se prefiere un menor riesgo de autoinmunidad)
2. Probabilidad de defecto de secuenciación (típicamente se prefiere una menor probabilidad de defectos)
3. Probabilidad de inmunogenicidad (típicamente se prefiere una mayor probabilidad de

inmunogenicidad)

4. Probabilidad de presentación (típicamente se prefiere una mayor probabilidad de presentación)

5. Expresión génica (típicamente se prefiere una mayor expresión)

6. Cobertura de genes de HLA (una gran cantidad de moléculas de HLA implicadas en la presentación de un conjunto de neoantígenos puede reducir la probabilidad de que un tumor escape del ataque inmunitario mediante la regulación por disminución o la mutación de moléculas de HLA)

7. Cobertura de las clases de HLA (la cobertura tanto de HLA-I como de HLA-II puede aumentar la probabilidad de respuesta terapéutica y disminuir la probabilidad de escape tumoral).

[00387] Además, de manera opcional, los neoantígenos pueden despriorizarse (por ejemplo, excluirse) de la vacuna si se prevé que los presenten alelos de HLA perdidos o inactivados en la totalidad o parte del tumor del paciente. La pérdida de los alelos de HLA puede producirse ya sea mediante mutación somática, pérdida de heterocigosidad o deleción homocigota del locus. Se conocen en la técnica los métodos para la detección de la mutación somática de alelos de HLA, por ejemplo (Shukla et ál., 2015). Se describen también los métodos para la detección de LOH somática y deleción homocigota (incluso para un locus de HLA). (Carter et ál., 2012; McGranahan et ál., 2017; Van Loo et ál., 2010).

V.D. Alfavirus

V.D.1. Biología del alfavirus

[00388] Los alfavirus son miembros de la familia *Togaviridae*, y son virus de ARN monocatenario de sentido positivo. Los alfavirus también pueden denominarse ARN de autorreplicación o srRNA. Los miembros se clasifican típicamente como del viejo mundo, tales como virus Sindbis, río Ross, Mayaro, Chikungunya y bosque Semliki o como del nuevo mundo, tales como encefalitis equina del este, Aura, Fort Morgan, encefalitis equina venezolana y su cepa derivada TC-83 (Strauss Microbrial Review 1994). Un genoma de alfavirus natural tiene típicamente alrededor de 12 kb de longitud, cuyos primeros dos tercios contienen los genes que codifican proteínas no estructurales (nsP, por sus siglas en inglés) que forman complejos de replicación de ARN para la autorreplicación del genoma viral y cuyo último tercio contiene un casete de expresión subgenómica que codifica proteínas estructurales para la producción de viriones (Frolov RNA 2001).

[00389] Un ciclo de vida modelo de un alfavirus implica varias etapas diferentes (Strauss Microbrial Review 1994, Jose Future Microbiol 2009). Después de la unión del virus a una célula hospedadora, el virión se fusiona con membranas dentro de compartimientos endocíticos que resultan en la eventual liberación del ARN genómico en el citosol. El ARN genómico, que está en una orientación de cadena positiva y comprende un casquete 5’ metilguanilato y una cola 3’ polyA, se traduce para producir

proteínas no estructurales nsP1-4 que forman el complejo de replicación. Al principio de la infección, la cadena positiva se replica por el complejo en una plantilla de cadena negativa. En el modelo actual, el complejo de replicación se procesa adicionalmente a medida que avanza la infección y el complejo procesado resultante cambia a una transcripción de la cadena negativa en ARN genómico de cadena positiva de longitud completa, así como ARN subgenómico de cadena positiva 26S que contiene los genes estructurales. Se han identificado varios elementos de secuencia conservados (CSE, por sus siglas en inglés) de alfavirus para potencialmente cumplir una función en las diversas etapas de la replicación de ARN que incluyen: un complemento de la UTR 5' en la replicación de ARN de cadena positiva de una plantilla de cadena negativa, un CSE de 51 nucleótidos en la replicación de la síntesis de cadena negativa de la plantilla genómica, un CSE de 24 nucleótidos en la región de unión entre las nsP y el ARN 26S en la transcripción de ARN subgenómico de la cadena negativa y un CSE 3' de 19 nucleótidos en la síntesis de cadena negativa de la plantilla de cadena positiva.

[00390] Después de la replicación de las diversas especies de ARN, las partículas de virus típicamente se ensamblan en el ciclo de vida natural del virus. El ARN 26S se traduce y las proteínas resultantes se procesan adicionalmente para producir proteínas estructurales que incluyen proteína de la cápside, glicoproteínas E1 y E2 y dos pequeños polipéptidos E3 y 6K (Strauss 1994). La encapsidación de ARN viral se produce, con las proteínas de la cápside normalmente específicas solo para el ARN genómico envasado, después del montaje de viriones y gemación en la superficie de la membrana.

V.D.2. Alfavirus como vector de suministro

[00391] Los alfavirus se han modificado genéticamente previamente para usarse como sistemas de vector de expresión (Pushko 1997, Rheme 2004). Los alfavirus ofrecen varias ventajas, particularmente en un entorno de vacuna donde se puede desear la expresión de antígeno heterólogo. Debido a su capacidad de autorreplicarse en el citosol del hospedador, los vectores de alfavirus generalmente pueden producir altos números de copia del casete de expresión dentro de una célula que resulta en un alto nivel de producción de antígeno heterólogo. Además, los vectores son generalmente transitorios, dando como resultado una mejor bioseguridad así como una menor inducción de tolerancia inmunológica al vector. El público, en general, también carece de la inmunidad preexistente a los vectores de alfavirus en comparación con otros vectores virales estándar, como adenovirus humano. Los vectores basados en alfavirus también generalmente resultan en respuestas citotóxicas a las células infectadas. La citotoxicidad, en cierta medida, puede ser importante en un entorno de vacuna para provocar de manera adecuada una respuesta inmunitaria al antígeno heterólogo expresado. Sin embargo, el grado de citotoxicidad deseada puede ser un acto de equilibrio y por lo tanto se han desarrollado varios alfavirus atenuados, inclusive la cepa TC-83 de VEE. Por lo tanto, un ejemplo de un vector de expresión de

neoantígeno descrito en la presente puede utilizar una estructura principal de alfavirus que permite un alto nivel de expresión de neoantígeno, provoca una respuesta inmunitaria sólida a un neoantígeno, no provoca una respuesta inmunitaria al vector en sí mismo y puede usarse de manera segura. Además, el casete de expresión de neoantígeno puede diseñarse para provocar diferentes niveles de una respuesta inmunitaria mediante la optimización de las secuencias de alfavirus que usa el vector inclusive, de modo no taxativo, secuencias obtenidas de VEE o su derivado atenuado TC-83.

[00392] Se han diseñado varias estrategias de diseño del vector de expresión que usan secuencias de alfavirus (Pushko 1997). En una estrategia, un diseño de vector de alfavirus incluye insertar una segunda copia de los elementos de secuencia promotora 26S corriente abajo de los genes de proteína estructural, seguido de un gen heterólogo (Frolov 1993). Por lo tanto, además de las proteínas estructurales y no estructurales naturales, se produce un ARN subgenómico adicional que expresa la proteína heteróloga. En este sistema, están presentes todos los elementos para la producción de viriones infecciosos y, por lo tanto, pueden producirse rondas repetidas de infección del vector de expresión en células no infectadas.

[00393] Otro diseño de vector de expresión utiliza los sistemas de virus auxiliar (Pushko 1997). En esta estrategia, las proteínas estructurales son remplazadas por un gen heterólogo. Por lo tanto, después de la autorreplicación de ARN viral mediada por genes no estructurales que todavía están intactos, el ARN subgenómico 26S provee la expresión de la proteína heteróloga. Tradicionalmente, se suministran vectores adicionales que expresan las proteínas estructurales *in trans*, como mediante transfección conjunta de una línea celular, para producir virus infecciosos. En USPN 8.093.021, que se incorpora en la presente mediante esta referencia en su totalidad a todos los efectos, se describe un sistema en detalle. El sistema de vectores auxiliares proporciona el beneficio de limitar la posibilidad de formar partículas infecciosas y, por lo tanto, mejora la bioseguridad. Además, el sistema de vectores auxiliares reduce la longitud total del vector, potencialmente mejorando la eficacia de replicación y expresión. Por lo tanto, un ejemplo de un vector de expresión de neoantígeno descrito en la presente puede utilizar una estructura principal de alfavirus donde las proteínas estructurales se remplazan por un casete del neoantígeno, donde el vector resultante reduce las preocupaciones de bioseguridad a la vez que promueve la expresión eficaz debido a la reducción del tamaño del vector de expresión total.

V.D.3. Producción de alfavirus *in vitro*

[00394] Los vectores de suministro de alfavirus son generalmente polinucleótidos de ARN de sentido positivo. Un método conveniente conocido en la técnica para la producción de ARN es la transcripción *in vitro*, IVT. En esta técnica, una plantilla de ADN del vector deseado se produce en primer lugar por métodos conocidos por los expertos en la técnica, inclusive técnicas de biología molecular estándar como clonación, digestión por restricción, ligación, síntesis génica y reacción en cadena de la polimerasa

(PCR, por sus siglas en inglés). La plantilla de ADN contiene un promotor de polimerasa de ARN en el extremo 5' de la secuencia que se desea transcribir en ARN. Los promotores incluyen, de modo no taxativo, promotores de polimerasa de bacteriófagos como T3, T7 o SP6. Luego la plantilla de ADN se incuba con la enzima de polimerasa de ARN, agentes amortiguadores y nucleótidos (NTP) apropiados. El polinucleótido de ARN resultante puede opcionalmente modificarse de manera adicional, lo que incluye, de modo no taxativo, adición de una estructura de casquete 5' como 7-metilguanosina o una estructura relacionada y opcionalmente modificar el extremo 3' para que incluya una cola poliadenilada (polyA). Luego el ARN puede purificarse usando técnicas conocidas en el campo, como la extracción de fenol cloroformo.

V.D.4. Suministro mediante nanopartículas de lípidos

[00395] Un aspecto importante que se debe tener en cuenta en el diseño del vector de vacuna es la inmunidad contra el propio vector (Riley 2017). Esto puede presentarse en la forma de inmunidad preexistente al propio vector, como con determinados sistemas de adenovirus humanos, o en la forma de la generación de inmunidad al vector después de la administración de la vacuna. La última es una consideración importante si se realizan múltiples administraciones de la misma vacuna, como dosis de cebado y refuerzo separadas, o si el mismo sistema de vector de vacuna se usa para suministrar diferentes casetes del neoantígeno.

[00396] En el caso de vectores de alfavirus, el método de suministro estándar es el sistema de virus auxiliares descrito anteriormente que proporciona cápside, proteínas E1 y E2 *in trans* para producir partículas virales infecciosas. Sin embargo, cabe destacar que las proteínas E1 y E2 con frecuencia son los objetivos principales de los anticuerpos neutralizantes (Strauss 1994). Por lo tanto, la eficacia del uso de vectores de alfavirus para suministrar neoantígenos de interés a células objetivo puede reducirse si los anticuerpos neutralizantes se dirigen a partículas infecciosas.

[00397] Una alternativa al suministro de genes mediado por partículas es el uso de nanomateriales para suministrar vectores de expresión (Riley 2017). Cabe destacar que los vehículos de nanomaterial pueden estar hechos de materiales no inmunogénicos y generalmente evitan provocar inmunidad al propio vector de suministro. Estos materiales pueden incluir, de modo no taxativo, lípidos, nanomateriales inorgánicos y otros materiales poliméricos. Los lípidos pueden ser catiónicos, aniónicos o neutros. Los materiales pueden ser de origen natural o sintético, y en algunos casos biodegradables. Los lípidos pueden incluir grasas, colesterol, fosfolípidos, conjugados lipídicos que incluyen, de modo no taxativo, conjugados de polietilenglicol (PEG) (lípidos PEGilados), ceras, aceites, glicéridos y vitaminas solubles en grasas.

[00398] Las nanopartículas lipídicas (LNP, por sus siglas en inglés) son un sistema de suministro atractivo debido a la naturaleza anfifílica de los lípidos que permite la formación de membranas y

estructuras tipo vesículas (Riley 2017). En general, estos vehículos suministran el vector de expresión mediante su absorción en la membrana de células objetivo y liberación de ácido nucleico al citosol. Además, los LNP pueden modificarse o funcionalizarse adicionalmente para facilitar el direccionamiento de tipos celulares específicos. Otra consideración en el diseño de LNP es el equilibrio entre la eficacia del direccionamiento y la citotoxicidad. Las composiciones lipídicas generalmente incluyen mezclas definidas de lípidos catiónicos, neutros, aniónicos y anfipáticos. En algunos casos, se incluyen lípidos específicos para evitar la agregación de LNP, evitar la oxidación de lípidos o proporcionar grupos químicos funcionales que facilitan la unión de restos adicionales. Las composiciones lipídicas pueden incluir en el tamaño y la estabilidad total de LNP. En un ejemplo, la composición lipídica comprende dilinoleilmetil- 4-dimetilaminobutirato (MC3) o moléculas tipo MC3. Las composiciones lipídicas de MC3 y tipo MC3 pueden formularse para incluir uno o más lípidos diferentes, como PEG o lípidos conjugados con PEG, un esterol, o lípidos neutros.

[00399] Los vectores de ácido nucleico, como vectores de expresión, expuestos directamente a suero pueden tener varias consecuencias indeseables, inclusive la degradación del ácido nucleico por nucleasas séricas o la estimulación fuera del objetivo del sistema inmunitario por los ácidos nucleicos libres. Por lo tanto, la encapsulación del vector de alfavirus puede usarse para evitar la degradación, a la vez que también evita posibles efectos fuera del objetivo. En determinados ejemplos, un vector de alfavirus está completamente encapsulado en el vehículo de suministro, como dentro del interior acuoso de una LNP. El encapsulamiento del vector de alfavirus dentro de una LNP puede llevarse a cabo por técnicas conocidas por los expertos, como mezcla de microfluidos y generación de gotas llevada a cabo en un dispositivo generador de gotas de microfluidos. Estos dispositivos incluyen, de modo no taxativo, dispositivos de unión T estándar o dispositivos de enfoque de flujo. En un ejemplo, la formulación lipídica deseada, como composiciones que contienen MC3 o tipo MC3, se proporciona al dispositivo generador de gotas en paralelo con el vector de suministro de alfavirus y otros agentes deseados, de manera que el vector de suministro y los agentes deseados estén totalmente encapsulados en el interior de la LNP basada en MC3 o tipo MC3. En un ejemplo, el dispositivo generador de gotas puede controlar el intervalo de tamaño y distribución de tamaño de las LNP producidas. Por ejemplo, la LNP puede tener un tamaño que varía de 1 a 1000 nanómetros de diámetro, p. ej., 1, 10, 50, 100, 500 o 1000 nanómetros. Después de la generación de gotas, los vehículos de suministro que encapsulan los vectores de expresión pueden tratarse o modificarse adicionalmente para prepararlos para la administración.

V.E. Adenovirus de chimpancé (ChAd)

V.E.1. Suministro viral de adenovirus de chimpancé

[00400] Las composiciones de vacuna para suministro de uno o más neoantígenos (p. ej., mediante un casete del neoantígeno) pueden crearse al proporcionar secuencias de nucleótidos de adenovirus de

origen de chimpancé, una variedad de vectores novedosos y líneas celulares que expresan genes de adenovirus de chimpancé. Una secuencia de nucleótidos de un adenovirus C68 de chimpancé (también denominada en la presente ChAdV68) puede usarse en una composición de vacuna para el suministro de neoantígeno (remitirse a SEQ ID NO: 1). En USPN 6.083.716, que se incorpora en la presente mediante esta referencia en su totalidad a todos los efectos, se describe el uso de vectores derivados de adenovirus C68 en más detalle.

[00401] En un aspecto adicional, en la presente se proporciona un adenovirus recombinante que comprende una secuencia de ADN de un adenovirus de chimpancé como C68 y un casete del neoantígeno enlazado operativamente a secuencias reguladoras que dirigen su expresión. El virus recombinante puede infectar una célula de mamífero, preferentemente de ser humano, y puede expresar el producto de casete de neoantígeno en la célula. En este vector, el gen E1 y/o gen E3 y/o gen E4 de chimpancé natural puede eliminarse. Un casete de neoantígeno puede insertarse en cualquiera de estos sitios de eliminación de genes. El casete de neoantígeno puede incluir un neoantígeno contra el cual se desea una respuesta inmunitaria cebada.

[00402] En otro aspecto, en la presente se proporciona una célula de mamífero infectada con un adenovirus de chimpancé como C68.

[00403] En incluso un aspecto adicional, se proporciona una línea de células de mamíferos novedosa que expresa un gen de adenovirus de chimpancé (p, ej., de C68) o un fragmento funcional de esta.

[00404] En incluso un aspecto adicional, en la presente se proporciona un método para suministrar un casete del neoantígeno a una célula de mamífero que comprende la etapa de introducir en la célula una cantidad eficaz de un adenovirus de chimpancé, como C68, que se modificó genéticamente para expresar el casete del neoantígeno.

[00405] Incluso otro aspecto proporciona un método para provocar una respuesta inmunitaria en un hospedador mamífero para tratar un cáncer. El método puede comprender la etapa de administrarle al hospedador una cantidad eficaz de un adenovirus de chimpancé recombinante como C68, que comprende un casete del neoantígeno que codifica uno o más neoantígenos del tumor contra el cual se dirige la respuesta inmunitaria.

[00406] También se describe una célula de mamífero no simio que expresa un gen de adenovirus de chimpancé obtenido de la secuencia de la SEQ ID NO: 1. El gen puede seleccionarse del grupo que consiste en el adenovirus E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 de la SEQ ID NO: 1.

[00407] También se describe una molécula de ácido nucleico que comprende una secuencia de ADN de adenovirus de chimpancé que comprende un gen obtenido de la secuencia de SEQ ID NO: 1. El gen puede seleccionarse del grupo que consiste en dichos genes de adenovirus de chimpancé E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 de la SEQ ID NO: 1. En algunos aspectos, la molécula de ácido

nucleico comprende la SEQ ID NO: 1. En algunos aspectos, la molécula de ácido nucleico comprende la secuencia de SEQ ID NO: 1, que carece de al menos un gen que se selecciona del grupo que consiste en E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 de la SEQ ID NO: 1.

[00408] También se describe una vector que comprende una secuencia de ADN de adenovirus de chimpancé obtenido de la SEQ ID NO: 1 y un casete del neoantígeno enlazado operativamente a una o más secuencias reguladoras que dirigen la expresión del casete en una célula hospedadora heteróloga, opcionalmente donde la secuencia de ADN de adenovirus de chimpancé comprende al menos los elementos *cis* necesarios para la replicación y encapsidación de viriones, donde los elementos *cis* flanquean el casete del neoantígeno y las secuencias reguladoras. En algunos aspectos, la secuencia de ADN de adenovirus de chimpancé comprende un gen que se selecciona del grupo que consiste en las secuencias de genes E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 de la SEQ ID NO: 1. En algunos aspectos, el vector puede carecer del gen E1A y/o E1B.

[00409] En la presente también se describe una célula hospedadora transfectada con un vector descrito en la presente como un vector C68 modificado genéticamente para expresar un casete de neoantígeno. En la presente también se describe una célula humana que expresa un gen seleccionado introducido allí mediante introducción de un vector descrito en la presente en la célula.

[00410] En la presente también se describe un método para suministrar un casete del neoantígeno a una célula de mamífero que comprende introducir en dicha célula una cantidad eficaz de un vector descrito en la presente, como un vector C68 modificado genéticamente para expresar el casete del neoantígeno.

[00411] En la presente también se describe un método para producir un neoantígeno que comprende introducir un vector descrito en la presente en una célula de mamífero, cultivar la célula en condiciones adecuadas y producir el neoantígeno.

V.E.2. Líneas celulares de complemento que expresan E1

[00412] Para generar adenovirus (Ad) de chimpancé recombinantes eliminados en cualquiera de los genes descritos en la presente, la función de la región de genes eliminada, si es esencial para la replicación e infectividad del virus, puede suministrarse al virus recombinante por un virus o línea celular auxiliar, es decir, una línea celular de complemento o envasado. Por ejemplo, para generar un vector de adenovirus de chimpancé con replicación defectuosa, se puede usar una línea celular que expresa los productos del gen E1 del adenovirus de humano o chimpancé; dicha línea celular puede incluir HEK293 o variantes de esta. Puede seguirse el protocolo para la generación de las líneas celulares que expresan los productos del gen E1 de chimpancé (Ejemplos 3 y 4 de USPN 6.083.716) para generar una línea celular que expresa cualquier gen de adenovirus de chimpancé seleccionado.

[00413] Se puede usar un ensayo de aumento de AAV para identificar una línea celular que expresa

E1 de adenovirus de chimpancé. Este ensayo es útil para identificar la función de E1 en líneas celulares elaboradas usando los genes E1 de otros adenovirus no caracterizados, p. ej., de otras especies. Ese ensayo se describe en el Ejemplo 4B de USPN 6.083.716.

[00414] Un gen de adenovirus de chimpancé seleccionado, p. ej., E1 puede estar bajo el control de transcripción de un promotor para la expresión en una línea celular original seleccionada. Los promotores inducibles o constitutivos pueden emplearse a estos fines. Entre los promotores inducibles se incluye el promotor de la metalotionina de oveja, inducible por zinc, o el promotor del virus del tumor mamario del ratón (MMTV, por sus siglas en inglés), inducible por un glucocorticoide, en particular, dexametasona. Otros promotores inducibles, como los identificados en la solicitud de patente internacional WO95/13392, incorporada en la presente mediante referencia, también se pueden usar en la producción de líneas celulares de envoltura. También pueden emplearse los promotores constitutivos que controlan la expresión del gen de adenovirus de chimpancé.

[00415] Una célula original puede seleccionarse para la generación de una línea celular novedosa que expresa cualquier gen C68 deseado. Sin limitación, dicha línea celular original puede ser las células HeLa [n.° de acceso de ATCC CCL 2], A549 [n.° de acceso de ATCC CCL 185], KB [CCL 17], Detroit [p. ej., Detroit 510, CCL 72] y WI-38 [CCL 75] Otras líneas celulares originales adecuadas pueden obtenerse de otras fuentes. Las líneas celulares originales pueden incluir CHO, HEK293 o variantes de estas, 911, HeLa, A549, LP-293, PER.C6 o AE1-2a.

[00416] Una línea celular que expresa E1 puede ser útil en la generación de vectores con E1 de adenovirus de chimpancé recombinante eliminado. Las líneas celulares construidas usando básicamente los mismos procedimientos que expresan uno o más productos de genes de adenovirus de chimpancé diferente son útiles en la generación de vectores de adenovirus de chimpancé recombinante eliminados en los genes que codifican esos productos. Además, las líneas celulares que expresan otros productos de genes E1 Ad también son útiles para generar Ad recombinantes de chimpancé.

V.E.3. Partículas virales recombinantes como vectores

[00417] Las composiciones descritas en la presente pueden comprender vectores virales, que suministran al menos un neoantígeno a las células. Estos vectores comprenden una secuencia de ADN de adenovirus de chimpancé como C68 y un casete del neoantígeno enlazado operativamente a secuencias reguladoras que dirigen la expresión del casete. El vector C68 puede expresar el casete en una célula de mamífero infectada. El vector C68 puede eliminarse funcionalmente en uno o más genes virales. Un casete de neoantígeno comprende al menos un neoantígeno bajo el control de una o más secuencias reguladoras como un promotor. Virus auxiliares y/o líneas celulares de envasado opcionales pueden suministrarle al vector viral de chimpancé cualquier producto necesario de genes adenovirales eliminados.

[00418] La expresión «eliminada funcionalmente» significa que una cantidad suficiente de la región del gen se elimina o altera de otro modo, p. ej., por mutación o modificación, de modo que la región del gen ya no sea capaz de producir uno o más productos funcionales de expresión génica. Las mutaciones o modificaciones que pueden dar como resultado deleciones funcionales incluyen, de modo no taxativo, mutaciones sin sentido, tales como la introducción de codones de detención prematuros y la eliminación de codones de inicio canónicos y no canónicos, mutaciones que alteran el corte y empalme del ARNm u otros procesos transcripcionales, o combinaciones de estos. Si se desea, puede retirarse toda la región del gen.

[00419] Las modificaciones de las secuencias de ácido nucleico que forman los vectores descritos en la presente, inclusive eliminaciones, inserciones y otras mutaciones de secuencia, pueden generarse usando técnicas de biología molecular estándar que están dentro del alcance de la presente invención.

V.E.4. Construcción del vector de plásmido viral

[00420] Los vectores C68 de adenovirus de chimpancé útiles en esta invención incluyen adenovirus recombinantes y defectuosos, es decir, secuencias de adenovirus de chimpancé eliminadas funcionalmente en los genes E1a o E1b y que opcionalmente portan otras mutaciones, p. ej., mutaciones sensibles a la temperatura o eliminaciones en otros genes. Se anticipa que estas secuencias de chimpancé también son útiles para formar vectores híbridos de otros adenovirus y/o secuencias de virus adenoasociadas. Los vectores de adenovirus homólogos preparados a partir de adenovirus humanos se describen en la bibliografía publicada [véase, por ejemplo, Kozarsky I y II, citado anteriormente y referencias citadas allí, patente estadounidense n.° 5.240.846].

[00421] En la construcción de vectores C68 de adenovirus de chimpancé útiles para el suministro de un casete del neoantígeno a una célula de humano (u otro mamífero), se puede emplear un rango de secuencias de ácido nucleico de adenovirus en los vectores. Un vector que comprende secuencias de adenovirus C68 de chimpancé puede usarse junto con un virus auxiliar para producir una partícula de virus recombinante infecciosa. El virus auxiliar proporciona productos génicos esenciales necesarios para la infectividad viral y la propagación del vector adenoviral de chimpancé mínimo. Cuando solo se realiza una o más eliminaciones seleccionadas de genes de adenovirus de chimpancé en un vector viral que de otro modo era funcional, los productos génicos eliminados pueden suministrarse en el proceso de producción de vectores virales al propagar el virus en una línea celular de envasado seleccionada que proporciona las funciones del gen eliminado in trans.

V.E.5. Adenovirus mínimo recombinante

[00422] Un virus Ad C68 mínimo de chimpancé es una partícula viral que contiene únicamente los elementos cis de adenovirus necesarios para la replicación y encapsidación. Es decir, el vector contiene

las secuencias de repetición terminal invertida (ITR, por su sigla en inglés) 3’ y 5’ de actuación en cis de los adenovirus (que actúan como orígenes de replicación) y los dominios naturales de potenciamiento/envoltura 5’ (que contienen secuencias necesarias para envolver genomas de Ad lineales, así como elementos potenciadores del promotor de E1). Remitirse, por ejemplo, a las técnicas descritas para la elaboración de un vector de Ad humano «mínimo» en la solicitud de patente internacional WO96/13597, que se incorpora a la presente mediante esta referencia.

V.E.6. Otros adenovirus defectuosos

[00423] Los adenovirus también pueden contener más que las secuencias de adenovirus mínimo de chimpancé. Estos otros vectores de Ad pueden caracterizarse a través de eliminaciones de diversas partes de las regiones génicas del virus, así como de partículas infecciosas del virus que se forman mediante el uso opcional de los virus auxiliares y/o las líneas celulares de envoltura.

[00424] A modo de ejemplo, pueden formarse vectores adecuados eliminando la totalidad o una parte suficiente del gen temprano retrasado E1b y el gen temprano inmediato E1a de adenovirus C68, a los efectos de eliminar sus funciones biológicas normales. Los virus con eliminación de E1 y deficiencia de replicación tienen la capacidad de replicarse y producir un virus infeccioso al cultivarse en una línea celular de complemento, transformada con adenovirus de chimpancé, que contenga los genes funcionales E1a y E1b de adenovirus, los cuales proporcionan los productos génicos correspondientes en trans. En función de las homologías de secuencias de adenovirus conocidas, se anticipa que, así como sucede en los adenovirus humanos con eliminación de E1 recombinante de la técnica, el adenovirus recombinante de chimpancé resultante tenga la capacidad de infectar muchos tipos de células y pueda expresar uno o más neoantígenos, pero no pueda replicarse en la mayoría de las células que no contengan el ADN de la región E1 de chimpancé, salvo que la célula presente una multiplicidad muy elevada de infección.

[00425] A modo de ejemplo adicional, se puede eliminar la totalidad o una parte del gen temprano retrasado E3 de adenovirus C68 de la secuencia de adenovirus de chimpancé que forma parte del virus recombinante.

[00426] Los vectores de adenovirus C68 de chimpancé también pueden construirse con eliminación del gen E4. Otro vector adicional puede contener una eliminación en el gen temprano retrasado E2a.

[00427] Asimismo, pueden realizarse eliminaciones en muchos de los genes tardíos L1 a L5 del genoma de adenovirus C68 de chimpancé. De forma similar, para algunos fines pueden ser útiles las eliminaciones en los genes intermedios IX y IVa2. Se pueden realizar otras eliminaciones en los otros genes estructurales o no estructurales de adenovirus.

[00428] Las eliminaciones mencionadas en lo precedente pueden utilizarse de forma individual, es decir, es posible que una secuencia de adenovirus contenga únicamente eliminaciones de E1. De forma

alternativa, se pueden utilizar en cualquier combinación eliminaciones de genes enteros o partes de estos que sean eficaces para destruir o reducir su actividad biológica. Por ejemplo, en un ejemplo de vector, la secuencia de adenovirus C68 puede presentar eliminaciones de los genes E1 y el gen E4, de los genes E1, E2a y E3, de los genes E1 y E3 o de los genes E1, E2a y E4, con o sin eliminación de E3, y así sucesivamente. Tal como se indicó anteriormente, estas eliminaciones pueden utilizarse en combinación con otras mutaciones, como mutaciones sensibles a la temperatura, para lograr un resultado deseado.

[00429] El casete que comprende uno o más neoantígenos se puede insertar opcionalmente en cualquier región eliminada del virus de Ad C68 de chimpancé. De forma alternativa y si se desea, el casete puede insertarse en una región génica existente, para alterar la función de dicha región.

V.E.7. Virus auxiliares

[00430] Dependiendo del contenido de gen de adenovirus de chimpancé en los vectores virales que se utilicen para transportar el casete de neoantígeno, puede utilizarse un fragmento de virus sin replicación o adenovirus auxiliar para proporcionar una cantidad suficiente de secuencias génicas de adenovirus de chimpancé para producir una partícula viral recombinante ineficaz que contenga el casete.

[00431] Los virus auxiliares útiles contienen secuencias génicas de adenovirus selectas que no se encuentran presentes en la construcción del vector de adenovirus y/o que no se expresan en línea celular de envoltura en la cual se transfecta el vector. Un virus auxiliar puede presentar deficiencia de replicación y contener diversos genes de adenovirus además de las secuencias descritas anteriormente. El virus auxiliar puede utilizarse en combinación con las líneas celulares con expresión de E1 que se describen en la presente.

[00432] Para C68, el virus «auxiliar» puede ser un fragmento formado mediante el recorte del extremo terminal C del genoma C68 con SspI, lo cual elimina alrededor de 1300 pb del extremo izquierdo del virus. El virus recortado luego se transfecta de forma conjunta a la línea celular con expresión de E1 con el ADN plásmido, con lo cual se forma el virus recombinante mediante recombinación homóloga con las secuencias de C68 en el plásmido.

[00433] Los virus auxiliares también pueden formarse en conjugados policatiónicos, como se describe en Wu et ál, J. Biol. Chem., 264:16985-16987 (1989); K. J. Fisher y J. M. Wilson, Biochem. J., 299:49 (1º de abril de 1994). El virus auxiliar puede contener opcionalmente un gen informante. En la técnica se conocen diversos genes informantes de este tipo. La presencia de un gen informante en el virus auxiliar diferente al casete neoantigénico en el vector de adenovirus permite un control independiente tanto del vector de Ad como del virus auxiliar. Este segundo informante se utiliza para permitir la separación entre

el virus recombinante resultante y el virus auxiliar tras la purificación.

V.E.8. Ensamblaje de partícula viral e infección de una línea celular

[00434] El ensamblaje de las secuencias de ADN selectas del adenovirus, el casete neoantigénico y otros elementos vectoriales en diversos plásmidos intermedios y vectores transportadores, así como el uso de los plásmidos y vectores para producir una partícula viral recombinante, pueden lograrse utilizando técnicas convencionales. Estas técnicas incluyen técnicas convencionales de clonación de ADNc, técnicas de recombinación in vitro (p. ej., ensamblaje Gibson), uso de secuencias superpuestas de oligonucleótidos de los genomas de adenovirus, reacción en cadena de la polimerasa y cualquier método adecuados que brinde la secuencia de nucleótidos deseada. Se utilizan técnicas estándar de transfección y transfección conjunta, p. ej., técnicas de precipitación con CaPO4 o métodos de transfección mediada por liposomas, como lipofectamina. Otros métodos convencionales que se utilizan incluyen recombinación homóloga de los genomas virales, formación de placas de virus en superposición de agar, métodos de medición de la generación de señales y similares.

[00435] Por ejemplo, tras la construcción y el ensamblaje del vector viral que contiene el casete neoantigénico deseado, el vector puede transfectarse in vitro y en presencia de un virus auxiliar a la línea celular de envoltura. Se produce recombinación homóloga entre las secuencias del vector y el auxiliar, lo cual permite la replicación y envoltura de las secuencias neoantigénicas de adenovirus en el vector en cápsides víricas que producen partículas vectoriales del vector recombinantes.

[00436] Los adenovirus C68 de chimpancé recombinante resultantes son útiles en la transferencia de un casete neoantigénico a una célula selecta. En los experimentos in vivo con el virus recombinante cultivado en las líneas celulares de envoltura, el adenovirus recombinante de chimpancé con eliminación de E1 demuestra utilidad en la transferencia de un casete a una célula que no sea de chimpancé, preferentemente de ser humano.

V.E.9. Uso de los vectores de virus recombinante

[00437] De esta forma, el adenovirus recombinante C68 de chimpancé que contiene el casete neoantigénico (producido mediante cooperación del vector de adenovirus y el virus auxiliar o el vector de adenovirus y la línea celular de envoltura, como se describió anteriormente) proporciona un vehículo de transferencia génica eficaz que puede suministrarle el o los neoantígenos a un sujeto in vivo o ex vivo.

[00438] Los vectores recombinantes descritos anteriormente se administran a humanos de acuerdo con los métodos publicados para terapia génica. Se le puede administrar a un paciente un vector viral de chimpancé que presente un casete neoantigénico, preferentemente suspendido en una solución biológicamente compatible o un vehículo de suministro farmacéuticamente aceptable. Un vehículo adecuado incluye solución salina estéril. A tales efectos, pueden utilizarse otras soluciones isotónicas

estériles acuosas y no acuosas, así como suspensiones estériles acuosas y no acuosas, para inyección conocidas por los expertos en la técnica y que se sepa que son portadores farmacéuticamente aceptables.

[00439] Los vectores de adenovirus de chimpancé se administran en cantidades suficiente para la transducción de las células humanas y para proporcionar niveles suficientes de transferencia y expresión del neoantígeno como para brindar un beneficio terapéutico sin efectos fisiológicos adversos indebidos o con efectos fisiológicos aceptables en medicina, los cuales pueden ser determinados por los expertos en la técnica médica. Las vías de administración convencionales y farmacéuticamente aceptables incluyen, de modo no taxativo, suministro directo en el hígado, intranasal, intravenoso, intramuscular, subcutáneo, intradérmico, oral y otras vías de administración parenteral. Si se desea, las vías de administración se pueden combinar.

[00440] Las dosificaciones del vector viral dependerán principalmente de factores como la afección que se esté tratando, la edad, el peso y el estado de salud del paciente, y, por lo tanto, varía según el paciente. La dosificación se ajusta para equilibrar el beneficio terapéutico en relación con cualquier efecto secundario y estas dosis pueden variar dependiendo de la aplicación terapéutica para la cual se utilice el vector recombinante. Se pueden controlar los niveles de expresión del o de los neoantígenos para determinar la frecuencia de administración de las dosis.

[00441] Los adenovirus recombinantes y con deficiencia de replicación pueden administrarse en una «cantidad farmacéuticamente eficaz», es decir, una cantidad de adenovirus recombinante eficaz en una vía de administración para la transfección de las células deseadas y para proporcionar niveles de expresión suficiente del gen selecto para brindar un beneficio de vacunación, es decir, determinado nivel ponderable de inmunidad de protección. Los vectores de C68 que comprendan un casete neoantigénico pueden administrarse de forma conjunta con un adyuvante. El adyuvante puede estar separado del vector (p. ej., alumbre) o codificado dentro del vector, en particular si el adyuvante es una proteína. En la técnica se conocen los adyuvantes.

[00442] Las vías de administración convencionales y farmacéuticamente aceptables incluyen, de modo no taxativo, intranasal, intramuscular, intratraqueal, subcutánea, intradérmica, rectal, oral y otras vías de administración parenteral. Si se desea, es posible combinar las vías de administración o ajustarlas dependiendo del inmunógeno o de la enfermedad. Por ejemplo, en la profilaxis de la rabia, se prefieren las vías subcutánea, intratraqueal e intranasal. La vía de administración dependerá principalmente de la naturaleza de la enfermedad que se esté tratando.

[00443] Se pueden controlar los niveles de inmunidad del o de los neoantígenos para determinar la necesidad de dosis de refuerzo, si la hubiera. Tras la evaluación de las titulaciones de anticuerpo en suero, por ejemplo, puede resultar deseable realizar inmunizaciones de refuerzo opcionales.

VI. Métodos terapéuticos y de elaboración

[00444] También se proporciona un método para inducir una respuesta inmunitaria específica para el tumor en un sujeto, realizar la vacunación contra un tumor, tratar y/o aliviar un síntoma de cáncer en un sujeto mediante la administración a dicho sujeto de uno o más neoantígenos tales como múltiples neoantígenos identificados a través de los métodos descritos en la presente.

[00445] En algunos aspectos, un sujeto cuenta con un diagnóstico de cáncer o presenta un riesgo de desarrollo de cáncer. Un sujeto puede ser un ser humano, un perro, un gato, un caballo o cualquier animal en el cual se desee una respuesta inmunitaria específica para el tumor. Un tumor puede ser un tumor sólido, como de mama, ovarios, próstata, pulmonar, de riñón, gástrico, de colon, testicular, de cabeza y cuello, páncreas, cerebro, melanoma y otros tumores de órganos tisulares, así como tumores hematológicos, tales como linfomas y leucemias, incluidas leucemia mielógena aguda, leucemia mielógena crónica, leucemia linfocítica crónica, leucemia linfocítica de linfocitos T y linfomas de linfocitos B.

[00446] Se puede administrar un neoantígeno en una cantidad suficiente para inducir una respuesta de CTL.

[00447] El neoantígeno puede administrarse solo o en combinación con otros agentes terapéuticos. Por ejemplo, el agente terapéutico es un agente quimioterapéutico, de radiación o inmunoterapia. Se puede administrar cualquier tratamiento terapéutico adecuado para un cáncer particular.

[00448] Además, un sujeto puede recibir adicionalmente un agente antiinmunosupresor/inmunoestimulador, como un inhibidor de punto de control. Por ejemplo, el sujeto puede recibir adicionalmente un anticuerpo anti-CTLA, anti-PD-1 o anti-PD-L1. El bloqueo de CTLA-4 o PD-L1 por parte de anticuerpos puede potenciar la respuesta inmunitaria a células cancerosas en el paciente. En particular, se ha demostrado que el bloqueo de CTLA-4 es eficaz cuando se sigue un protocolo de vacunación.

[00449] Se pueden determinar la cantidad óptima de cada neoantígeno que se incluya en la composición de vacuna y el régimen de dosificación óptimo. Por ejemplo, se pueden preparar un neoantígeno o su variante para inyección intravenosa (i.v.), inyección subcutánea (s.c.), inyección intradérmica (i.d.), inyección intraperitoneal (i.p.), inyección intramuscular (i.m.). Los métodos de inyección incluyen s.c, i.d., i.p., i.m. e i.v. Los métodos de inyección de ADN o ARN incluyen i.d., i.m., s.c, i.p. e i.v. Los expertos en la técnica conocen otros métodos de administración de la composición de vacuna.

[00450] Se puede compilar una vacuna de manera tal que la selección, el número y/o la cantidad de neoantígenos presentes en la composición sea específica del tejido, el cáncer y/o el paciente. Por ejemplo, la selección exacta de péptidos puede estar guiada por patrones de expresión de las proteínas

originales en un tejido específico. La selección puede depender del tipo específico de cáncer, el estado de la enfermedad, regímenes de tratamiento anteriores, el estado inmunológico del paciente y, por supuesto, el haplotipo de HLA del paciente. Además, una vacuna puede contener componentes individualizados, según las necesidades personales del paciente particular. Los ejemplos incluyen variación de la selección de neoantígenos de acuerdo con la expresión del neoantígeno en el paciente particular o ajustes de tratamientos secundarios tras una primera ronda o un primer esquema de tratamiento.

[00451] Para poder utilizar una composición como vacuna para el cáncer, se pueden evitar neoantígenos con autopéptidos expresados en grandes cantidades en tejidos normales o estos pueden estar presentes en cantidades bajas en una composición descrita en la presente. Por otra parte, si se sabe que el tumor de un paciente expresa cantidades elevadas de un determinado neoantígeno, la composición farmacéutica respectiva para el tratamiento de este cáncer puede estar presente en cantidades elevadas y/o se puede incluir más de un neoantígeno específico para este neoantígeno o vía de este neoantígeno particular.

[00452] Se le pueden administrar composiciones que comprendan un neoantígeno a un individuo que ya padezca el cáncer. En aplicaciones terapéuticas, las composiciones se le administran a un paciente en una cantidad suficiente para provocar una respuesta de CTL eficaz contra el antígeno tumoral y para curar o al menos detener parcialmente los síntomas y/o las complicaciones. Una cantidad adecuada para lograr este resultado se define como «dosis terapéuticamente eficaz». Las cantidades eficaces para este uso dependerán, p. ej., de la composición, la forma de administración, la etapa y gravedad de la enfermedad que se esté tratando, el peso y estado general de salud del paciente, así como del criterio del médico tratante. Se debe tener en cuenta que las composiciones generalmente pueden emplearse en estados graves de enfermedad, es decir, situaciones con riesgo de vida o potencialmente mortales, especialmente cuando el cáncer presenta metástasis. En estos casos y en función de la minimización de sustancias extrañas y la naturaleza no tóxica relativa de un neoantígeno, es posible que el médico tratante administre excesos sustanciales de estas composiciones, lo cual puede resultar deseable.

[00453] Para el uso terapéutico, la administración puede comenzar a partir de la detección o extirpación quirúrgica de tumores. A esto le siguen dosis de refuerzo hasta que al menos los síntomas disminuyan sustancialmente y durante un período posterior.

[00454] Las composiciones farmacéuticas (por ejemplo, composiciones de vacuna) para el tratamiento terapéutico se prevén para administración parenteral, tópica, nasal, oral o local. Las composiciones farmacéuticas se pueden administrar por vía parenteral, p. ej., intravenosa, subcutánea, intradérmica o intramuscular. Las composiciones pueden administrarse en el lugar de la extirpación quirúrgica para inducir una respuesta inmunitaria local al tumor. En la presente se describen

composiciones para administración parenteral, que comprenden una solución del neoantígeno y las composiciones de vacunas se disuelven o suspenden en un portador aceptable, p. ej., un portador acuoso. Puede utilizarse diversos portadores acuosos, p. ej., agua, agua amortiguada, solución salina al 0,9 %, glicina al 0,3 %, ácido hialurónico y similares. Estas composiciones farmacéuticas pueden esterilizarse mediante técnicas de esterilización convencionales conocidas o pueden filtrarse en condiciones estériles. Las soluciones acuosas resultantes pueden envasarse para uso en la forma en que se encuentran o liofilizarse, en cuyo caso la preparación liofilizada se combina con una solución estéril antes de la administración. Las composiciones pueden contener sustancias auxiliares farmacéuticamente aceptables necesarias para acercarse a las condiciones fisiológicas, tales como agentes amortiguadores y de ajuste del pH, agentes de ajuste de tonicidad, agentes humectantes y similares, por ejemplo, acetato de sodio, lactato de sodio, cloruro de sodio, cloruro de potasio, cloruro de calcio, monolaurato de sorbitán, oleato de trietanolamina, etc.

[00455] Los neoantígenos también pueden administrarse mediante liposomas, los cuales los dirigen hacia un tejido celular particular, como un tejido linfoide. Los liposomas también son útiles para aumentar la semivida. Los liposomas incluyen emulsiones, espumas, micelas, monocapas insolubles, cristales líquidos, dispersiones de fosfolípidos, capas laminares y similares. En estas preparaciones, el neoantígeno que se desea suministrar se incorpora como parte de un liposoma, solo o en conjunto con una molécula que se una, p. ej., a un receptor con prevalencia entre las células linfoides, como anticuerpos monoclonales que se unan al antígeno CD45, o con otras composiciones terapéuticas o inmunogénicas. Por lo tanto, los liposomas rellenos con un neoantígeno deseado se pueden dirigir al lugar de las células linfoides, donde los liposomas luego suministran las composiciones terapéuticas/inmunogénicas. Los liposomas pueden formarse a partir de lípidos estándar que forman vesículas, los cuales generalmente incluyen fosfolípidos con carga neutra y negativa y un esterol, como colesterol. La selección de lípidos generalmente está guiada, p. ej., por consideraciones de tamaño del liposoma, labilidad ácida y estabilidad de los liposomas en el torrente sanguíneo. Diversos métodos para preparar liposomas se encuentran disponibles, como se describe, p. ej., en Szoka et ál., Ann. Rev. Biophys. Bioeng. 9; 467 (1980), patentes estadounidenses n.º 4.235.871, 4.501.728, 4.501.728, 4.837.028 y 5.019.369.

[00456] Para el direccionamiento hacia las células inmunológicas, un ligando que se incorpore en el liposoma puede incluir, p. ej., anticuerpos o fragmentos de estos específicos para determinantes de superficie celular de las células del sistema inmunológico deseadas. Se puede administrar una suspensión de liposomas por vía intravenosa, local, tópica, etc., en una dosis que varíe según la forma de administración, el péptido que se suministre y la etapa de la enfermedad que se esté tratando, entre otros.

[00457] Con fines de inmunización o terapéuticos, también se le pueden administrar al paciente ácidos nucleicos que codifiquen un péptido y, opcionalmente, uno o más de los péptidos descritos en la presente. Se utilizan de forma conveniente diversos métodos para el suministro de los ácidos nucleicos al paciente. Por ejemplo, el ácido nucleico puede suministrarse de forma directa como «ADN sin conjugar». Este enfoque se describe, por ejemplo, en Wolff et ál., Science 247: 1465-1468 (1990), así como en las patentes estadounidenses n.º 5.580.859 y 5.589.466. Los ácidos nucleicos también pueden administrarse utilizando suministro balístico, como se describe, por ejemplo, en la patente estadounidense n.º 5.204.253. Es posible administrar partículas compuestas únicamente por ADN. De forma alternativa, el ADN puede encontrarse adherido a partículas, como partículas de oro. Los enfoques para suministro de secuencias de ácido nucleico pueden incluir vectores virales, vectores de ARNm y vectores de ADN con o sin electroporación.

[00458] Los ácidos nucleicos también pueden suministrarse en complejo con compuestos catiónicos, como lípidos catiónicos. Los métodos de suministro de genes mediados por lípidos se describen, por ejemplo, en 9618372WOAWO 96/18372; 9324640WOAWO 93/24640; Mannino & Gould-Fogerite, BioTechniques 6(7): 682-691 (1988); la patente estadounidense n.º 5.279.833 Rose, la patente estadounidense n.º 5.279.833; 9106309WOAWO 91/06309; y en Felgner et ál., Proc. Natl. Acad. Sci. EUA 84: 7413-7414 (1987).

[00459] También pueden incluir los neoantígenos en plataformas de vacuna basadas en vectores virales, como vaccinia, viruela aviar, alfavirus de autorreplicación, marabavirus, adenovirus (remitirse, p. ej., a Tatsis et ál., Adenoviruses, *Molecular Therapy* (2004) 10, 616—629) o lentivirus, incluidos, de modo no taxativo, lentivirus de segunda o tercera generación o híbridos de segunda y tercera generación y lentivirus recombinantes de cualquier generación diseñados para dirigirse a receptores o tipos de células específicos (remitirse, p. ej., a Hu et ál., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, *Immunol Rev.* (2011) 239(1): 45-61, Sakuma et ál., Lentiviral vectors: basic to translational, *Biochem J.* (2012) 443(3):603-18, Cooper et ál., Rescue of splicing-mediated intron loss maximizes expression in lentiviral vectors containing the human ubiquitin C promoter, *Nucl. Acids Res.* (2015) 43 (1): 682-690, Zufferey et ál., Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, *J. Virol.* (1998) 72 (12): 9873-9880). Dependiendo de la capacidad de envasado de las plataformas de vacunas basadas en vectores virales antedichas, este enfoque puede suministrar una o más secuencias de nucleótidos que codifiquen uno o más péptidos neoantigénicos. Las secuencias pueden estar flanqueadas por secuencias sin mutación, pueden estar separadas por enlazadores o pueden estar precedidas por una o más secuencias dirigidas al compartimiento subcelular (remitirse, p. ej., a Gros et ál., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen et ál., Targeting of cancer neoantigens with

donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu et ál., Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions, *Clin Cancer Res.* (2014) 20(13):3401-10). Tras la introducción en un hospedador, las células infectadas expresan los neoantígenos y, de esa forma, provocan una respuesta inmunitaria del hospedador (p. ej., CTL) contra el o los péptidos. Los vectores de vaccinia y métodos útiles para los protocolos de inmunización se encuentran descritos, p. ej., en la patente estadounidense n.º 4.722.848. Otro vector es BCG (Bacilos Calmette-Guerin). Los vectores de BCG se encuentran descritos en Stover et ál. (Nature 351:456-460 (1991)). Los expertos en la técnica podrán observar una amplia variedad de vectores de vacunas adicionales, útiles para la inmunización o administración terapéutica de los neoantígenos, p. ej., vectores de Salmonella typhi y similares, a partir de la descripción en la presente.

[00460] Un medio de administración de ácidos nucleicos utiliza construcciones de minigenes que codifican uno o múltiples epítopos. Para crear una secuencia de ADN que codifique los epítopos de CTL selectos (minigenes) para expresión en células humanas, las secuencias de aminoácidos de los epítopos se someten a traducción inversa. Se utiliza una tabla de uso de codones humanos para guiar la selección de codones para cada aminoácido. Estas secuencias de ADN que codifican epítopos se unen de forma directa y así crean una secuencia de polipéptidos continua. Para optimizar la expresión y/o inmunogenicidad, se pueden incorporar elementos adicionales en el diseño de minigenes. Los ejemplos de secuencias de aminoácidos que podrían someterse a traducción inversa y que se incluyen en la secuencia de minigenes incluyen: linfocitos T auxiliares, epítopos, una secuencia líder (señal) y una señal de retención del retículo endoplásmico. Adicionalmente, se puede mejorar la presentación de epítopos de CTL en el MHC incluyendo secuencias flanqueantes de origen natural o sintéticas (p. ej., poli-alanina) adyacentes a los epítopos de CTL. La secuencia de minigenes se convirtieron en ADN mediante ensamblaje de oligonucleótidos que codifican las cadenas más y menos del minigén. Los oligonucleótidos superpuestos (30 a 100 bases de longitud) se sintetizan, fosforilan, purifican e hibridan en condiciones adecuadas mediante técnicas conocidas. Los extremos de los oligonucleótidos se unen utilizando ligasa de ADN T4. Este minigén sintético que codifica el polipéptido epitópico de CTL luego puede clonarse en un vector de expresión deseado.

[00461] Puede prepararse ADN plásmido purificado para inyección utilizando diversas formulaciones. La más sencilla de estas es la reconstitución de ADN liofilizado en solución salina amortiguada con fosfato (PBS) estéril. Se describieron diversos métodos y es posible que se creen nuevas técnicas. Tal como se indicó anteriormente, los ácidos nucleicos se formulan de manera conveniente con lípidos catiónicos. Además, los glucolípidos, liposomas fusogénicos, péptidos y compuestos que se indican de forma colectiva como protectores, interactivos, sin condensación (PINC) también podrían formarse en complejo para ADN plásmido purificado, para afectar variables como la estabilidad, la dispersión

intramuscular o el tráfico hacia tipos de células u órganos específicos.

[00462] También se describe un método para elaborar una vacuna tumoral, que comprende realizar las etapas de un método descrito en la presente, y producir una vacuna tumoral que comprenda múltiples neoantígenos o un subconjunto de los múltiples neoantígenos.

[00463] Los neoantígenos que se describen en la presente pueden elaborarse utilizando métodos conocidos en la técnica. Por ejemplo, un método para producir un neoantígeno o un vector (p. ej., un vector que incluya al menos una secuencia que codifique uno o más neoantígenos) descrito en la presente puede incluir el cultivo de una célula hospedadora en condiciones adecuadas para expresar el neoantígeno o el vector, donde la célula hospedadora comprende al menos un polinucleótido que codifique el neoantígeno o vector, y la purificación del neoantígeno o vector. Los métodos de purificación estándar incluyen técnicas cromatográficas, técnicas electroforéticas, inmunológicas, de precipitación, diálisis, filtración, concentración y cromatoenfoque.

[00464] Las células hospedadoras pueden incluir una célula de ovario de hámster chino (CHO), célula de NS0, levadura o célula de HEK293. Las células hospedadoras pueden transformarse con uno o más polinucleótidos que comprendan al menos una secuencia de ácido nucleico que codifique un neoantígeno o vector descrito en la presente, opcionalmente donde el polinucleótido aislado comprende además una secuencia de promotor enlazada de forma operativa a al menos una secuencia de ácido nucleico que codifique el neoantígeno o vector. En determinadas modalidades, el polinucleótido aislado puede ser ADNc.

VII. Administración y uso de neoantígenos

[00465] Se puede utilizar un protocolo de vacunación para suministrarle a un sujeto dosis de uno o más neoantígenos. Se puede utilizar una vacuna inicial y una vacuna de refuerzo para suministro de dosis al sujeto. La vacuna inicial puede basarse en C68 (p. ej., las secuencias ilustradas en las SEQ ID NO: 1 o 2) o srRNA (p. ej., las secuencias ilustradas en las SEQ ID NO: 3 o 4) y la vacuna de refuerzo puede basarse en C68 (p. ej., las secuencias ilustradas en las SEQ ID NO: 1 o 2) o srRNA (p. ej., las secuencias ilustradas en las SEQ ID NO: 3 o 4). Cada vector suele incluir un casete con neoantígenos. Los casetes pueden incluir alrededor de 20 neoantígenos, separados por separadores como la secuencia natural que suele rodear cada antígeno u otras secuencias separadoras no natural, como AAY. Los casetes también incluyen antígenos del MHCII, como un antígeno del toxoide del tétano y antígeno de PADRE, que pueden considerarse antígenos universales de clase II. Los casetes también incluyen una secuencia de direccionamiento, como una secuencia de direccionamiento de ubiquitina. Asimismo, cada dosis de la vacuna puede administrarse al sujeto en conjunto (p. ej., de forma simultánea, antes o después) con un inhibidor de punto de control (CPI, por sus siglas en inglés). Los CPI pueden utilizar aquellos que inhiben CTLA4, PD1 y/o PDL1, como anticuerpos o partes de unión a

antígenos de estos. Estos anticuerpos pueden incluir tremelimumab o durvalumab.

[00466] El sujeto puede recibir una inyección de vacuna inicial (p. ej., por vía intramuscular). Se pueden utilizar inyecciones bilaterales por dosis. Por ejemplo, es posible utilizar una o más inyecciones de ChAdV68 (C68) (p. ej., dosis total de $1x10^{12}$ partículas virales); es posible utilizar una o más inyecciones de ARN autorreplicante (srRNA) a una dosis de vacunación baja, seleccionada del intervalo de 0,001 a 1 ug de ARN, en particular 0,1 o 1 ug; o es posible utilizar una o más inyecciones de srRNA a una dosis de vacunación elevada, seleccionada del intervalo de 1 a 100 ug de ARN, en particular 10 o 100 ug.

[00467] Se puede inyectar una vacuna de refuerzo (vacuna reforzadora) (p. ej., por vía intramuscular) tras la vacunación inicial. La vacuna de refuerzo puede administrarse aproximadamente cada 1, 2, 3, 4, 5, 6, 7, 8, 9 o 10 semanas, p. ej., cada 4 semanas y/u 8 semanas tras la vacuna inicial. Se pueden utilizar inyecciones bilaterales por dosis. Por ejemplo, es posible utilizar una o más inyecciones de ChAdV68 (C68) (p. ej., dosis total de $1x10^{12}$ partículas virales); es posible utilizar una o más inyecciones de ARN autorreplicante (srRNA, por sus siglas en inglés) a una dosis de vacunación baja, seleccionada del intervalo de 0,001 a 1 ug de ARN, en particular 0,1 o 1 ug; o es posible utilizar una o más inyecciones de srRNA a una dosis de vacunación elevada, seleccionada del intervalo de 1 a 100 ug de ARN, en particular 10 o 100 ug.

[00468] También se le puede administrar al sujeto anti-CTLA-4 (p. ej., tremelimumab). Por ejemplo, el anti-CTLA4 puede administrarse por vía subcutánea cerca del lugar de inyección intramuscular de la vacuna (dosis inicial de ChAdV68 o baja de srRNA), para garantizar el drenaje hacia el mismo nódulo linfático. Tremelimumab es un inhibidor de mAb de IgG2 humana selectivo de CTLA-4. La dosis subcutánea de anti-CTLA-4 objetivo (tremelimumab) suele ser de 70 a 75 mg (en particular, 75 mg), con un intervalo de dosis, p. ej., de 1 a 100 mg o 5 a 420 mg.

[00469] En determinados casos, se puede utilizar un anticuerpo anti-PD-L1 como durvalumab (MEDI 4736). Durvalumab es un mAb de IgG1 humana selectivo y de gran afinidad que bloque la unión de PD-L1 a PD-1 y CD80. Durvalumab suele administrarse a 20 mg/kg i.v. cada 4 semanas.

[00470] Se puede realizar control inmunológico antes, durante y/o después de la administración de la vacuna. Este control puede brindar información sobre la seguridad y eficacia, entre otros parámetros.

[00471] Para realizar el control inmunológico, se suelen utilizar PBMC. Las PBMC pueden aislarse antes de la vacunación inicial y después de la vacunación inicial (p. ej., 4 semanas y 8 semanas). Las PBMC pueden extraerse inmediatamente antes de las vacunaciones de refuerzo y después de cada vacunación de refuerzo (p. ej., 4 semanas y 8 semanas).

[00472] El protocolo de control inmunológico puede incluir evaluación de las respuestas de linfocitos T. Las respuestas de linfocitos T pueden medirse utilizando uno o más métodos conocidos en la técnica,

como ELISpot, tinción de citocinas intracelulares, secreción de citocinas y captura de superficie celular, proliferación de linfocitos T, tinción multimérica de MHC o ensayos de citotoxicidad. Es posible controlar las respuestas de linfocitos T a los epítopos codificados en las vacunas a partir de las PBMC, midiendo la inducción de citocinas, como IFN-gamma, utilizando un ensayo ELISpot. Es posible controlar las respuestas específicas de linfocitos T CD4 o CD8 a los epítopos codificados en las vacunas a partir de las PBMC, midiendo la inducción de citocinas capturadas a nivel intracelular o extracelular, como IFN-gamma, utilizando un ensayo de citometría de flujo. Las respuestas específicas de linfocitos T CD4 o CD8 a los epítopos codificados en las vacunas pueden controlarse a partir de PBMC midiendo las poblaciones de linfocitos T con expresión de receptores de linfocitos T específicos para los epítopos/complejos MHC de clase I, mediante tinción multimérica de MHC. Las respuestas específicas de linfocitos T CD4 o CD8 a los epítopos codificados en las vacunas pueden controlarse a partir de PBMC midiendo la expansión de poblaciones de linfocitos T ex vivo tras la incorporación de 3H-timidina, bromodeoxiuridina y carboxifluoresceína-diacetato–succinimidiléster (CFSE). La capacidad de reconocimiento de antígenos y la actividad lítica de linfocitos T derivados de PBMC específicos con respecto a epítopos codificados en las vacunas pueden evaluarse de forma funcional a través de un ensayo de liberación de cromo o ensayos colorimétricos alternativos de citotoxicidad.

VIII. Identificación de neoantígenos

VIII.A. Identificación de candidatos neoantigénicos

[00473] Se han descrito y aplicado métodos de investigación para análisis NGS de tumor y transcriptomas y exomas normales en el espacio de identificación de neoantígenos. [6,14,15] El ejemplo posterior considera determinadas optimizaciones para una mayor sensibilidad y especificidad con respecto a la identificación de neoantígenos en el entorno clínico. Estas optimizaciones pueden agruparse en dos áreas, las relacionadas con los procesos de laboratorio y las relacionadas con el análisis de datos de NGS.

VIII.A.1. Optimizaciones de procesos de laboratorio

[00474] Las mejoras de los procesos que se establecen en la presente abordan dificultadas en el hallazgo de neoantígenos con gran exactitud a partir de muestras clínicas con bajo contenido tumoral y pequeños volúmenes mediante la ampliación de conceptos desarrollados para la evaluación confiable de genes impulsores del cáncer en paneles de cáncer objetivo[16] y la configuración de transcriptoma y exoma completos necesaria para la identificación de neoantígenos. De forma específica, estas mejoras incluyen:

1. Direccionamiento de una cobertura promedio exclusiva profunda (>500x) en todo el exoma tumoral, para detectar mutaciones presentes a una baja frecuencia de alelos mutantes debido al contenido tumoral bajo o el estado de subclonación.

2. Direccionamiento de una cobertura uniforme en todo el exoma tumoral, con <5 % de bases abarcadas a <100x, de modo tal de evitar la menor cantidad de neoantígenos posible, por ejemplo:

a. utilizando sondas de captura basadas en ADN con la sonda individual QC[17]

b. incluyendo señuelos adicionales para regiones de mala cobertura

3. dirigiendo una cobertura uniforme en todo el exoma normal, en el que <5 % de bases se encuentren abarcadas a <20x, para contar con la menor cantidad posible de neoantígenos sin clasificación de estado somático/de línea germinal (y, por lo tanto, que no se puedan utilizar como TSNA)

4. Para minimizar la cantidad total de secuenciación necesaria, se diseñan sondas de captura de secuencias para codificar regiones de genes únicamente, dado que el ARN no codificante no puede dar lugar a neoantígenos. Otras optimizaciones incluyen:

a. sondas complementarias para genes de HLA ricos en GC y con baja captura en la secuenciación de exomas estándar[18]

b. exclusión de genes que se prevé que generen poca o nula cantidad de neoantígenos candidatos, debido a factores como expresión insuficiente, digestión subóptima por parte del proteasoma o características de secuencia inusuales.

5. De la misma forma, se secuencia el ARN tumoral a gran profundidad (lecturas de >100 M) para permitir la detección de variantes, cuantificación de genes y expresión de variantes de corte y empalme («isoformas»), así como detección de fusión. Las muestras de ARN de FFPE se extraen utilizando enriquecimiento a base de sondas[19], con sondas iguales o similares a las utilizadas para la captura de exomas en el ADN.

VIII.A.2. Optimizaciones de análisis de datos de NGS

[00475] Las mejoras en los métodos de análisis abordan la especificidad y sensibilidad subóptimas de los enfoques de investigación habituales de lectura de mutaciones y, de forma específica, consideran adaptaciones pertinentes para la identificación de neoantígenos en el entorno clínico. Estas incluyen:

1. Uso del genoma humano de referencia HG38 o una versión posterior para alineación, dado que este contiene múltiples ensamblajes de regiones de MHC que reflejan más correctamente el polimorfismo de la población, a diferencia de las liberaciones de genomas anteriores.

2. Superación de las limitaciones de lectores de variantes individuales[20] aunando los resultados de diferentes programas[5]

a. Es posible detectar indeles y variantes de nucleótidos individuales a partir de ADN tumoral, ARN tumoral y ADN normal con un conjunto de herramientas que incluyen: programas basados en comparaciones de ADN tumoral y normal, como Strelka [21] y Mutect [22]; y programas que incorporan ADN tumora, ARN tumoral y ADN normal, como UNCeqR, lo cual resulta particularmente ventajoso en

muestras de baja pureza[23].

b. Los indeles se determinan con programas que realizan un reensamblaje local, como Strelka y ABRA [24].

c. Las reorganizaciones estructurales se determinan utilizando herramientas específicas, como Pindel [25] o Breakseq [26].

3. Para detectar y prevenir los cambios de muestras, se comparan los lectores de variantes de las muestras para el mismo paciente en una cantidad seleccionadas de puntos polimórficos.

4. Por ejemplo, se lleva a cabo una extensa filtración de lectores distorsionados, mediante:

a. eliminación de variantes halladas en ADN normal, posiblemente con parámetros de detección relajados en casos de baja cobertura y con un criterio de proximidad permisivo en el caso de indeles

b. eliminación de variantes debido a una baja calidad de correspondencia o baja calidad base[27].

c. Eliminación de variantes derivadas de distorsiones recurrentes en la secuenciación, incluso en caso de no observarse en la forma normal correspondiente[27]. Los ejemplos incluyen variantes que se detectan principalmente en una cadena.

d. Eliminación de variantes detectadas en un conjunto de controles sin relación[27]

5. Lectura exacta de HLA a partir de exomas normales, mediante uno de seq2HLA [28], ATHLATES [29] u Optitype, junto con combinación de datos de secuencia de ARN y exomas[28]. Posibles optimizaciones adicionales incluyen implementación de un ensayo específico de tipificación de HLA, como secuenciación de ADN de lectura prolongada[30], o adaptación de un método de unión de fragmentos de ARN para mantener la continuidad[31].

6. Se realiza una fuerte detección de neo-ORF a partir de variantes de corte y empalme específicas del tumor mediante ensamblaje de transcripciones de datos de secuenciación de ARN con CLASS [32], Bayesembler [33], StringTie [34] o un programa similar en su modo con guía de referencia (es decir, utilizando estructuras de transcripción conocidas en lugar de intentar recrear las transcripciones por completo a partir de cada experimento). Si bien suele utilizarse Cufflinks [35] para tales fines, este muchas veces produce cantidades inverosímilmente considerables de variantes de corte y empalme, muchas de ellas mucho más cortas que el gen de longitud completa, y es posible que no logre extraer controles positivos simples. Las secuencias de codificación y el potencial de degradación mediada por mutaciones terminadoras se determinan con herramientas como SpliceR[36] y MAMBA[37], con reintroducción de secuencias mutantes. La expresión génica se determina con una herramienta como Cufflinks[35] o Express (Roberts y Pachter, 2013). Los niveles relativos y/o conteos de expresión específica de mutantes y tipo salvaje se determinan con herramientas creadas a tales efectos, como ASE[38] o HTSeq[39]. Las posibles etapas de filtración incluyen:

a. Eliminación de neo-ORF candidatos que se consideran con expresión insuficiente.

b. Eliminación de neo-ORF candidatos que se prevé que desencadenen una degradación mediada por mutaciones terminadoras (NMD).

7. Los antígenos candidato observados únicamente en ARN (p. ej., neoORF) que no pueden verificarse de forma directa como específicos del tumor se separan en una categoría de probabilidad de especificidad tumoral según parámetros adicionales, por ejemplo, considerando lo siguiente:

a. presencia de mutaciones en el punto de corte y empalme o cambio de marco con actuación en *cis* exclusiva de ADN tumoral que actúen de confirmación

b. presencia de mutación con actuación en *trans* exclusiva de ADN tumoral en un factor de corte y empalme que actúe como verificación. Por ejemplo, en tres experimentos publicados de forma independiente con SF3B1 con mutación R625, los genes que presentaban el corte y empalme más diferencial fueron concordantes incluso pese a que un experimento examinaba pacientes con melanoma uveal[40], el segundo una línea celular de melanoma uveal[41] y el tercero pacientes con cáncer de mama[42].

c. Para las isoformas novedosas de corte y empalme, presencia de lecturas de unión de corte y empalme «novedosas» en los datos de secuenciación de ARN que actúa como verificación.

d. Para reorganizaciones novedosas, presencia de lecturas de yuxta-exón en el ADN tumoral ausentes en el ADN normal que actúe como verificación.

e. Ausencia de un compendio de expresión génica como GTEx[43] (es decir, menor probabilidad de origen de línea germinal)

8. Complemento del análisis basado en alineación de genomas de referencia con comparación de lecturas normales y tumoral es de ADN ensambladas (o k-meros de dichas lecturas) de forma directa para evitar distorsiones y errores basados en alineaciones y anotaciones. (p. ej., para variantes somáticas que surjan cerca de variantes de línea germinal o indeles en contexto repetido).

[00476] En las muestras con ARN poliadenilado, se evalúa la presencia de ARN viral y microbiano en los datos de secuenciación de ARN utilizando RNA CoMPASS[44] o un método similar, para identificar factores adicionales que puedan pronosticar la respuesta del paciente.

VIII.B. Aislamiento y detección de péptidos de HLA

[00477] Se realizó aislamiento de moléculas peptídicas de HLA con métodos clásicos de inmunoprecipitación (IP) tras el análisis y la solubilización de la muestra de tejido (55-58). Se utilizó un lisado aclarado para IP específica de HLA.

[00478] La inmunoprecipitación se realizó con anticuerpos acoplados a microesferas, con el anticuerpo específico para moléculas de HLA. Para una inmunoprecipitación de pan-HLA de clase I, se utilizó un anticuerpo pan-CR de clase I, para HLA-DR de clase II, se utilizó un anticuerpo HLA-DR. El anticuerpo se enlaza de forma covalente a microesferas de sefarosa-NHS durante la incubación nocturna. Tras el enlace covalente, se lavaron las microesferas y se separaron en alícuotas para la IP.

(59, 60). También pueden llevarse a cabo inmunoprecipitaciones con anticuerpos que no están unidos de forma covalente a esferas. Típicamente, esto se realiza mediante el uso de sefarosa o esferas magnéticas recubiertas con Proteína A y/o Proteína G para mantener al anticuerpo en la columna. Se enumeran más adelante algunos anticuerpos que pueden usarse para enriquecer de forma selectiva el complejo de MHC/péptido.

Nombre del anticuerpo	Especificidad
W6/32	HLA-A, B, C clase I
L243	Clase II – HLA-DR
Tu36	Clase II – HLA-DR
LN3	Clase II – HLA-DR
Tu39	Clase II – HLA-DR, DP, DQ

[00479] Se agregó el lisado tisular aclarado a las microesferas de anticuerpo para la inmunoprecipitación. Tras la inmunoprecipitación, se retiraron las microesferas del lisado y se almacenó dicho lisado para experimentos adicionales, incluidas IP adicionales. Las microesferas de IP se lavaron para eliminar la unión no específica y se realizó elución del complejo de HLA/péptido de las microesferas con técnicas estándar. Se retiraron los componentes proteicos de los péptidos utilizando una columna de centrifugado de peso molecular o fraccionamiento C18. Los péptidos resultantes se llevan a sequedad con evaporación en SpeedVac y, en algunos casos, se almacenan a -20 C antes del análisis de MS.

[00480] Los péptidos secos se reconstituyen en un amortiguador de HPLC adecuado para cromatografía de fase inversa y se cargan en una columna de HPLC microcapilar C-18 para elución de gradiente en un espectrómetro de masas Fusion Lumos (Thermo). Se recopilaron los espectros de MS1 de masa/carga de los péptidos (m/z) en el detector Orbitrap a resolución elevada, seguido por lecturas de resolución baja MS2 recopilados en el detector con trampa iónica tras la fragmentación HCD del ion seleccionado. Asimismo, se pueden obtener espectros de MS2 utilizando métodos de fragmentación CID o ETD o cualquier combinación de las tres técnicas, para obtener una mayor cobertura de aminoácidos del péptido. Los espectros de mS2 también pueden medirse con exactitud de masa de alta resolución en el detector Orbitrap.

[00481] Los espectros de MS2 de cada análisis se buscan en una base de datos de proteínas mediante Comet (61, 62) y se califica la identificación de péptidos utilizando Percolator (63-65). Se lleva a cabo una secuenciación adicional mediante el uso de PEAKS studio (Bioinformatics Solutions Inc.) y pueden usarse otros motores de búsqueda o métodos de secuenciación que incluyen coincidencia espectral y secuenciación *de novo* (97).

VIII.B.1. Límite en MS de los estudios de detección para avalar la secuenciación generalizada de péptidos de HLA.

[00482] Utilizando el péptido YVYVADVAAK, se determinaron los límites de detección con cantidades diferentes de péptido cargado en la columna de LC. Las cantidades de péptido analizadas fueron de 1 pmol, 100 fmol, 10 fmol, 1 fmol y 100 amol. (Tabla 1) Los resultados se muestran en la figura 1F. Estos resultados indican que el límite inferior de detección (LoD, por sus siglas en inglés) se encuentra en el intervalo atomolar (10^{-18}), que el intervalo dinámico abarca cinco órdenes de magnitud y que la señal a ruido parece suficiente para la secuenciación en intervalos femtomolares (10^{-15}).

Tabla 1

m/z del péptido	cargado en columna	Copias/célula en células 1e9
566,830	1 pmol	600
562,823	100 fmol	60
559,816	10 fmol	6
556,810	1 fmol	0,6
553,802	100 amol	0,06

IX. Modelo de presentación

IX.A. Descripción general del sistema

[00483] La figura 2A es una vista general de un entorno 100 para identificar las probabilidades de presentación de péptidos en pacientes, de acuerdo con una modalidad. El entorno 100 proporcionar el contexto para la introducción de un sistema de identificación de presentación 160, que incluye en sí mismo un almacenamiento de información de presentación 165.

[00484] El sistema de identificación de presentación 160 es un modelo informático representado en un sistema informático como se mencionó anteriormente con respecto a la figura 14, el cual recibe secuencias peptídicas asociadas con un conjunto de alelos de MHC y determina las probabilidades de que uno o más del conjunto de alelos de MHC asociados presenten las secuencias peptídicas. El sistema de identificación de presentación 160 puede aplicarse a alelos de MHC tanto clase I como clase II. Esto resulta útil en diversos contextos. Un caso de uso específico del sistema de identificación de presentación 160 es la capacidad de recibir secuencias de nucleótidos de neoantígenos candidato con un conjunto de alelos de MHC de las células tumorales de un paciente 110 y determinar las probabilidades que uno o más de los alelos de MHC asociados del tumor presenten los neoantígenos candidato y/o inducir respuestas inmunogénicas en el sistema inmunológico del paciente 110. Los

neoantígenos candidato con mayores probabilidades, según determine el sistema 160, pueden seleccionarse para inclusión en una vacuna 118, y dicha respuesta inmunitaria antitumoral puede ser provocada por el sistema inmunológico del paciente 110 que proporcione las células tumorales.

[00485] El sistema de identificación de presentación 160 determina las probabilidades de presentación en uno o más modelos de presentación. De forma específica, los modelos de presentación generan probabilidades de que un conjunto de alelos de MHC asociados presente secuencias peptídicas específicas, y se generan en función de la información de presentación almacenada 165. Por ejemplo, los modelos de presentación pueden generar probabilidades de que el conjunto de alelos HLA-A*02:01, HLA-A*03:01, HLA-B*07:02, HLA-B*08:03, HLA-C*01:04, HLA-A*06:03, HLA-B*01:04 en la superficie celular de la muestra presente una secuencia peptídica «YVYVADVAAK (SEQ ID NO: 59)». La información de presentación 165 contiene información sobre si los péptidos se unen a diferentes tipos de alelos de MHC, de modo tal que los alelos de MHC presenten dichos péptidos, lo cual se determina en los modelos dependiendo de las posiciones de los aminoácidos en las secuencias peptídicas. El modelo de presentación puede prever la presentación de una secuencia peptídica no reconocida en asociación con un conjunto de alelos de MHC asociados, en función de la información de presentación 165. Tal como se mencionó anteriormente, los modelos de presentación pueden aplicarse a alelos de MHC tanto clase I como clase II.

IX.B. Información de presentación

[00486] La figura 2 ilustra un método para obtener información de presentación de acuerdo con una modalidad. La información de presentación 165 incluye dos categorías generales de información: información sobre interacción con alelos e información sobre ausencia de interacción con alelos. La información sobre interacción con alelos incluye información que afecta la presentación de secuencias peptídicas dependientes del tipo de alelo de MHC. La información sobre ausencia de interacción con alelos incluye información que afecta la presentación de secuencias peptídicas independientes del tipo de alelo de MHC.

IX.B.1. Información sobre interacción con alelos

[00487] La información sobre interacción con alelos incluye principalmente secuencias peptídicas que se sabe que se presentan en una o más moléculas de MHC identificadas de humanos, ratones, etc. Cabe señalar que esto puede incluir o no datos obtenidos de muestras tumorales. Las secuencias peptídicas que se presentan pueden identificarse a partir de células con expresión de un único alelo de MHC. En este caso, las secuencias peptídicas presentadas suelen recolectarse de líneas celulares de alelos individuales modificados para expresar un alelo de MHC predeterminado y que se exponen posteriormente a una proteína sintética. Los péptidos que se presentan en el alelo de MHC se aíslan utilizando técnicas como elución en ácido e identificación a través de espectrometría de masas. La figura

2B muestra un ejemplo de esto, en el que el ejemplo de péptido YEMFNDKS (SEQ ID NO: 60), que se presenta en el alelo MHC predeterminado HLA-A*01:01, se aísla e identifica a través de espectrometría de masas. La FIG. 2D muestra otro ejemplo de esto, donde el péptido representativo YEMFNDKSQRAPDDKMF (SEQ ID NO: 61), presentado en el alelo de MHC predeterminado HLA-DRB1*12:01, se aísla e identifica a través de espectrometría de masas. Dado que en esta situación se identifican péptidos a través de células modificadas para expresión de una única proteína de MHC predeterminada, se conoce de forma definitiva la asociación directa entre un péptido presentado y la proteína de MHC a la cual se enlaza.

[00488] Las secuencias peptídicas que se presentan también pueden recolectarse de células con expresión de múltiples alelos de MHC. Generalmente, en seres humanos se expresan 6 tipos diferentes de moléculas de MHC-I y hasta 12 tipos diferentes de MHC-II para una célula. Estas secuencias peptídicas presentadas pueden identificarse a partir de líneas celulares con alelos múltiples que se modifican para expresar múltiples alelos de MHC predeterminados. Estas secuencias peptídicas presentadas también pueden identificarse a partir de muestras de tejido, ya sea de muestras de tejido normal o muestras de tejido tumoral. En este caso particular, las moléculas de MHC pueden someterse a inmunoprecipitación de tejido normal o tumoral. Los péptidos que se presentan en los múltiples alelos de MHC se pueden aislar de forma similar utilizando técnicas como elución en ácido e identificación a través de espectrometría de masas. La figura 2C muestra un ejemplo de esto, en el que los seis ejemplos de péptidos YEMFNDKSF (SEQ ID NO: 62), HROEIFSHDFJ (SEQ ID NO: 63), FJIEJFOESS (SEQ ID NO: 64), NEIOREIREI (SEQ ID NO: 65), JFKSIFEMMSJDSSU (SEQ ID NO: 66), and KNFLENFIESOFI (SEQ ID NO: 67), se presentan en los alelos de MHC identificados HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, HLA-C*01:03 y HLA-C*01:04 y se aíslan e identifican a través de espectrometría de masas. En otro ejemplo, la FIG. 2C muestra donde los seis péptidos ilustrativos, YEMFNDKSF (SEQ ID NO: 62), HROEIFSHDFJ (SEQ ID NO: 63), FJIEJFOESS (SEQ ID NO: 64), NEIOREIREI (SEQ ID NO: 65), JFKSIFEMMSJDSSUIFLKSJFIEIFJ (SEQ ID NO: 68), y KNFLENFIESOFI (SEQ ID NO: 67), se presentan en alelos del MHC clase I identificados HLA-A*01:01, HLA-A*02:01, HLA-B*07:02, HLA-B*08:01, y alelos del MHC clase II identificados HLA-DRB1*10:01, HLA-DRB1:11:01y se aíslan e identifican a través de espectrometría de masas. A diferencia de las líneas celulares de alelos individuales, en estos ejemplos puede desconocerse la asociación directa entre un péptido que se presente y la proteína de MHC a la cual se enlace, dado que los péptidos enlazados se aíslan de las moléculas de MHC antes de la identificación.

[00489] La información sobre interacción con alelos también puede incluir corriente iónica en espectrometría de masas, que depende tanto de la concentración de complejos de moléculas MHC-péptidos como la eficacia de ionización de péptidos. La eficacia de ionización varía entre los péptidos de

forma dependiente de la secuencia. En términos generales, la eficacia de ionización varía entre los péptidos en aproximadamente dos órdenes de magnitud, mientras que la concentración de los complejos de MHC-péptido varía en un espectro mayor.

[00490] La información sobre interacción con alelos también incluye mediciones o pronósticos de afinidad de unión entre un alelo de MHC específico y un péptido específico (94, 95, 96). Uno o más modelos de afinidad pueden generar dichos pronósticos. Por ejemplo, volviendo al ejemplo de la figura 1D, la información de presentación 165 puede incluir un pronóstico de afinidad de unión de 1000 nM entre el péptido YEMFNDKSF (SEQ ID NO: 62), y el alelo HLA-A*01:01 clase I. El MHC presenta pocos péptidos con una IC50 > 1000 nm y los menores valores de IC50 aumentan la probabilidad de presentación. La información de presentación 165 puede incluir una predicción de afinidad de unión entre el péptido KNFLENFIESOFI (SEQ ID NO: 67), y el alelo clase II HLA-DRB1:11:01.

[00491] La información sobre interacción con alelos también puede incluir mediciones o pronósticos de estabilidad del complejo de MHC. Uno o más modelos de estabilidad pueden generar dichos pronósticos. Hay una mayor probabilidad de presentación de complejos de MHC-péptidos más estables (es decir, complejos con semividas más prolongadas) a una mayor cantidad de copias en células tumorales y en células con presentación de antígeno que tengan contacto con el antígeno de la vacuna. Por ejemplo, volviendo al ejemplo de la figura 2C, la información de presentación 165 puede incluir un pronóstico de estabilidad de semivida de 1 h para la molécula clase I HLA-A*01:01. La información de presentación 165 puede incluir también una predicción de estabilidad de una semivida para la molécula clase II HLA-DRB1:11:01.

[00492] La información sobre interacción con alelos también puede incluir la tasa medida o prevista de reacción de formación para el complejo de MHC-péptido. Existe una mayor probabilidad de presentación de los complejos que se forman a una tasa mayor en la superficie celular a concentración elevada.

[00493] La información sobre interacción con alelos también puede incluir la secuencia y longitud del péptido. Generalmente se prefieren moléculas de MHC de clase I en relación con péptidos presentes con longitudes de entre 8 y 15 péptidos. El 60 al 80 % de los péptidos que se presentan tienen una longitud 9. En la figura 5 se muestran histogramas de longitudes de péptidos presentados de diversas líneas celulares. Las moléculas clase II típicamente prefieren presentar péptidos con longitudes entre 6-30 péptidos.

[00494] La información sobre interacción con alelos también puede incluir la presencia de motivos de secuencias de cinasa en el péptido codificado por el neoantígeno, así como la ausencia o presencia de modificaciones postraducción específicas en el péptido codificado por el neoantígeno. La presencia de motivos de cinasa afecta la probabilidad de modificación postraducción, lo cual puede potenciar la unión

del MHC o interferir con ella.

[00495] La información sobre interacción con alelos también puede incluir los niveles de actividad o expresión de proteínas involucradas en el proceso de modificación postraducción, p. ej., cinasas (según la medición o el pronóstico a partir de la secuenciación de ARN, espectrometría de masas u otros métodos).

[00496] La información sobre interacción con alelos también puede incluir la probabilidad de presentación de péptidos con secuencia similar en células de otros individuos que expresen el alelo de MHC particular, según la evaluación con proteónica de espectrometría de masas u otros medios.

[00497] La información sobre interacción con alelos también puede incluir los niveles de expresión del alelo de MHC particular en el individuo particular (p. ej., según la medición con secuenciación de ARN o espectrometría de masas). Existe una mayor probabilidad de presentación de péptidos que se unen con mayor fuerza a un alelo de MHC expresado a mayores niveles con respecto a péptidos que se unen con mayor fuerza a un alelo de MHC expresado a menor nivel.

[00498] La información sobre interacción con alelos también puede incluir la probabilidad global de presentación independiente de la secuencia peptídica codificada por neoantígenos para el alelo de MHC particular en otros individuos con expresión del alelo de MHC particular.

[00499] La información sobre interacción con alelos también puede incluir la probabilidad global de presentación independiente de la secuencia peptídica para alelos de MHC en la misma familia de moléculas (p. ej., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) en otros individuos. Por ejemplo, las moléculas de HLA-C suelen expresarse en menores niveles que las moléculas de HLA-A o HLA-B y, por lo tanto, la presentación de un péptido en HLA-C en principio es menos probable que la presentación en HLA-A o HLA-B. Como otro ejemplo, HLA-DP típicamente se expresa a niveles menores que HLA-DR o HLA-DQ; por consiguiente, la presentación de un péptido mediante HLA-DP es una presentación previa menos probable que la presentación mediante HLA-DR o HLA-DQ.

[00500] La información sobre interacción con alelos también puede incluir la secuencia de proteína del alelo de MHC particular.

[00501] También se puede realizar un modelo de cualquier información sobre ausencia de interacción con alelos de MHC mencionada en la sección posterior como información sobre interacción con alelos de MHC.

IX.B.2. Información sobre ausencia de interacción con alelos

[00502] La información sobre ausencia de interacción con alelos puede incluir secuencias de extremo C que flanqueen el péptido codificado por el neoantígeno en su secuencia proteica original. Para MHC-I, las secuencias flanquantes de extremo C pueden afectar el procesamiento proteosómica de los péptidos. No obstante, la secuencia flanqueante de extremo C se escinde del péptido mediante el

proteasoma antes del transporte del péptido hacia el retículo endoplásmico y se encuentra con los alelos de MHC en las superficies de las células. Por consiguiente, las moléculas de MHC no reciben información sobre la secuencia flanqueante de extremo C y, por lo tanto, el efecto de la secuencia flanqueante de extremo C no puede variar dependiendo del tipo de alelo de MHC. Por ejemplo, volviendo al ejemplo de la figura 2C, la información de presentación 165 puede incluir la secuencia flanqueante de extremo C FOEIFNDKSLDKFJI (SEQ ID NO: 69), del péptido presentado FJIEJFOESS (SEQ ID NO: 64), que se identifica a partir de la proteína original del péptido.

[00503] La información sobre ausencia de interacción con alelos también puede incluir mediciones de cuantificación de ARNm. Por ejemplo, pueden obtenerse datos de cuantificación de ARNm para las mismas muestras que proporcione los datos para formación de espectrometría de masas. Tal como se describe posteriormente con respecto a la figura 13H, se identificó la expresión de ARN como un potente predictor de la presentación de péptidos. En una modalidad, las mediciones de cuantificación de ARNm se identifican a partir de la herramienta de software RSEM. La implementación detallada de la herramienta de software RSEM puede encontrarse en Bo Li y Colin N. Dewey. *RSEM: cuantificación de transcripción exacta a partir de datos de secuenciación de ARN con o sin genoma de referencia.* **BMC Bioinformatics**, 12:323, agosto de 2011. En una modalidad, la cuantificación de ARNm se mide en unidades de fragmentos por kilobase de transcripción por millones de lecturas con correspondencia (FPKM, en inglés).

[00504] La información sobre ausencia de interacción con alelos también puede incluir secuencias de extremo N que flanqueen el péptido en su secuencia proteica original.

[00505] La información de no interacción de alelos puede incluir también el gen de origen de la secuencia de péptidos. El gen de origen puede definirse como la familia de proteínas Ensembl de la secuencia de péptidos. En otros ejemplos, el gen de origen puede definirse como el ADN de origen o el ARN de origen de la secuencia de péptidos. Por ejemplo, el gen de origen puede representarse como una cadena de nucleótidos que codifican una proteína o, de manera alternativa, representarse de forma más categórica en función de un conjunto nombrado de secuencias de ADN o ARN conocidas que se sabe que codifican proteínas específicas. En otro ejemplo, la información de no interacción de alelos puede incluir también la transcripción o isoforma de origen o conjunto posible de transcripciones o isoformas de origen de la secuencia de péptidos extraídos de una base de datos tal como Ensembl o RefSeq.

[00506] La información sobre ausencia de interacción con alelos también puede incluir la presencia de motivos de escisión de proteasa en el péptido, opcionalmente ponderada de acuerdo con la expresión de proteasas correspondientes en las células tumorales (según la medición con secuenciación de ARN o espectrometría de masas). Existe una menor probabilidad de presentación de péptidos que contienen

motivos de escisión de proteasa, dado que estos se degradan con mayor rapidez en las proteasas y, por lo tanto, son menos estables en la célula.

[00507] La información de no interaccción con alelos puede incluir también la velocidad de rotación de la proteína fuente según lo medido en el tipo de célula adecuado. La velocidad de rotación más rápida (es decir, la semivida más baja) aumenta la probabilidad de presentación, sin embargo, la energía predictiva de esta característica es baja si se mide en un tipo de célula diferente.

[00508] La información de no interaccción con alelos puede incluir también la longitud de la proteína fuente, opcionalmente considerando las variantes de empalme específicas («isoformas») que se expresan a nivel más alto en las células tumorales según se mide mediante la espectrometría de masas de proteomas o secuenciación de RNA o según lo predice la anotación de la línea germinal o mutaciones de empalme somáticas detectadas en los datos de secuencia de RNA o DNA.

[00509] La información de no interaccción con alelos puede incluir también el nivel de expresión de la proteasoma, la inmunoproteasoma, la timoproteasoma u otras proteasas en las células tumorales (que se pueden medir mediante secuenciación de RNA, espectrometría de masas de proteomas o inmunohistoquímica). Diferentes proteasomas tienen diferentes preferencias de sitio de escisión. Se dará más peso a las preferencias de escisión de cada tipo de proteasoma en proporción a su nivel de expresión.

[00510] La información de no interaccción con alelos puede incluir también la expresión del gen fuente del péptido (por ejemplo, según lo medido mediante secuenciación de RNA o espectrometría de masas). Las posibles optimizaciones incluyen ajustar la expresión medida para justificar la presencia de células estromales y los linfocitos que infiltrantes de tumores dentro de la muestra tumoral. Es más probable que se presenten los péptidos de genes expresados a nivel más alto. Los péptidos de genes con niveles no detectables de expresión se pueden excluir de la consideración.

[00511] La información de no interaccción con alelos puede incluir también la probabilidad de que el RNAm fuente del péptido codificado por neoantígenos se someterá a degradación mediada por mutaciones terminadoras como se predice por medio de un modelo de degradación mediada por mutaciones terminadoras, por ejemplo, el modelo de Rivas et ál, Science 2015.

[00512] La información que interactúa con alelos puede incluir también la expresión específica del tejido típica del gen fuente del péptido durante las diversas etapas del ciclo celular. Los genes que se expresan a un nivel bajo general (según lo medido por la secuenciación de RNA o proteómica de espectrometría de masas) pero que se sabe que se expresan a un nivel alto durante las etapas específicas del ciclo celular producen probablemente más péptidos presentados que genes que se expresan establemente en niveles muy bajos.

[00513] La información que interactúa con alelos puede incluir también un catálogo completo de

características de la proteína fuente como se da en, por ejemplo, uniProt o PDB http://www.rcsb.org/pdb/home/home.do. Estas características pueden incluir, entre otros: las estructuras secundarias y terciarias de la proteína, localización subcelular 11, términos de ontología génica (GO). Específicamente, esta información puede contener anotaciones que actúan a nivel de la proteína, por ejemplo, longitud de la UTR 5' y las anotaciones que actúan a nivel de los residuos específicos, por ejemplo, motivo de hélice entre los residuos 300 y 310. Estas características pueden incluir también motivos de giro, motivos de lámina y residuos desordenados.

[00514] La información de no interaccción con alelos puede incluir también características que describen las propiedades del dominio de la proteína fuente que contiene el péptido, por ejemplo: estructura secundaria o terciaria (por ejemplo, hélice alfa vs. Lámina beta); empalme alternativo.

[00515] La información que interactúa con alelos puede incluir también características que describen la presencia o ausencia de un punto de acceso de presentación en la posición del péptido en la proteína fuente del péptido.

[00516] La información de no interaccción con alelos puede incluir también la probabilidad de la presentación de péptidos de la proteína fuente del péptido en cuestión en otros individuos (después de ajustar el nivel de expresión de la proteína fuente en aquellos individuos y la influencia de los tipos HLA diferentes de esos individuos).

[00517] La información de no interaccción con alelos puede incluir también la probabilidad de que el péptido no sea detectado o se sobrerepresente mediante espectrometría de masas debido a inclinaciones técnicas.

[00518] La expresión de diversos módulos/vías génicas según lo medido a través de un ensayo de expresión génica tales como secuenciación de RNA, microarreglo/s, panel/es dirigido/s tal como Nanostring o representantes de genes únicos/múltiples de módulos génicos medidos mediante ensayos tales como RT-PCR (que no necesita contener la proteína fuente del péptido) que son informativos sobre el estado de células tumorales, estroma o linfocitos infiltrantes de tumores (TIL).

[00519] La información de no interaccción con alelos puede incluir también la cantidad de copias del gen fuente del péptido en las células tumorales. Por ejemplo, a los péptidos de genes que se someten a eliminación homocigota en células tumorales se les pueden asignar una probabilidad de presentación de cero.

[00520] La información de no interaccción con alelos puede incluir también la probabilidad de que el péptido se una al TAP o la afinidad de unión prevista o medida del péptido al TAP. Es más probable que se presenten péptidos que son más probables que se unan al TAP, o péptidos que se unen a TAP con afinidad más alta mediante MHC-I.

[00521] La información de no interaccción con alelos puede incluir también el nivel de expresión

de TAP en las células tumorales (que se puede medir mediante secuenciación de RNA, espectrometría de masas de proteomas, inmunohistoquímica). Para HMC-I, los niveles de expresión de TAP más altos aumentan la probabilidad de presentación de todos los péptidos.

[00522] La información de no interaccción con alelos puede incluir también la presencia o ausencia de mutaciones tumorales, que incluyen, de modo no taxativo:

i. Mutaciones conductoras en genes conductores del cáncer conocidos tales como EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3.

ii. En genes que codifican las proteínas implicadas en la maquinaria de presentación de antígenos (por ejemplo, B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 o cualquiera de los genes que codifican los componentes de la proteasoma o inmunoproteasoma). Los péptidos cuya presentación se basa en un componente de la maquinaria de presentación de antígenos que se somete a la mutación de pérdida de función en el tumor tiene menor probabilidad de presentación.

[00523] La presencia o ausencia de polimorfismos de línea germinal funcional, que incluyen, de modo no taxativo:

i. En genes que codifican las proteínas implicadas en la maquinaria de presentación de antígenos (por ejemplo, B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 o cualquiera de los genes que codifican los componentes de la proteasoma o inmunoproteasoma).

[00524] La información de no interaccción con alelos puede incluir también el tipo de tumor (por ejemplo, NSCLC, melanoma).

[00525] La información de no interaccción con alelos puede incluir también la funcionalidad conocida de alelos de HLA, como se refleja mediante, por ejemplo, sufijos de alelos de HLA. Por ejemplo, el sufijo N en el nombre del alelo HLA-A*24:09N indica un alelo nulo que no se expresa y, por lo tanto, es poco probable que presente epítopos; la nomenclatura de sufijos de alelos HLA se describe en https://www.ebi.ac.uk/ipd/imgt/hla/nomenclature/suffixes.html.

[00526] La información de no interaccción con alelos puede incluir también el subtipo de tumor (por ejemplo, cáncer de pulmón escamoso vs. el no escamoso).

[00527] La información de no interaccción con alelos puede incluir también el historial de tabaquismo.

[00528] La información de no interaccción con alelos puede incluir también el historial de

quemaduras de sol, la exposición solar o la exposición a otros mutágenos.

[00529] La información de no interaccción con alelos puede incluir también la expresión típica del gen fuente del péptido en el tipo de tumor pertinente o subtipo clínico, opcionalmente estratificado mediante mutación conductora. Es más probable que se presenten genes que se expresan típicamente a niveles altos en el tipo de tumor pertinente.

[00530] La información de no interaccción con alelos puede incluir también la frecuencia de la mutación en todos los tumores o en tumores del mismo tipo o en tumores de individuos con al menos un alelo MHC compartido o en tumores del mismo tipo en individuos con al menos un alelo MHC compartido.

[00531] En el caso de un péptido específico para el tumor mutado, la lista de características utilizadas para predecir una probabilidad de preservación puede incluir también la anotación de la mutación (por ejemplo, cambio de sentido, translectura, cambio de marco, fusión, etc.) o si la mutación se predice para que provoque degradación mediada por mutaciones terminadoras (NMD). Por ejemplo, a los péptidos de segmentos de proteína que no se traducen en las células tumorales debido a las mutaciones de detención temprana homocigotas se les puede asignar una probabilidad de presentación de cero. La NMD provoca traducción del RNAm disminuido, lo cual disminuye la probabilidad de presentación.

IX.C. Sistema de identificación de presentación

[00532] La figura 3 es un diagrama de bloques de nivel alto que ilustra los componentes de lógica informática del sistema de identificación de presentación 160, de acuerdo con una modalidad. En esta modalidad de ejemplo, el sistema de identificación de presentación 160 incluye un módulo de gestión de datos 312, un módulo de codificación 314, un módulo de entrenamiento 316 y un módulo de predicción 320. El sistema de identificación de presentación 160 se compone también de un almacenamiento de datos de entrenamiento 170 y un almacenamiento de modelos de presentación 175. Algunas modalidades del sistema de gestión de modelo 160 tienen módulos diferentes a los que se describen aquí. De manera similar, las funciones se pueden distribuir entre los módulos de una manera diferente a la que se describe aquí.

IX.C.1. Módulo de gestión de datos

[00533] El módulo de gestión de datos 312 genera conjuntos de datos de entrenamiento 170 desde la información de presentación 165. Cada conjunto de datos de entrenamiento contiene múltiples instancias de datos, en las que cada instancia de datos *i* contiene un conjunto de variables independientes $\boldsymbol{z}^i$ que incluyen al menos una secuencia peptídica presentada o no presentada $\boldsymbol{p}^i$**,** uno o más alelos MHC asociados $\boldsymbol{a}^i$ que se asocian con la secuencia peptídica $\boldsymbol{p}^i$, y una variable dependiente y^i que representa la información que el sistema de identificación de presentación 160 está interesado en

predecir para valores nuevos de variables independientes.

[00534] En una implementación particular a la que se hace referencia a lo largo de lo que resta de la memoria descriptiva, la variable dependiente y^i es un nivel binario que indica si el péptido $\boldsymbol{p}^i$ se presentó mediante el uno o más alelos MHC asociados $\boldsymbol{a}^i$. Sin embargo, se aprecia que en otras implementaciones, la variable dependiente $\boldsymbol{y}^i$ puede representar cualquier otro tipo de información que el sistema de identificación de presentación 160 está interesado en predecir dependiendo de las variables independientes $\boldsymbol{z}^i$. Por ejemplo, en otra implementación, la variante dependiente $\boldsymbol{y}^i$ también puede ser un valor numérico que indica la corriente de iones de espectrometría de masas identificada para la instancia de datos.

[00535] La secuencia peptídica $\boldsymbol{p}^i$ para la instancia de datos $\boldsymbol{i}$ es una secuencia de aminoácidos k_i, en la cual k_i puede variar entre las instancias de datos i dentro de un intervalo. Por ejemplo, ese intervalo puede ser de 8-15 para la clase I de MHC o 6-30 para la clase II de MHC. En una implementación específica del sistema 160, todas las secuencias peptídicas $\boldsymbol{p}^i$ en un conjunto de datos de entrenamiento pueden tener la misma longitud, por ejemplo, 6. La cantidad de aminoácidos en una secuencia peptídica puede variar dependiendo del tipo de alelos de MHC (por ejemplo, alelos de MHC en humanos, etc.). Los alelos de MHC $\boldsymbol{a}^i$ para la instancia de datos i indican qué alelos de MHC estaban presentes en asociación con la secuencia peptídica correspondiente $\boldsymbol{p}^i$**.**

[00536] El módulo de gestión de datos 312 puede incluir también variables que interactúan con alelos adicionales, tales como predicciones de afinidad de unión $\boldsymbol{b}^i$ y estabilidad $\boldsymbol{s}^i$ junto con las secuencias peptídicas $\boldsymbol{p}^i$ y alelos de MHC asociados $\boldsymbol{a}^i$ incluidos en los datos de entrenamiento 170. Por ejemplo, los datos de entrenamiento 170 pueden contener predicciones de afinidad de unión $\boldsymbol{b}^i$ entre un péptido $\boldsymbol{p}^i$ y cada una de las moléculas de MHC asociadas que se indican en $\boldsymbol{a}^i$. Como ejemplo adicional, los datos de entrenamiento 170 pueden contener predicciones de estabilidad $\boldsymbol{s}^i$ para cada uno de los alelos de MHC que se indican en $\boldsymbol{a}^i$.

[00537] El módulo de gestión de datos 312 puede incluir también variables que no interactúan con alelos $\boldsymbol{w}^i$, tal como secuencias flanqueantes de extremo C y mediciones de cuantificación de RNA junto con las secuencias peptídicas $\boldsymbol{p}^i$.

[00538] El módulo de gestión de datos 312 identifica también secuencias peptídicas que no son presentadas por los alelos de MHC para generar los datos de entrenamiento 170. Generalmente, esto implica identificar las secuencias «más largas» de la proteína fuente que incluye las secuencias peptídicas presentadas antes de la presentación. Cuando la información de presentación contiene líneas celulares modificadas genéticamente, el módulo de gestión de datos 312 identifica una serie de secuencias peptídicas en la proteína sintética a la cual se expusieron las células que no se presentaron en los alelos de MHC de las células. Cuando la información de presentación contiene muestras de tejido,

el módulo de gestión de datos 312 identifica proteínas fuente de las cuales se originaron secuencias peptídicas presentadas e identifica una serie de secuencias peptídicas en la proteína fuente que no se presentaron en alelos de MHC de las células de muestras de tejido.

[00539] El módulo de gestión de datos 312 puede generar también de forma artificial péptidos con secuencias aleatorias de aminoácidos e identifica las secuencias generadas como péptidos no presentados en alelos de MHC. Esto se puede lograr al generar de forma aleatoria secuencias peptídicas que permiten que el módulo de gestión de datos 312 genere fácilmente grandes cantidades de datos sintéticos para péptidos no presentados en alelos de MHC. Dado que en realidad un pequeño porcentaje de secuencias peptídicas son presentadas por alelos de MHC, es muy probable que las secuencias peptídicas generadas de forma sintética no hayan sido presentadas por alelos de MHC incluso si fueron incluidas en proteínas procesadas por células.

[00540] La FIG. 4A ilustra un conjunto de ejemplo de datos de entrenamiento 170A, de acuerdo con una modalidad. Específicamente, las 3 primeras instancias de datos en los datos de entrenamiento 170A indican información de presentación de péptidos de una línea celular de único alelo que implica el alelo HLA-C*01:03 y 3 secuencias peptídicas QCEIOWARE (SEQ ID NO: 70), FIEUHFWI (SEQ ID NO: 71), y FEWRHRJTRUJR (SEQ ID NO: 72). La cuarta instancia de datos en los datos de entrenamiento 170A indica la información peptídica de una línea celular con múltiples alelos que implica los alelos HLA-B*07:02, HLA-C*01:03, HLA-A*01:01y una secuencia peptídica QIEJOEIJE (SEQ ID NO: 73). La primera instancia de datos indica que la secuencia peptídica QCEIOWARE (SEQ ID NO: 70), no fue presentada por el alelo HLA-C*01:03. Como se trata en los dos párrafos anteriores, la secuencia peptídica puede generarse de forma aleatoria por medio del módulo de gestión de datos 312 o puede identificarse desde una proteína fuente de péptidos presentados. Los datos de entrenamiento 170A incluyen también una predicción de afinidad de unión de 1000nM y una predicción de estabilidad de una semivida de 1 hora para el par de alelo-secuencia peptídica. Los datos de entrenamiento 170A incluyen también variables que no interactúan con alelos, tal como la secuencia flanqueante de extremo C del péptido FJELFISBOSJFIE (SEQ ID NO: 74), y una medición de cuantificación de RNAm de 10^2 TPM. La cuarta instancia de datos indica que la secuencia peptídica QIEJOEIJE (SEQ ID NO: 73), no fue presentada mediante uno de los alelos HLA-B*07:02, HLA-C*01:03 o HLA-A*01:01. Los datos entrenamiento 170A incluyen también las predicciones de afinidad de unión y predicciones de estabilidad para cada uno de los alelos, así como también las secuencia flanqueante del extremo C del péptido y la medición de cuantificación de RNAm para el péptido.

[00541] La FIG. 4B ilustra otro conjunto de datos de entrenamiento ilustrativo 170A, de acuerdo con una modalidad. Específicamente, las primeras instancias de datos en los datos de entrenamiento 170A indican información de presentación de péptidos a partir de una línea celular de alelo simple que

implica el alelo de clase II HLA-DRB3:01:01 y la secuencia de péptidos QCEIOWAREFLKEIGJ (SEQ ID NO: 75). La primera instancia de datos indica que la secuencia de péptidos QCEIOWAREFLKEIGJ (SEQ ID NO: 75), no se presentó mediante el alelo HLA-DRB3:01:01.

IX.C.2. Módulo de codificación

[00542] El módulo de codificación 314 codifica la información incluida en los datos de entrenamiento 170 en una representación numérica que se puede usar para generar el uno o más modelos de presentación. En una implementación, el módulo de codificación 314 codifica secuencias *one-hot* (por ejemplo, secuencias peptídicas o secuencias flanqueante de extremo C) en un alfabeto de aminoácidos de 20 letras predeterminado. Específicamente, se representa una secuencia peptídica $\boldsymbol{p}^i$ con aminoácidos k_i como un vector de fila de elementos $20 \cdot k_i$, donde un único elemento entre $p^i_{20\cdot(j-1)+1}$, $p^i_{20\cdot(j-1)+2}$, …, $p^i_{20\cdot j}$ que corresponde al alfabeto del aminoácido en la posición *j-th* de la secuencia peptídica tiene un valor de 1. De otro modo, los elementos restantes tienen un valor de 0. Como ejemplo, para un alfabeto dado {A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y}, la secuencia peptídica EAF de 3 aminoácidos para la instancia de datos *i* puede ser representada por el vector de fila de 60 elementos $\boldsymbol{p}^i$ =[0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]. La secuencia flanqueante de extremo C $\boldsymbol{c}^i$ se puede codificar de forma similar a como se describió anteriormente, así como también las secuencias de proteínas $\boldsymbol{d}_h$ para alelos de MHC, y otros datos de secuencia en la información de presentación.

[00543] Cuando los datos de entrenamiento 170 contienen secuencias de diferentes longitudes de aminoácidos, el módulo de codificación 314 puede codificar además los péptidos en vectores de igual longitud al agregar un caracter PAD para extender el alfabeto predeterminado. Por ejemplo, esto se puede realizar al rellenar con ceros las secuencias peptídicas con un caracter PAD hasta que la longitud de la secuencia peptídica alcance la secuencia peptídica con la longitud más grande en los datos de entrenamiento 170. Por lo tanto, cuando la secuencia peptídica con la longitud más larga tiene aminoácidos k_{max}, el módulo de codificación 314 representa numéricamente cada secuencia como un vector de fila de elementos $(20+1)\cdot k_{max}$. Como ejemplo, para el alfabeto dado {PAD, A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y} y una longitud de aminoácidos máxima de $k_{max}=5$, la misma secuencia peptídica de ejemplo EAF de 3 aminoácidos se puede representar por medio del vector de fila de 105 elementos $\boldsymbol{p}^i$ =[1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]. La secuencia flanqueante de extremo C $\boldsymbol{c}^i$ u otros datos de secuencia se pueden codificar de forma similar como se describió anteriormente. Por lo tanto, cada variable o columna independiente en la secuencia peptídica $\boldsymbol{p}^i$ o $\boldsymbol{c}^i$ representa la presencia de un aminoácido particular en una posición

particular de la secuencia.

[00544] Aunque el método anterior de datos de secuencia codificadora se describió en referencia a las secuencias que tienen secuencias de aminoácidos, el método puede extenderse de manera similar a otros tipos de datos de secuencia, tales como datos de secuencia de DNA o RNA y similares.

[00545] El módulo de codificación 314 codifica también los uno o más alelos de MHC $\boldsymbol{a}^i$ para la instancia de datos *i* como un vector de fila de elementos *m*, en el que cada elemento *h=1, 2, …, m* corresponde a un único alelo de MHC identificado. Los elementos que corresponden a los alelos de MHC identificados para la instancia de datos *i* tienen un valor de 1. De otro modo, los elementos restantes tienen un valor de 0. Como ejemplo, los alelos HLA-B*07:02 y HLA-C*01:03 para una instancia de datos *i* que corresponde a una línea celular de múltiples alelos entre *m=4* tipos de alelos de MHC identificados únicos {HLA-A*01:01, HLA-C*01:08, HLA-B*07:02, HLA-C*01:03} pueden ser representados por el vector de fila de 4 elementos $\boldsymbol{a}^i$=[0 0 1 1], en el cual a_3^i=1 y a_4^i=1. A modo de otro ejemplo, los elementos correspondientes a los alelos de MHC identificados para la instancia de datos *i* tienen un valor de 1. Por el contrario, los elementos restantes tienen un valor de 0. A modo de ejemplo, los alelos HLA-B*07:02 y HLA-DRB1*10:01 para una instancia de datos *i* correspondiente a líneas celulares de múltiples alelos entre *m=4* tipos de alelos de MHC únicos identificados {HLA-A*01:01, HLA-C*01:08, HLA-B*07:02, HLA-DRB1*10:01 } pueden representarse mediante el vector fila de 4 elementos $\boldsymbol{a}^i$=[0 0 1 1], en la cual a_3^i=1 y a_4^i=1. Aunque el ejemplo se describe en la presente con 4 tipos de alelos de MHC identificados, la cantidad de tipos de alelos de MHC clase I puede ser de cientos o miles en la práctica. Como se trató previamente, cada instancia de datos *i* contiene típicamente a lo sumo 6 tipos de alelos de MHC diferentes en asociación con la secuencia peptídica $\boldsymbol{p}_i$ y/o un máximo de 4 tipos distintos de alelos del MHC clase II DR asociados con la secuencia peptídica $\boldsymbol{p}_i$ y/o un máximo de 12 tipos distintos de alelos del MHC clase II asociados con la secuencia peptídica $\boldsymbol{p}_i$.

[00546] El módulo de codificación 314 codifica también la etiqueta y_i para cada instancia de datos *i* como una variable binaria que tiene valores del conjunto de {0, 1}, en el que un valor de 1 indica que el péptido $\boldsymbol{x}^i$ se presentó por medio de uno de los alelos de MHC asociados $\boldsymbol{a}^i$, y un valor de 0 indica que el péptido $\boldsymbol{x}^i$ no fue presentado por ninguno de los alelos de MHC asociado $\boldsymbol{a}^i$. Cuando la variable dependiente y_i representa la corriente de iones de espectrometría de masas, el módulo de codificación 314 puede escalar adicionalmente los valores usando diversas funciones, tal como la función log que tiene un intervalo de [-∞, ∞] para valores actuales de iones entre [0, ∞].

[00547] El módulo de codificación 314 puede representar un par de variables que interactúan con alelos $\boldsymbol{x}_h^i$ f para el péptido p_i y un alelo de MHC asociado *h* como un vector de fila en el que las representaciones numéricas de variables que interactúan con alelos están concatenadas unas detrás de otras. Por ejemplo, el módulo de codificación 314 puede representar $\boldsymbol{x}_h^i$ como un vector de fila igual a

[p^i], **[p^i $b_h{}^i$]**, **[p^i $s_h{}^i$]**, o **[p^i $b_h{}^i$ $s_h{}^i$]**, donde $b_h{}^i$ es la predicción de afinidad de unión para el péptido p_i y alelo de MHC asociado h, y de forma similar para $s_h{}^i$ para la estabilidad. De manera alternativa, una o más combinación de variables que interactúan con alelos se pueden almacenar de forma individual (por ejemplo, como vectores o matrices individuales).

[00548] En un caso, el módulo de codificación 314 representa la información de afinidad de unión al incorporar valores medidos o previstos para la afinidad de unión en las variables que interactúan con alelos $\boldsymbol{x_h{}^i}$.

[00549] En un caso, el módulo de codificación 314 representa la información de estabilidad de unión al incorporar valores medidos o previstos para la estabilidad de unión en las variables que interactúan con alelos $\boldsymbol{x_h{}^i}$,

[00550] En un caso, el módulo de codificación 314 representa la información de tasa de asociación de unión al incorporar valores medidos o previstos para la tasa de asociación de unión en las variables que interactúan con alelos $\boldsymbol{x_h{}^i}$.

[00551] En un caso, para los péptidos presentados por moléculas del MHC clase I, el módulo de codificación 314 representa la longitud de péptidos como un vector $\boldsymbol{T_k}$=[$\mathbb{1}(L_k$=8) $\mathbb{1}(L_k$=9) $\mathbb{1}(L_k$=10) $\mathbb{1}(L_k$=11) $\mathbb{1}(L_k$=12) $\mathbb{1}(L_k$=13) $\mathbb{1}(L_k$=14) $\mathbb{1}(L_k$=15)] donde $\mathbb{1}$ es la función de indicador, y L_k indica la longitud del péptido $\boldsymbol{p_k}$. El vector $\boldsymbol{T_k}$ puede incluirse en las variables que interactúan con alelos $\boldsymbol{x_h{}^i}$. En otra instancia, para los péptidos presentados por moléculas del MHC clase II, el módulo codificante 314 representa la longitud del péptido como un vector $\boldsymbol{T_k}$=[$\mathbb{1}(L_k$=6) $\mathbb{1}(L_k$=7) $\mathbb{1}(L_k$=8) $\mathbb{1}(L_k$=9) $\mathbb{1}(L_k$=10) $\mathbb{1}(L_k$=11) $\mathbb{1}(L_k$=12) $\mathbb{1}(L_k$=13) $\mathbb{1}(L_k$=14) $\mathbb{1}(L_k$=15) $\mathbb{1}(L_k$=16) $\mathbb{1}(L_k$=17) $\mathbb{1}(L_k$=18) $\mathbb{1}(L_k$=19) $\mathbb{1}(L_k$=20) $\mathbb{1}(L_k$=21) $\mathbb{1}(L_k$=22) $\mathbb{1}(L_k$=23) $\mathbb{1}(L_k$=24) $\mathbb{1}(L_k$=25) $\mathbb{1}(L_k$=26) $\mathbb{1}(L_k$=27) $\mathbb{1}(L_k$=28) $\mathbb{1}(L_k$=29) $\mathbb{1}(L_k$=30)] donde $\mathbb{1}$ es la función indicadora y L_k indica la longitud del péptido $\boldsymbol{p_k}$. El vector $\boldsymbol{T_k}$ puede incluirse en las variables que interactúan con alelos $\boldsymbol{x_h{}^i}$.

[00552] En un caso, el módulo de codificación 314 representa la información de expresión del RNA de alelos de MHC al incorporar niveles de expresión basados en la secuenciación de RNA de alelos de MHC en las variables que interactúan con alelos $\boldsymbol{x_h{}^i}$.

[00553] De manera similar, el módulo de codificación 314 puede representar las variables que no interactúan con alelos $\boldsymbol{w^i}$ como un vector de fila en el que las representaciones numéricas de variables que no interactúan con alelos están concatenadas unas detrás de otras. Por ejemplo, $\boldsymbol{w^i}$ puede ser un vector de fila igual a **[$\boldsymbol{c^i}$]** o **[$\boldsymbol{c^i}$ $\boldsymbol{m^i}$ $\boldsymbol{w^i}$]** en el que $\boldsymbol{w^i}$ es un vector de fila que representa cualquier otra variable que no interactúa con alelos además de la secuencia flanqueante de extremo C del péptido $\boldsymbol{p^i}$ y la medición de cuantificación del RNAm $\boldsymbol{m^i}$ asociada con el péptido. De manera alternativa, una o más combinación de variables que no interactúan con alelos se pueden almacenar de forma individual (por ejemplo, como vectores o matrices individuales).

[00554] En un caso, el módulo de codificación 314 representa la velocidad de rotación de la

proteína fuente para una secuencia peptídica al incorporar la velocidad de rotación o semivida en las variables que no interactúan con alelos w^i.

[00555] En un caso, el módulo de codificación 314 representa la longitud de la proteína fuente o isoforma al incorporar la longitud en las variables que no interactúan con alelos w^i.

[00556] En un caso, el módulo de codificación 314 representa la activación de la inmunoproteasoma al incorporar la expresión promedio de las subunidades de proteasoma específicas para la inmunoproteasoma que incluyen las subunidades $\beta 1_i$, $\beta 2_i$, $\beta 5_i$ en las variables que no interactúan con alelos w^i.

[00557] En un caso, el módulo de codificación 314 representa la abundancia de la secuenciación de RNA de la proteína fuente del péptido o gen o transcripción de un péptido (unidades cuantificadas de FPKM, TPM mediante técnicas tales como RSEM) puede ser incorporando la abundancia de la proteína fuente en las variables que no interactúan con alelos w^i.

[00558] En un caso, el módulo de codificación 314 representa la probabilidad de que la transcripción de origen de un péptido se someterá a degradación mediada por mutaciones terminadoras (NMD) según lo estimado por el modelo en, por ejemplo, Rivas et. ál. *Science,* 2015 al incorporar esta probabilidad en las variables que no interactúan con alelos w^i.

[00559] En un caso, el módulo de codificación 314 representa el estado de activación de un módulo o vía génica evaluada mediante secuenciación de RNA, por ejemplo, al cuantificar la expresión de los genes en la vía en unidades de TPM usando, por ejemplo, RSEM para cada uno de los genes en la vía luego computando una estadística general, por ejemplo, el promedio, a lo largo de los genes en la vía. El promedio se puede incorporar en las variables que no interactúan con alelos w^i.

[00560] En un caso, el módulo de codificación 314 representa la cantidad de copias del gen fuente al incorporar la cantidad de copias en las variables que no interactúan con alelos w^i.

[00561] En un caso, el módulo de codificación 314 representa la afinidad de unión a TAP al incluir la afinidad de unión a TAP prevista o medida (por ejemplo, en unidades molares) en las variables que no interactúan con alelos w^i.

[00562] En un caso, el módulo de codificación 314 representa los niveles de expresión de TAP al incluir los niveles de expresión de TAP medidos mediante secuenciación de RNA (y cuantificados en unidades de TPM mediante, por ejemplo, RSEM) en las variables que no interactúan con alelos w^i.

[00563] En un caso, el módulo de codificación 314 representa mutaciones tumorales como un vector de variables indicadoras (es decir, $d^k = 1$ si el péptido p^k viene de una muestra con una mutación KRAS G12D y 0 de otro modo) en las variables que no interactúan con alelos w^i.

[00564] En un caso, el módulo de codificación 314 representa polimorfismos de línea germinal en genes de presentación de antígenos como un vector de variables indicadoras (es decir, $d^k = 1$ si el

péptido p^k viene de una muestra con un polimorfismo de línea germinal específico en el TAP). Estas variables indicadoras se pueden incluir en las variables que no interactúan con alelos w^i.

[00565] En un caso, el módulo de codificación 314 representa el tipo de tumor como un vector codificado *one-hot* de una longitud en el alfabeto de tipos de tumores (por ejemplo, NSCLC, melanoma, cáncer colorrectal, etc.). Estas variables codificadas *one-hot* se pueden incluir en las variables que no interactúan con alelos w^i.

[00566] En un caso, el módulo de codificación 314 representa los sufijos de alelos de MHC al tratar los alelos de HLA de 4 dígitos con diferentes sufijos. Por ejemplo, HLA-A*24:09N se considera un alelo diferente de HLA-A*24:09 para los fines del modelo. De manera alternativa, la probabilidad de presentación por medio de un alelo de MHC con sufijo N puede establecerse a cero para todos los péptidos porque los alelos de HLA que terminan en el sufijo N no se expresan.

[00567] En un caso, el módulo de codificación 314 representa el subtipo de tumor como un vector codificado *one-hot* de una longitud en el alfabeto de subtipos de tumores (por ejemplo, adenocarcinoma de pulmones, carcinoma de células escamosas de pulmón, etc.). Estas variables codificadas *one-hot* se pueden incluir en las variables que no interactúan con alelos w^i.

[00568] En un caso, el módulo de codificación 314 representa el historial de tabaquismo como una variable indicadora binaria (d^k = 1 si el paciente tiene un historial de tabaquismo y 0 de otro modo) que se puede incluir en las variables que no interactúan con alelos w^i. De manera alternativa, el historial de tabaquismo se puede codificar como una variable codificada *one-hot* de una longitud en un alfabeto de gravedad de tabaquismo. Por ejemplo, el estado de tabaquismo se puede clasificar en una escala de 1-5, donde 1 indica no fumadores y 5 indica los fumadores empedernidos actuales. Debido a que el historial de tabaquismo es principalmente relevante para tumores de pulmón, cuando se entrena un modelo en múltiples tipos de tumor, esta variable se puede definir también para que sea igual a 1 si el paciente tiene un historial de tabaquismo y el tipo de tumor es un tumor de pulmón y 0 si es de otro modo.

[00569] En un caso, el módulo de codificación 314 representa el historial de quemaduras de sol como una variable indicadora binaria (d^k = 1 si el paciente tiene un historial de quemaduras de sol graves y 0 si es de otro modo) que se puede incluir en las variables que no interactúan con alelos w^i. Debido a que la quemadura de sol grave es principalmente relevante para melanomas, cuando se entrena un modelo en múltiples tipos de tumor, esta variable se puede definir también para que sea igual a 1 si el paciente tiene un historial de quemaduras de sol graves y el tipo de tumor es melanoma y 0 si es de otro modo.

[00570] En un caso, el módulo de codificación 314 representa la distribución de niveles de expresión de un gen o transcripción particular para cada gen o transcripción en el genoma humano como

estadística general (por ejemplo, promedio, mediana) de distribución de niveles de expresión al usar bases de datos de referencia tal como la TCGA. Específicamente, para un péptido $\boldsymbol{p}^k$ en una muestra con muestra de tipo tumoral, se puede incluir no solamente el nivel de expresión de transcripción o génica medido del gen o transcripción de origen del péptido $\boldsymbol{p}^k$ en las variables que no interactúan con alelos $\boldsymbol{w}^i$, sino también la expresión de transcripción o génica media y/o mediana del gen o transcripción de origen del péptido $\boldsymbol{p}^k$ en melanomas según lo medido por TCGA.

[00571] En un caso, el módulo de codificación 314 representa el tipo de mutación como una variable codificada *one-hot* de una longitud en el alfabeto de tipos de mutaciones (por ejemplo, cambio de marco, inducción por NMD, etc.). Estas variables codificadas *one-hot* se pueden incluir en las variables que no interactúan con alelos $\boldsymbol{w}^i$.

[00572] En un caso, el módulo de codificación 314 representa las características de nivel proteico de la proteína como el valor de la anotación (por ejemplo, longitud de la 5' UTR) de la proteína fuente en las variables que no interactúan con alelos $\boldsymbol{w}^i$. En otro caso, el módulo de codificación 314 representa las anotaciones de nivel de residuos de la proteína fuente para el péptido $\boldsymbol{p}^i$ al incluir una variable indicadora que es igual a 1 si el péptido $\boldsymbol{p}^i$ se superpone con un motivo de hélice y 0 de otro modo, o que es igual a 1 si la proteína $\boldsymbol{p}^i$ se encuentra completamente incluida dentro de un motivo de hélice en las variables que no interactúan con alelos $\boldsymbol{w}^i$. En otro caso, una característica que representa la proporción de residuos en el péptido $\boldsymbol{p}^i$ que se encuentra incluida dentro de una anotación de motivo de hélice puede incluirse en las variables que no interactúan con alelos $\boldsymbol{w}^i$.

[00573] En un caso, el módulo de codificación 314 representa el tipo de proteínas o isoformas en el proteoma humano como un vector indicador o^k que tiene una longitud igual a la cantidad de proteínas o isoformas en el proteoma humano y el elemento correspondiente o^k_i es 1 si el péptido $\boldsymbol{p}^k$ proviene de una proteína i y 0 otro modo.

[00574] En un caso, el módulo de codificación 314 representa el gen de origen G=gen($\boldsymbol{p}^i$) del péptido $\boldsymbol{p}^i$ como variable categórica con L categorías posibles, donde L denota el límite superior de la cantidad de genes de origen indexados 1, 2, …, L.

[00575] El módulo de codificación 314 puede representar también un conjunto general de variables $\boldsymbol{z}^i$ para el péptido $\boldsymbol{p}^i$ y un alelo de MHC asociado h tiene un vector de fila en el que las representaciones numéricas de variables que interactúan con alelos $\boldsymbol{x}^i$ y las variables que no interactúan con alelos $\boldsymbol{w}^i$ están concatenadas unas detrás de otras. Por ejemplo, el módulo de codificación 314 puede representar $\boldsymbol{z}_h^i$ como un vector de fila igual a $\boldsymbol{[x_h^i\ w^i]}$ o $\boldsymbol{[w_i\ x_h^i]}$.

X. Módulo de entrenamiento

[00576] El módulo de entrenamiento 316 construye uno o más modelos de presentación que generan probabilidades de si secuencias peptídicas se presentarán por medio de alelos de MHC

asociados con las secuencias peptídicas. Específicamente, dada una secuencia peptídica $\boldsymbol{p^k}$ y un conjunto de alelos de MHC $\boldsymbol{a^k}$ asociados con la secuencia peptídica $\boldsymbol{p^k}$, cada modelo de presentación genera una estimación u_k que indica una probabilidad de que la secuencia peptídica $\boldsymbol{p^k}$ será presentada por uno o más alelos de MHC asociados $\boldsymbol{a^k}$.

X.A. Visión general

[00577] El módulo de entrenamiento 316 construye los uno o más modelos de presentación de acuerdo con los conjuntos de datos de entrenamiento almacenados en el almacenamiento 170 generados a partir de la información de presentación almacenada en 165. Generalmente, independientemente del tipo específico de modelo de presentación, todos los modelos de presentación capturan la dependencia entre las variables independientes y variables dependientes en los datos de entrenamiento 170 de modo que se minimiza una función de pérdida. Específicamente, la función de pérdida ℓ ($y_{i\in S}$, $u_{i\in S}$; $\boldsymbol{\theta}$) representa las discrepancias entre los valores de variables dependientes $y_{i\in S}$ para una o más instancias de datos S en los datos de entrenamiento 170 y las probabilidades estimadas $u_{i\in S}$ para las instancias de datos S generadas por el modelo de presentación. En una implementación particular a la que se hace referencia a lo largo del resto de la memoria descriptiva, la función de pérdida ($y_{i\in S}$, $u_{i\in S}$; $\boldsymbol{\theta}$) es la función de probabilidad log negativa dada por la ecuación (1a) de la siguiente manera:

$$\ell(y_{i\in S}, u_{i\in S}; \boldsymbol{\theta}) = \sum_{i\in S} (y_i \log u_i + (1 - y_i) \log(1 - u_i)) \,. \quad (1a)$$

Sin embargo, en la práctica, se puede utilizar otra función de pérdida. Por ejemplo, cuando se hacen las predicciones para la corriente de iones de espectrometría de masas, la función de pérdida es la pérdida cuadrada media dada por la ecuación 1b es de la siguiente manera:

$$\ell(y_{i\in S}, u_{i\in S}; \theta) = \sum_{i\in S} (\|y_i - u_i\|_2^2) \,. \quad (1b)$$

[00578] El modelo de presentación puede ser un modelo paramétrico en el que uno o más parámetros $\boldsymbol{\theta}$ especifican matemáticamente la dependencia entre las variables independientes y las variables dependientes. Típicamente, diversos parámetros de modelos de presentación de tipo paramétrico que minimizan la función de pérdida ($y_{i\in S}$, $u_{i\in S}$; $\boldsymbol{\theta}$) se determinan a través de los algoritmos de optimización numéricos basados en gradientes tales como algoritmos de gradientes, algoritmos de gradientes estocásticas y similares. De manera alternativa, el modelo de presentación puede ser un modelo no paramétrico en el que la estructura de modelo se determina a partir de los datos de entrenamiento 170 y no se basa estrictamente en un conjunto fijo de parámetros.

X.B. Modelos por alelo

[00579] El módulo de entrenamiento 316 puede construir los modelos de presentación para predecir las probabilidades de presentación de los péptidos por cada alelo. En este caso, el módulo de

entrenamiento 316 puede entrenar los modelos de presentación de acuerdo con las instancias de datos *S* en los datos de entrenamiento 170 generados a partir de células que expresan alelos de MHC únicos.

[00580] En una implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ para un alelo específico *h* en:

$$u_k^h = \Pr\left(\boldsymbol{p^k} \text{ presentados; alelo de}\text{MHC } h\right) = f\left(g_h\left(\boldsymbol{x_h^k}; \boldsymbol{\theta_h}\right)\right), (2)$$

donde la secuencia peptídica $\boldsymbol{x_h^k}$ indica las variables que interactúan con alelos codificadas para el péptido $\boldsymbol{p^k}$ y el alelo de MHC correspondiente *h*, *f*(·) es cualquier función, y en la presente se hace referencia a una función de transformación para la conveniencia de la descripción. Asimismo, $g_h(\cdot)$ es cualquier función, en la presente se hace referencia a una función de dependencia para la conveniencia de la descripción y genera puntuaciones de dependencia para las variables que interactúan con alelos $\boldsymbol{x_h^k}$ de acuerdo con un conjunto de parámetros $\boldsymbol{\theta_h}$ determinados por el alelo de MHC *h*. Los valores para el conjunto de parámetros $\boldsymbol{\theta_h}$ para cada alelo de MHC *h* se puede determinar al minimizar la función de pérdida con respecto a $\boldsymbol{\theta_h}$, donde *i* es cada instancia en el subconjunto *S* de datos de entrenamiento 170 generados a partir de células que expresan el único alelo de MHC *h*.

[00581] El resultado de la función de dependencia $g_h(\boldsymbol{x_h^k};\boldsymbol{\theta_h})$ representa una puntuación de dependencia para el alelo de MHC *h* que indica si el alelo de MHC *h* presentará el neoantígeno correspondiente de acuerdo con al menos las características que interactúan con alelos $\boldsymbol{x_h^k}$, y en particular, de acuerdo con las posiciones de aminoácidos de la secuencia peptídica del péptido $\boldsymbol{p^k}$. Por ejemplo, la puntuación de dependencia para el alelo de MHC *h* puede tener un valor alto si el alelo de MHC *h* es probable que presente el péptido $\boldsymbol{p^k}$ y puede tener un valor bajo si la presentación no es probable. La función de transformación *f*(·) transforma el resultado, y más específicamente, transforma la puntuación de dependencia generada por $g_h(\boldsymbol{x_h^k};\boldsymbol{\theta_h})$ en este caso, en un valor adecuado para indicar la probabilidad de que el péptido $\boldsymbol{p^k}$ será presentado por un alelo de MHC.

[00582] En una implementación particular a la que se hace referencia en todo lo que resta de la memoria descriptiva, *f*(·) es una función que tiene el intervalo dentro de [0, 1] durante un intervalo de dominio adecuado. En un ejemplo, *f*(·) es la función expit dada por:

$$f(z) = \frac{\exp(z)}{1 + \exp(z)}. (4)$$

Como ejemplo adicional, *f*(·) puede ser también la función tangente hiperbólica dada por:

$$f(z) = \tanh(z) \quad (5)$$

cuando los valores para el dominio z es igual o mayor que 0. De manera alternativa, cuando se hacen las predicciones para la corriente de iones de espectrometría de masas que tienen valores fuera del intervalo [0, 1], *f*(·) puede ser cualquier función tal como la función de identidad, la función exponencial, la función log y similares.

[00583] Por lo tanto, la probabilidad por alelo de que una secuencia peptídica $\boldsymbol{p^k}$ sea presentada por un alelo de MHC *h* puede generarse al aplicar la función de dependencia $g_h(\cdot)$ para el alelo de MHC h a la versión codificada de la secuencia peptídica $\boldsymbol{p^k}$ para generar la puntuación de dependencia correspondiente. La puntuación de dependencia se puede transformar mediante la función de transformación $f(\cdot)$ para generar una probabilidad por alelos de que la secuencia peptídica $\boldsymbol{p^k}$ se presentará por el alelo de MHC *h*.

X.B.1 Funciones de dependencia para variables que interactúan con alelos

[00584] En una implementación particular a la que se hace referencia en toda la memoria descriptiva, la función de dependencia $g_h(\cdot)$ es una función afín dada por:

$$g_h\left(\boldsymbol{x_h^i}; \boldsymbol{\theta_h}\right) = \boldsymbol{x_h^i} \cdot \boldsymbol{\theta_h}. (6)$$

que combina de forma lineal cada variable que interactúa con alelos en $\boldsymbol{x_h^k}$ con un parámetro correspondiente en el conjunto de parámetros $\boldsymbol{\theta_h}$ determinados por el alelo de MHC asociado *h*.

[00585] En otra implementación particular a la que se hace referencia en toda la memoria descriptiva, la función de dependencia $g_h(\cdot)$ es una función de red dada por:

$$g_h\left(\boldsymbol{x_h^i}; \boldsymbol{\theta_h}\right) = NN_h(\boldsymbol{x_h^i}; \boldsymbol{\theta_h}). (7)$$

representado por un modelo de red $NN_h(\cdot)$ que tiene una serie de nodos colocados en una o más capas. Un nodo puede estar conectado a otros nodos a través de conexiones, donde cada uno tiene un parámetro asociado en el conjunto de parámetros $\boldsymbol{\theta_h}$. Un valor en un nodo particular puede estar representado como una suma de los valores de nodos conectados al nodo particular ponderado por el parámetro asociado mapeado por una función de activación asociada con el nodo particular. Al contrario de la función afín, los modelos de red son ventajosos porque el modelo de presentación puede incorporar no linealidad y datos de proceso que tienen diferentes longitudes de secuencias de aminoácidos. Específicamente, a través del modelado no lineal, los modeles de red pueden capturar la interacción entre los aminoácidos en diferentes posiciones en una secuencia peptídica y cómo esta interacción afecta la presentación de péptidos.

[00586] En general, los modelos de red $NN_h(\cdot)$ se pueden estructurar como redes de previsión tales como redes neurales artificiales (ANN), redes neurales convolucionales (CNN), redes neurales profundas (DNN) y/o redes recurrentes, tal como redes de memoria de corto plazo (LSTM), redes recurrentes bidireccionales, redes recurrentes bidireccionales profundas y similares.

[00587] En un caso al que se hacer referencia en todo lo que resta de la memoria descriptiva, cada alelo de MHC en *h=1,2,…, m* se asocia con un modelo de red separada y $NN_h(\cdot)$ indica los resultado/s de un modelo de red asociado con el alelo de MHC *h*.

[00588] La figura 5 ilustra un modelo de red de ejemplo $NN_3(\cdot)$ en asociación con un alelo de MHC arbitrario h=3. Como se ilustra en la figura 5, el modelo de red $NN_3(\cdot)$ para el alelo de MHC $h=3$ incluye tres nodos de entrada en la capa $l=1$, cuatro nodos en la capa $l=2$, dos nodos en la capa $l=3$ y un nodo de salida en la capa $l=4$. El modelo de red $NN_3(\cdot)$ se asocia con un conjunto de diez parámetros $\theta_3(1)$, $\theta_3(2)$, …, $\theta_3(10)$. El modelo de red $NN_3(\cdot)$ recibe los valores de entrada (instancias de datos individuales que incluyen datos secuencias de polipéptidos codificados y cualquier otro dato de entrenamiento utilizado) para tres variables que interactúan con alelos $x_3^k(1)$, $x_3^k(2)$ y $x_3^k(3)$ para el alelo de MHC $h=3$ y produce el valor $NN_3(x_3^k)$. La función de red también puede incluir uno o más modelos de red, donde cada uno toma distintas variables de interacción de alelos como entrada.

[00589] En otro caso, los alelos de MHC identificados $h=1, 2, \ldots, m$ se asocian con un modelo de red único $NN_H(\cdot)$, y $NN_h(\cdot)$ indica uno o más resultados del modelo de red único asociado con el alelo de MHC h. En dicho caso, el conjunto de parámetros $\boldsymbol{\theta_h}$ puede corresponder a un conjunto de parámetros para el modelo de red único, y por lo tanto, el conjunto de parámetros $\boldsymbol{\theta_h}$ se puede compartir por todos los alelos de MHC.

[00590] La figura 6A ilustra un modelo de red de ejemplo $NN_H(\cdot)$ compartido por los alelos de MHC $h=1,2, \ldots,m$. Como se ilustra en la figura 6A, el modelo de red $NN_H(\cdot)$ incluye m nodos de salida donde cada uno corresponde a un alelo de MHC. El modelo de red $NN_3(\cdot)$ recibe las variables que interactúan con alelos $\boldsymbol{x_3^k}$ para el alelo de MHC $h=3$ y produce valores m que incluyen el valor $NN_3(x_3^k)$ que corresponde al alelo de MHC $h=3$.

[00591] En incluso otro caso, el modelo de red único $NN_H(\cdot)$ puede ser un modelo de red que produce una puntuación de dependencia dadas las variables que interactúan con alelos $\boldsymbol{x_h^k}$ y la secuencia de proteínas codificada $\boldsymbol{d_h}$ de un alelo de MHC h. En dicho caso, el conjunto de parámetros $\boldsymbol{\theta_h}$ puede corresponder de nuevo a un conjunto de parámetros para el modelo de red único, y por lo tanto, el conjunto de parámetros $\boldsymbol{\theta_h}$ puede ser compartido por todos los alelos de MHC. Por lo tanto, en dicho caso, $NN_h(\cdot)$ puede indicar el resultado del modelo de red único $NN_H(\cdot)$ dadas las entradas $[\boldsymbol{x_h^k}\ \boldsymbol{d_h}]$ al modelo de red único. Dicho modelo de red es ventajoso porque las probabilidades de presentación de péptidos para los alelos de MHC que no eran conocidas en los datos de entrenamiento se pueden predecir justo mediante identificación de su secuencia de proteínas.

[00592] La figura 6 B ilustra un modelo de red de ejemplo $NN_H(\cdot)$ compartido por los alelos de MHC. Como se ilustra en la figura 6B, el modelo de red $NN_H(\cdot)$ recibe las variables que interactúan con alelos y la secuencia de proteínas del alelo de MHC $h=3$ como entrada, y produce una puntuación de dependencia $NN_3(x_3^k)$ que corresponde al alelo de MHC $h=3$.

[00593] En incluso otro caso, la función de dependencia $\boldsymbol{g_h(\cdot)}$ se puede expresar como:

$$g_h\left(\boldsymbol{x_h^k}; \boldsymbol{\theta_h}\right) = g'_h\left(\boldsymbol{x_h^k}; \boldsymbol{\theta'_h}\right) + \theta_h^0$$

donde $g'_h(\boldsymbol{x_h^k};\boldsymbol{\theta'_h})$ es la función afín con un conjunto de parámetros $\boldsymbol{\theta'_h}$, la función de red o similares, con un parámetro de inclinación $\boldsymbol{\theta_h^0}$ en el conjunto de parámetros para variables que interactúan con alelos para el alelo de MHC que representa una probabilidad de referencia de la presentación para el alelo de MHC *h.*

[00594] En otra implementación, el parámetro de inclinación $\boldsymbol{\theta_h^0}$ puede compartirse de acuerdo con la familia génica del alelo de MHC *h.* Es decir, el parámetro de inclinación $\boldsymbol{\theta_h^0}$ para el alelo de MHC *h* puede ser igual a $\boldsymbol{\theta_{gen(h)}^0}$, donde el gen(*h*) es la familia génica del alelo de MHC *h.* Por ejemplo, los alelos de MHC clase I HLA-A*02:01, HLA-A*02:02 y HLA-A*02:03 se pueden asignar a la familia génica de «HLA-A» y se puede compartir el parámetro de inclinación $\boldsymbol{\theta_h^0}$ para cada uno de estos alelos de MHC. A modo de otro ejemplo, los alelos del MHC clase II HLA-DRB1:10:01, HLA-DRB1:11:01 y HLA-DRB3:01:01 pueden asignarse a la familia de genes de “HLA-DRB,” y el parámetro de sesgo $\boldsymbol{\theta_h^0}$ para cada uno de estos alelos de MHC puede ser compartido.

[00595] Volviendo a la ecuación (2), como ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ sea presentado por el alelo de MHC *h=3*, entre *m=4* diferentes alelos de MHC identificados usando la función de dependencia afín $g_h(\cdot)$, puede generarse por:

$$u_k^3 = f\left(\boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right),$$

donde $\boldsymbol{x_3^k}$ son las variables que interactúan con alelos identificadas para el alelo de MHC *h=3*, y $\boldsymbol{\theta_3}$ son el conjunto de parámetros determinados por el alelo de MHC *h=3* a través de la minimización de función de pérdida.

[00596] Como ejemplo adicional, la probabilidad de que el péptido $\boldsymbol{p^k}$ sea presentado por el alelo de MHC *h=3*, entre *m=4* diferentes alelos de MHC identificados usando funciones de transformación de red separadas $g_h(\cdot)$, puede generarse por:

$$u_k^3 = f\left(NN_3(\boldsymbol{x_3^k};\boldsymbol{\theta_3})\right),$$

donde $\boldsymbol{x_3^k}$ son las variables que interactúan con alelos identificadas para el alelo de MHC *h=3*, y $\boldsymbol{\theta_3}$ son el conjunto de parámetros determinados por el modelo de red $NN_3(\cdot)$ asociado con el alelo de MHC *h=3*.

[00597] La figura 7 ilustra la generación de una probabilidad de presentación para el péptido $\boldsymbol{p^k}$ en asociación con el alelo de MHC *h*=3 usando un modelo de red de ejemplo $NN_3(\cdot)$. Como se ilustra en la figura 7, el modelo de red $NN_3(\cdot)$ recibe las variables que interactúan con alelos $\boldsymbol{x_3^k}$ para el alelo de MHC *h=3* y genera el resultado $NN_3(\boldsymbol{x_3^k})$. El resultado es mapeado por la función $\boldsymbol{f(\cdot)}$ para generar la probabilidad de presentación estimada u_k.

X.B.2. Por alelo con variables que no interactúan con alelos

[00598] En una implementación, el módulo de entrenamiento 316 incorpora variables que no

interactúan con alelos y modela la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ mediante:

$$u_k^h = \Pr\left(\boldsymbol{p^k} \text{ presentado}\right) = f\left(g_w\left(\boldsymbol{w^k}; \boldsymbol{\theta_w}\right) + g_h\left(\boldsymbol{x_h^i}; \boldsymbol{\theta_h}\right)\right), (8)$$

donde $\boldsymbol{w^k}$ indica las variables que no interactúan con alelos codificadas para el péptido $\boldsymbol{p^k}$, $g_w(\cdot)$ es una función para las variables que no interactúan con alelos $\boldsymbol{w^k}$ de acuerdo con un conjunto de parámetros $\boldsymbol{\theta_w}$ determinados para las variables que no interactúan con alelos. Específicamente, los valores para el conjunto de parámetros $\boldsymbol{\theta_h}$ para cada alelo de MHC *h* y el conjunto de parámetros $\boldsymbol{\theta_w}$ para las variables que no interactúan con alelos se pueden determinar al minimizar la función de pérdida con respecto a $\boldsymbol{\theta_h}$ y $\boldsymbol{\theta_w}$, donde *i* es cada instancia en el subconjunto *S* de datos de entrenamiento 170 generados a partir de células que expresan alelos de MHC únicos.

[00599] El resultado de la función de dependencia $g_w(\boldsymbol{w^k};\boldsymbol{\theta_w})$ representa una puntuación de dependencia para las variables que no interactúan con alelos que indica si el péptido $\boldsymbol{p^k}$ será presentado por uno o más alelos de MHC de acuerdo con el impacto de las variables que no interactúan con alelos. Por ejemplo, la puntuación de dependencia para las variables que no interactúan con alelos puede tener un valor alto si el péptido $\boldsymbol{p^k}$ se asocia con una secuencia flanqueante de extremo C que es conocida por impactar positivamente la presentación del péptido $\boldsymbol{p^k}$ y puede tener un valor bajo si el péptido $\boldsymbol{p^k}$ se asocia con una secuencia flanqueante de extremo C que es conocida por impactar negativamente la presentación del péptido $\boldsymbol{p^k}$.

[00600] De acuerdo con la ecuación (8), la probabilidad por alelo de que una secuencia peptídica p^k sea presentada por un alelo de MHC *h* puede generarse al aplicar la función $g_h(\cdot)$ para el alelo de MHC *h* a la versión codificada de la secuencia peptídica $\boldsymbol{p^k}$ para generar la puntuación de dependencia correspondiente para variables que interactúan con alelos. La función $g_w(\cdot)$ para las variables que no interactúan con alelos se aplica también a la versión codificada de las variables que no interactúan con alelos para generar la puntuación de dependencia para las variables que no interactúan con alelos. Ambas puntuaciones se combinan y la puntuación combinada se transforma mediante la función de transformación $f(\cdot)$ para generar una probabilidad por alelos de que la secuencia peptídica $\boldsymbol{p^k}$ será presentada por el alelo de MHC *h*.

[00601] De manera alternativa, el módulo de entrenamiento 316 puede incluir variables que no interactúan con alelos $\boldsymbol{w^k}$ en la predicción al agregar las variables que no interactúan con alelos $\boldsymbol{w^k}$ a las variables que interactúan con alelos $\boldsymbol{x_h^k}$ en la ecuación (2). Por lo tanto, la probabilidad de presentación puede estar dada por:

$$u_k^h = \Pr\left(\boldsymbol{p^k} \text{ presentados; alelo } h\right) = f\left(g_h\left([\boldsymbol{x_h^k}\, \boldsymbol{w^k}]; \boldsymbol{\theta_h}\right)\right). (9)$$

X.B.3 Funciones de dependencia para variables que no interactúan con alelos

[00602] De manera similar a la función de dependencia $g_h(\cdot)$ para las variables que interactúan

con alelos, la función de dependencia $g_w(\cdot)$ para las variables que no interactúan con alelos puede ser una función afín o una función de red en la que un modelo de red separado se asocia con las variables que no interactúan con alelos $\boldsymbol{w^k}$.

[00603] Específicamente, la función de dependencia $g_w(\cdot)$ es una función dada por:

$$g_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}) = \boldsymbol{w^k} \cdot \boldsymbol{\theta_w}.$$

que combinan de forma lineal las variables que no interactúan con alelos en $\boldsymbol{w^k}$ con un parámetro correspondiente en el conjunto de parámetros $\boldsymbol{\theta_w}$.

[00604] La función de dependencia $\boldsymbol{g_w(\cdot)}$ puede ser también una función de red dada por:

$$g_h(\boldsymbol{w^k}; \boldsymbol{\theta_w}) = NN_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}).$$

representada por un modelo de red $NN_w(\cdot)$ que tiene un parámetro asociado en el conjunto de parámetros $\boldsymbol{\theta_w}$. La función de red también puede incluir uno o más modelos de red, donde cada uno toma distintas variables de no interacción de alelos como entrada.

[00605] En otro caso, la función de dependencia $\boldsymbol{g_w(\cdot)}$ para las variables que no interactúan con alelos puede estar dada por:

$$g_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}) = g'_w(\boldsymbol{w^k}; \boldsymbol{\theta'_w}) + h(m^k; \theta_w^m), \quad (10)$$

Donde $g'_w(\boldsymbol{w^k}; \boldsymbol{\theta'_w})$ es la función afín, la función de red con el conjunto de parámetros que no interactúan con alelos $\boldsymbol{\theta'_w}$, o similares, m^k es la medición de cuantificación de RNAm para el péptido $\boldsymbol{p^k}$, $h(\cdot)$ es una función que transforma la medición de cuantificación, y θ_w^m es un parámetro en el conjunto de parámetros para las variables que no interactúan con alelos que se combina con la medición de cuantificacion de RNAm para generar una puntuación de dependencia para la medición de cuantificación de RNAm. En una modalidad particular a la que se hace referencia en todo el resto de la memoria descriptiva, $h(\cdot)$ es la función log, sin embargo, en la práctica $h(\cdot)$ puede ser cualquiera de una variedad de funciones diferentes.

[00606] En incluso otro caso, la función de dependencia $\boldsymbol{g_w(\cdot)}$ para las variables que no interactúan con alelos puede estar dada por:

$$g_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}) = g'_w(\boldsymbol{w^k}; \boldsymbol{\theta'_w}) + \boldsymbol{\theta_w^o} \cdot \boldsymbol{o^k}, \quad (11)$$

donde $g'_w(\boldsymbol{w^k}; \boldsymbol{\theta'_w})$ es la función afín, la función de red con el conjunto de parámetros que no interactúan con alelos $\boldsymbol{\theta'_w}$, o similares, $\boldsymbol{o^k}$ es el vector indicador descrito anteriormente que representa proteínas e isoformas en el proteoma humano para el péptido $\boldsymbol{p^k}$, y $\boldsymbol{\theta_w^o}$ es un conjunto de parámetros en el conjunto de parámetros para variables que no interactúan con alelos que se combina con el vector indicador. En una variación, cuando la dimensionalidad de $\boldsymbol{o^k}$ y el conjunto de parámetros $\boldsymbol{\theta_w^o}$ son considerablemente altos, un término de regularización de parámetros, tal como $\lambda \cdot ||\boldsymbol{\theta_w^o}||$, donde $||\cdot||$ representa la norma L1, norma L2, una combinación o similares, se puede agregar a la función de pérdida al determinar el valor

de los parámetros. El valor óptimo del hiperparámetro λ puede determinarse a través de métodos adecuados.

[00607] En aun otro caso, la función de dependencia $g_w(\cdot)$ de las variables de no interacción de alelos puede otorgarse mediante:

$$g_w(\boldsymbol{w^k};\boldsymbol{\theta_w}) = g'_w(\boldsymbol{w^k};\boldsymbol{\theta'_w}) + \sum_{l=1}^{L} \mathbb{1}\left(\text{gene}(\boldsymbol{p^k} = l)\right) \cdot \boldsymbol{\theta_w^l}, \qquad (12)$$

donde $g'_w(\boldsymbol{w^k};\boldsymbol{\theta'_w})$ es la función afín, la función de red con el conjunto de parámetros de no interacción de alelos $\boldsymbol{\theta'_w}$, o similar, $\mathbb{1}$ (gen($\boldsymbol{p^k}$=*l*)) es la función indicadora que equivale a 1 si el péptido $\boldsymbol{p^k}$ es del gen de origen *l* como se describió anteriormente con respecto a las variables de no interacción de alelos, y $\boldsymbol{\theta_w^l}$ es un parámetro que indica la "antigenicidad" del gen de origen *l.* En una variación, cuando *L* es significativamente elevado y, por lo tanto, la cantidad de parámetros $\boldsymbol{\theta_w^{l=1, 2, \ldots, L}}$ es significativamente elevada, puede agregarse un término de regularización de parámetros, tal como $\lambda \cdot ||\boldsymbol{\theta_w^l}||$, donde ||·|| representa la norma L1, norma L2, una combinación o similar, a la función de pérdida al determinar el valor de los parámetros. El valor óptimo del hiperparámetro λ puede determinarse a través de los métodos adecuados.

[00608] En la práctica, los términos adicionales de cualquiera de las ecuaciones (10), (11) y (12) pueden combinarse para generar la función de dependencia $g_w(\cdot)$ para las variables de no interacción de alelos. Por ejemplo, el término $h(\cdot)$ que indica la medición de cuantificación de ARNm en la ecuación (10) y el término que indica la antigenicidad del gen de origen en la ecuación (12) pueden sumarse junto con cualquier otra función de red o afín para generar la función de dependencia para variables de no interacción de alelos.

[00609] Volviendo a la ecuación (8), como ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ sea presentado por el alelo de MHC *h=3*, entre *m=4* diferentes alelos de MHC identificados usando las funciones de transformación afines $g_h(\cdot)$, $g_w(\cdot)$, puede generarse por:

$$u_k^3 = f\left(\boldsymbol{w^k} \cdot \boldsymbol{\theta_w} + \boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right),$$

donde $\boldsymbol{w^k}$ son las variables que no interactúan con alelos identificadas para el péptido $\boldsymbol{p^k}$, y $\boldsymbol{\theta_w}$ son el conjunto de parámetros determinados para las variables que no interactúan con alelos.

[00610] Como ejemplo adicional, la probabilidad de que el péptido $\boldsymbol{p^k}$ sea presentado por el alelo de MHC *h=3*, entre *m=4* diferentes alelos de MHC identificados usando las funciones de transformación de red $g_h(\cdot)$, $g_w(\cdot)$, puede generarse por:

$$u_k^3 = f\left(NN_w(\boldsymbol{w^k};\boldsymbol{\theta_w}) + NN_3(\boldsymbol{x_3^k};\boldsymbol{\theta_3})\right)$$

donde $\boldsymbol{w^k}$ son las variables que interactúan con alelos identificadas para el péptido $\boldsymbol{p^k}$, y $\boldsymbol{\theta_w}$ son el conjunto de parámetros determinados para las variables que no interactúan con alelos.

[00611] La figura 8 ilustra la generación de una probabilidad de presentación para el péptido $\boldsymbol{p^k}$ en asociación con el alelo de MHC h=3 usando modelos de red de ejemplo $NN_3(\cdot)$ y $NN_w(\cdot)$. Como se ilustra en la figura 8, el modelo de red $NN_3(\cdot)$ recibe las variables que interactúan con alelos $\boldsymbol{x_3^k}$ para el alelo de MHC h=3 y genera el resultado $NN_3(\boldsymbol{x_3^k})$. El modelo de red $NN_w(\cdot)$ recibe las variables que no interactúan con alelos $\boldsymbol{w^k}$ para el péptido $\boldsymbol{p^k}$ y genera el resultado $NN_w(\boldsymbol{w^k})$. Los resultados se combinan y mapean por la función $f(\cdot)$ para generar la probabilidad de presentación estimada u_k.

X.C. Modelos de alelos múltiples

[00612] El módulo de entrenamiento 316 puede construir también los modelos de presentación para predecir las probabilidades de presentación de péptidos en una configuración de alelos múltiples donde dos o más alelos de MHC están presentes. En este caso, el módulo de entrenamiento 316 puede entrenar los modelos de presentación de acuerdo con las instancias de datos S en los datos de entrenamiento 170 generados a partir de células que expresan alelos de MHC únicos, células que expresan alelos de MHC múltiples o una combinación de estos.

X.C.1. Ejemplo 1: Máximo por modelos por alelos

[00613] En una implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ en asociación con un conjunto de alelos de MHC múltiples H en función de las probabilidades de presentación $u_k^{h\in H}$ determinadas para cada uno de los alelos de MHC h en el conjunto H determinado de acuerdo con las células que expresan alelos únicos, como se describió anteriormente en conjunto con ecuaciones (2)-(11). Específicamente, la probabilidad de presentación u_k puede ser cualquier función de $u_k^{h\in H.}$ En una implementación, como se muestra en la ecuación (12), la función es la función máxima y la probabilidad de presentación u_k puede determinarse como el máximo de las probabilidades de presentación para cada alelo de MHC h en el conjunto H.

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados; alelos } H\right) = \max\left(u_k^{h\in H}\right).$$

X.C.2. Ejemplo 2.1: Modelos de función de sumas

[00614] En una implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ mediante:

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados}\right) = f\left(\sum_{h=1}^{m} a_h^k \cdot g_h\left(\boldsymbol{x_h^k}; \boldsymbol{\theta_h}\right)\right), \quad (13)$$

donde los elementos a_h^k son 1 para los múltiples alelos de MHC H asociados con la secuencia peptídica $\boldsymbol{p^k}$ y $\boldsymbol{x_h^k}$ indica las variables que interactúan con alelos codificadas para el péptido $\boldsymbol{p^k}$ y los alelos de MHC correspondientes. Los valores para el conjunto de parámetros $\boldsymbol{\theta_h}$ para cada alelo de MHC h se pueden determinar al minimizar la función de pérdida con respecto a $\boldsymbol{\theta_h}$, donde i es cada instancia en el

subconjunto *S* de datos de entrenamiento 170 generados a partir de células que expresan alelos de MHC únicos y/o células que expresan múltiples alelos de MHC. La función de dependencia $\boldsymbol{g_h}$ puede ser en forma de cualquiera de las funciones de dependencia $\boldsymbol{g_h}$ introducidas anteriormente en las secciones X.B.1.

[00615] De acuerdo con la ecuación (13), la probabilidad de presentación de que una secuencia peptídica $\boldsymbol{p^k}$ sea presentada por uno o más alelos de MHC *h* puede generarse al aplicar la función de dependencia $g_h(\cdot)$ a la versión codificada de la secuencia peptídica $\boldsymbol{p^k}$ para cada uno de los alelos de MHC *H* para generar la puntuación correspondiente para las variables que interactúan con alelos. Las puntuaciones para cada alelo de MHC *h* se combinan y transforman por medio de la función de transformación $f(\cdot)$ para generar la probabilidad de presentación de que la secuencia peptídica $\boldsymbol{p^k}$ será presentada por el conjunto de alelos de MHC *H.*

[00616] El modelo de presentación de la ecuación (13) es diferente al modelo por alelo de la ecuación (2), ya que la cantidad de alelos asociados para cada péptido $\boldsymbol{p^k}$ puede ser mayor que 1. En otras palabras, más de un elemento en a_h^k puede tener valores de 1 para los múltiples alelos de MHC asociados con la secuencia peptídica $\boldsymbol{p^k}$.

[00617] Como ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ sea presentado por el alelo de MHC *h=2, h=3*, entre *m=4* diferentes alelos de MHC identificados usando las funciones de transformación afines $g_h(\cdot)$, puede generarse por:

$$u_k = f\left(\boldsymbol{x_2^k} \cdot \boldsymbol{\theta_2} + \boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right),$$

donde $\boldsymbol{x_2^k}, \boldsymbol{x_3^k}$ son las variables que interactúan con alelos identificadas para el alelo de MHC *h=2, h=3*, y $\boldsymbol{\theta_2}, \boldsymbol{\theta_3}$ son el conjunto de parámetros determinados por el alelo de MHC *h=2, h=3*.

[00618] Como ejemplo adicional, la probabilidad de que el péptido $\boldsymbol{p^k}$ sea presentado por los alelos de MHC *h=2, h=3*, entre *m=4* diferentes alelos de MHC identificados usando las funciones de transformación de red $g_h(\cdot)$, $g_w(\cdot)$, puede generarse por:

$$\boldsymbol{u_k} = f\left(NN_2(\boldsymbol{x_2^k}; \boldsymbol{\theta_2}) + NN_3(\boldsymbol{x_3^k}; \boldsymbol{\theta_3})\right),$$

donde $NN_2(\cdot)$, $NN_3(\cdot)$ son los modelos de red identificados para alelos de MHC *h=2, h=3*, y $\boldsymbol{\theta_2}, \boldsymbol{\theta_3}$ son el conjunto de parámetros determinado por los alelos de MHC *h=2, h=3*.

[00619] La Figura 9 ilustra la generación de una probabilidad de presentación para el péptido $\boldsymbol{p^k}$ asociado a los alelos de MHC *h*=2, *h*=3 con el uso de modelos de red ilustrativos $NN_2(\cdot)$ y $NN_3(\cdot)$. Como se muestra en la Figura 9, el modelo de red $NN_2(\cdot)$ recibe las variables de interacción con el alelo $\boldsymbol{x_2^k}$ para el alelo de MHC *h=2* y genera el producto $NN_2(\boldsymbol{x_2^k})$ y el modelo de red $NN_3(\cdot)$ recibe las variables de interacción con el alelo $\boldsymbol{x_3^k}$ para el alelo de MHC *h=3* y genera el producto $NN_3(\boldsymbol{x_3^k})$. Los productos se combinan y se mapean con la función $f(\cdot)$ para generar la probabilidad de presentación estimada u_k.

X.C.3. Ejemplo 2.2: Modelos de función de sumas con variables de no interacción con alelos

[00620] En una implementación, el módulo de entrenamiento 316 incorpora variables de no interacción con alelos y modela la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ mediante:

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados}\right) = f\left(g_w\left(\boldsymbol{w^k}; \boldsymbol{\theta_w}\right) + \sum_{h=1}^{m} a_h^k \cdot g_h\left(\boldsymbol{x_h^k};\ \boldsymbol{\theta_h}\right)\right), (14)$$

donde $\boldsymbol{w^k}$ denota las variables codificadas de no interacción con alelos para el péptido $\boldsymbol{p^k}$. Específicamente, los valores para el conjunto de parámetros $\boldsymbol{\theta_h}$ para cada alelo de MHC *h* y el conjunto de parámetros $\boldsymbol{\theta_w}$ para las variables de no interacción con alelos se pueden determinar al minimizar la función de pérdida con respecto a $\boldsymbol{\theta_h}$ y $\boldsymbol{\theta_w}$, donde *i* es cada ejemplo en el subconjunto *S* de los datos de entrenamiento 170 generados a partir de las células que expresan los alelos individuales de MHC y/o células que expresan múltiples alelos de MHC. La función de dependencia $\boldsymbol{g_w}$ puede encontrarse en forma de cualquiera de las funciones de dependencia $\boldsymbol{g_w}$ introducidas anteriormente en las secciones X.B.3.

[00621] Por tanto, de acuerdo con la ecuación (14), la probabilidad de presentación de que una secuencia peptídica $\boldsymbol{p^k}$ se presente mediante uno o más de los alelos de MHC *H* se puede generar mediante la aplicación de la función $g_h(\cdot)$ a la versión codificada de la secuencia peptídica $\boldsymbol{p^k}$ para cada uno de los alelos de MHC *H* para generar la puntuación de dependencia correspondiente para las variables de interacción con alelos para cada uno de los alelos de MHC *h*. La función $g_w(\cdot)$ para las variables de no interacción con alelos también se aplica a la versión codificada de las variables de no interacción con alelos para generar la puntuación de dependencia para las variables de no interacción con alelos. Las puntuaciones se combinan y la puntuación combinada se transforma mediante la función de transformación $f(\cdot)$ para generar la probabilidad de presentación que la secuencia peptídica $\boldsymbol{p^k}$ va a presentar mediante los alelos de MHC *H.*

[00622] En el modelo de presentación de la ecuación (14), la cantidad de alelos asociados para cada péptido $\boldsymbol{p^k}$ puede ser mayor que 1. En otras palabras, más de un elemento en a_h^k puede tener valores de 1 para los múltiples alelos de MHC *H* asociados a la secuencia peptídica $\boldsymbol{p^k}$.

[00623] A modo de ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *h=2, h=3*, entre *m=4* alelos de MHC diferentes identificados con el uso de las funciones de transformación afín $g_h(\cdot)$, $g_w(\cdot)$, se puede generar mediante:

$$u_k = f\left(\boldsymbol{w^k} \cdot \boldsymbol{\theta_w} + \boldsymbol{x_2^k} \cdot \boldsymbol{\theta_2} + \boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right),$$

donde $\boldsymbol{w^k}$ son las variables identificadas de no interacción con alelos para el péptido $\boldsymbol{p^k}$ y $\boldsymbol{\theta_w}$ son el conjunto de parámetros determinado para las variables de no interacción con alelos.

[00624] A modo de otro ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos

de MHC *h=2, h=3*, entre *m=4* alelos de MHC identificados diferentes con el uso de las funciones de transformación en red $g_h(\cdot)$, $g_w(\cdot)$, se puede generar mediante:

$$u_k = f\left(NN_w\left(\boldsymbol{w^k};\boldsymbol{\theta_w}\right) + NN_2(\boldsymbol{x_2^k};\boldsymbol{\theta_2}) + NN_3(\boldsymbol{x_3^k};\boldsymbol{\theta_3})\right)$$

donde $\boldsymbol{w^k}$ son las variables de interacción con alelos para el péptido $\boldsymbol{p^k}$ y $\boldsymbol{\theta_w}$ son el conjunto de parámetros determinado para las variables de interacción con alelos.

[00625] La Figura 10 ilustra la generación de una probabilidad de presentación para el péptido $\boldsymbol{p^k}$ asociado a los alelos de MHC *h=2, h*=3 con el uso de modelos de red ilustrativos $NN_2(\cdot)$, $NN_3(\cdot)$ y $NN_w(\cdot)$. Como se muestra en la Figura 10, el modelo de red $NN_2(\cdot)$ recibe las variables de interacción con alelos $\boldsymbol{x_2^k}$ para el alelo de MHC *h=2* y genera el producto $NN_2(\boldsymbol{x_2^k})$. El modelo de red $NN_3(\cdot)$ recibe las variables de interacción con alelos $\boldsymbol{x_3^k}$ para el alelo de MHC *h=3* y genera el producto $NN_3(\boldsymbol{x_3^k})$. El modelo de red $NN_w(\cdot)$ recibe las variables de no interacción con alelos $\boldsymbol{w^k}$ para el péptido $\boldsymbol{p^k}$ y genera el producto $NN_w(\boldsymbol{w^k})$. Los productos se combinan y se mapean con la función $f(\cdot)$ para generar la probabilidad de presentación estimada u_k.

[00626] De manera alternativa, el módulo de entrenamiento 316 puede incluir variables de no interacción con alelos $\boldsymbol{w^k}$ en la predicción al agregar las variables de no interacción con alelos $\boldsymbol{w^k}$ a las variables de interacción con alelos $\boldsymbol{x_h^k}$ en la ecuación (15). Por tanto, la probabilidad de presentación se puede proporcionar con:

$$u_k = \Pr\left(\boldsymbol{p^k}\ \text{presentados}\right) = f\left(\sum_{h=1}^{m} a_h^k \cdot g_h\left([\boldsymbol{x_h^k}\,\boldsymbol{w^k}];\boldsymbol{\theta_h}\right)\right). \quad (15)$$

X.C.4. Ejemplo 3.1: Modelos con probabilidades implícitas por alelo

[00627] En otra implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ con:

$$u_k = \Pr\left(\boldsymbol{p^k}\ \text{presentados}\right) = r\left(s\left(\boldsymbol{v} = \left[a_1^k \cdot u'^1_k(\boldsymbol{\theta}) \ldots\ a_m^k \cdot u'^m_k(\boldsymbol{\theta})\right]\right)\right), \quad (16)$$

donde los elementos a_h^k son 1 para los múltiples alelos de MHC $h \in H$ asociados a la secuencia peptídica $\boldsymbol{p^k}$, u'^h_k es una probabilidad de presentación implícita por alelo para el alelo de MHC h, el vector $\boldsymbol{v}$ es un vector en el cual el elemento v_h corresponde a $a_h^k \cdot u'^h_k$, $s(\cdot)$ es una función que mapea los elementos de $\boldsymbol{v}$ y $r(\cdot)$ es una función de recorte que recorta el valor de la entrada de datos en un intervalo determinado. Tal como se describe detalladamente más adelante, $s(\cdot)$ puede ser la función de suma o la función de segundo orden, pero se aprecia que en otras modalidades, $s(\cdot)$ puede ser cualquier función tal como la función de máximo. Los valores para el conjunto de parámetros $\boldsymbol{\theta}$ para las probabilidades implícitas por alelo se pueden determinar al minimizar la función de pérdida con respecto a $\boldsymbol{\theta}$, donde *i* es cada ejemplo en el subconjunto *S* de datos de entrenamiento 170 generados a partir de células que expresan los alelos individuales de MHC y/o células que expresan múltiples alelos de MHC.

[00628] La probabilidad de presentación en el modelo de presentación de la ecuación (17) se modela como una función de probabilidades de presentación implícita por alelo u'^{h}_{k} donde cada una corresponde a la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante un alelo de MHC *h* individual. La probabilidad implícita por alelo es distinta de la probabilidad de presentación por alelo de la sección X.B ya que los parámetros para las probabilidades implícitas por alelo se pueden aprender de múltiples configuraciones de alelos, en las cuales la asociación directa entre un péptido presentado y el alelo de MHC correspondiente es desconocida, además de las configuraciones de alelos individuales. Por tanto, en una configuración de múltiples alelos, el modelo de presentación no solamente puede estimar si el péptido $\boldsymbol{p^k}$ será presentado por un conjunto de alelos de MHC *H* por completo, sino que también puede proporcionar probabilidades individuales $u'^{h\in H}_{k}$ que indiquen qué alelo de MHC *h* tiene más probabilidad de haber presentado el péptido $\boldsymbol{p^k}$. Una ventaja de esto es que el modelo de presentación puede generar las probabilidades implícitas sin datos de entrenamiento para las células que expresan los alelos individuales de MHC.

[00629] En una implementación particular a la que se hace referencia a lo largo del resto de la memoria descriptiva, $r(\cdot)$ es una función que tiene el intervalo [0, 1]. Por ejemplo, $r(\cdot)$ puede ser la función de recorte: $r(z) = \text{mín}\,(\text{máx}(z, 0), 1)$,

donde el valor mínimo entre *z* y 1 se selecciona como la probabilidad de presentación u_k. En otra implementación, $r(\cdot)$ es la función de tangente hiperbólica proporcionada con:

$r(z) = \tanh(z)$

cuando los valores para el dominio *z* son iguales o mayores que 0.

X.C.5. Ejemplo 3.2: Modelo de suma de funciones

[00630] En una implementación particular, $s(\cdot)$ es una función de suma y la probabilidad de presentación se proporciona mediante la suma de las probabilidades de presentación implícita por alelo:

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados}\right) = r\left(\sum_{h=1}^{m} a_h^k \cdot u'^{h}_{k}(\boldsymbol{\theta})\right). (17)$$

[00631] En una implementación, la probabilidad de presentación implícita por alelo para el aleo de MHC *h* se genera mediante:

$$u'^{h}_{k} = f\left(g_h\left(\boldsymbol{x_h^k}; \boldsymbol{\theta_h}\right)\right), (18)$$

de modo que la probabilidad de presentación se estima con:

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados}\right) = r\left(\sum_{h=1}^{m} a_h^k \cdot f\left(g_h\left(\boldsymbol{x_h^k}; \boldsymbol{\theta_h}\right)\right)\right). (19)$$

[00632] De acuerdo con la ecuación (19), la probabilidad de presentación de que una secuencia

peptídica $\boldsymbol{p^k}$ se presente mediante uno o más alelos de MHC *H* se puede generar mediante la aplicación de la función $g_h(\cdot)$ a la versión codificada de la secuencia peptídica $\boldsymbol{p^k}$ para cada uno de los alelos de MHC *H* para generar la puntuación de dependencia correspondiente para las variables de interacción con alelos. Cada puntuación de dependencia se transforma primero con la función $f(\cdot)$ para generar las probabilidades de presentación implícita por alelo u'^h_k. Las probabilidades por alelo u'^h_k se combinan y se puede aplicar la función de recorte a las probabilidades combinadas para recortar los valores en un intervalo [0, 1] para generar la probabilidad de presentación de que la secuencia peptídica $\boldsymbol{p^k}$ se presente mediante un conjunto de alelos de MHC *H.* La función de dependencia $\boldsymbol{g_h}$ puede encontrarse en forma de cualquiera de las funciones de dependencia $\boldsymbol{g_h}$ introducidas anteriormente en las secciones X.B.1.

[00633] A modo de ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *h=2, h=3*, entre *m=4* alelos de MHC identificados diferentes con el uso de las funciones de transformación afín $g_h(\cdot)$, se puede generar con:

$$u_k = r\left(f\left(\boldsymbol{x_2^k}\cdot\boldsymbol{\theta_2}\right) + f\left(\boldsymbol{x_3^k}\cdot\boldsymbol{\theta_3}\right)\right),$$

donde $\boldsymbol{x_2^k}$, $\boldsymbol{x_3^k}$ son las variables de interacción con alelos para los alelos de MHC *h=2, h=3* y $\boldsymbol{\theta_2}$, $\boldsymbol{\theta_3}$ son el conjunto de parámetros determinado para los alelos de MHC *h=2, h=3*.

[00634] A modo de otro ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *h=2, h=3*, entre *m=4* alelos de MHC identificados diferentes con el uso de las funciones de transformación en red $g_h(\cdot)$, $g_w(\cdot)$, se puede generar mediante:

$$u_k = r\left(f\left(NN_2(\boldsymbol{x_2^k};\boldsymbol{\theta_2})\right) + f\left(NN_3(\boldsymbol{x_3^k};\boldsymbol{\theta_3})\right)\right),$$

donde $NN_2(\cdot)$, $NN_3(\cdot)$ son los modelos de red identificados para los alelos de MHC *h=2, h=3* y $\boldsymbol{\theta_2}$, $\boldsymbol{\theta_3}$ son el conjunto de parámetros determinado para los alelos de MHC *h=2, h=3*.

[00635] La Figura 11 ilustra la generación de una probabilidad de presentación para el péptido $\boldsymbol{p^k}$ asociado a los alelos de MHC *h*=2, *h*=3 con el uso de modelos de red ilustrativos $NN_2(\cdot)$ y $NN_3(\cdot)$. Como se muestra en la Figura 9, el modelo de red $NN_2(\cdot)$ recibe las variables de interacción con el alelo $\boldsymbol{x_2^k}$ para el alelo de MHC *h=2* y genera el producto $NN_2(\boldsymbol{x_2^k})$ y el modelo de red $NN_3(\cdot)$ recibe las variables de interacción con el alelo $\boldsymbol{x_3^k}$ para el alelo de MHC *h=3* y genera el producto $NN_3(\boldsymbol{x_3^k})$. Cada producto se mapea con la función $f(\cdot)$ y se combina para generar la probabilidad de presentación estimada u_k.

[00636] En otra implementación, cuando se hacen predicciones para el registro de las corrientes iónicas por espectrometría de masas, $r(\cdot)$ es la función logarítmica y $f(\cdot)$ es la función exponencial.

X.C.6. Ejemplo 3.3: Modelos de funciones de sumas con variables de no interacción con alelos

[00637] En una implementación, la probabilidad de presentación implícita por alelo para el aleo de MHC *h* se genera mediante:

$$u'^h_k = f\left(g_h\left(\boldsymbol{x_h^k};\boldsymbol{\theta_h}\right) + g_w\left(\boldsymbol{w^k};\boldsymbol{\theta_w}\right)\right), \quad (20)$$

de modo que la probabilidad de presentación se genere con:

$$u_k = \Pr(\boldsymbol{p^k} \text{ presentados}) = r\left(\sum_{h=1}^{m} a_h^k \cdot f\left(g_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}) + g_h(\boldsymbol{x_h^k}; \boldsymbol{\theta_h})\right)\right), \quad (21)$$

para incorporar el impacto de las variables de no interacción con alelos en la presentación del péptido.

[00638] De acuerdo con la ecuación (21), la probabilidad de presentación de que una secuencia peptídica $\boldsymbol{p^k}$ se presente con uno o más alelos de MHC *H* se puede generar mediante la aplicación de la función $g_h(\cdot)$ a la versión codificada de la secuencia peptídica $\boldsymbol{p^k}$ para cada uno de los alelos de MHC *H* para generar la puntuación de dependencia correspondiente para las variables de interacción con alelos para cada alelo de MHC *h*. La función $g_w(\cdot)$ para las variables de no interacción con alelos también se aplica a la versión codificada de las variables de no interacción con alelos para generar la puntuación de dependencia para las variables de no interacción con alelos. La puntuación para las variables de no interacción con alelos se combina para cada una de las puntuaciones de dependencia para las variables de interacción con alelos. Cada una de las puntuaciones combinadas se transforman mediante la función $f(\cdot)$ para generar las probabilidades de presentación implícita por alelo. Las probabilidades implícitas se combinan y es posible aplicar la función de recorte a los productos combinados para recortar los valores en un intervalo [0,1] para generar la probabilidad de presentación de que la secuencia peptídica $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *H.* La función de dependencia $\boldsymbol{g_w}$ puede encontrarse en forma de cualquiera de las funciones de dependencia $\boldsymbol{g_w}$ introducidas anteriormente en las secciones X.B.3.

[00639] A modo de ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *h=2, h=3*, entre *m=4* alelos de MHC diferentes identificados con el uso de las funciones de transformación afín $g_h(\cdot)$, $g_w(\cdot)$, se puede generar mediante:

$$u_k = r\left(f\left(\boldsymbol{w^k} \cdot \boldsymbol{\theta_w} + \boldsymbol{x_2^k} \cdot \boldsymbol{\theta_2}\right) + f\left(\boldsymbol{w^k} \cdot \boldsymbol{\theta_w} + \boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right)\right),$$

donde $\boldsymbol{w^k}$ son las variables identificadas de no interacción con alelos para el péptido $\boldsymbol{p^k}$ y $\boldsymbol{\theta_w}$ son el conjunto de parámetros determinado para las variables de no interacción con alelos.

[00640] A modo de otro ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *h=2, h=3*, entre *m=4* alelos de MHC identificados diferentes con el uso de las funciones de transformación en red $g_h(\cdot)$, $g_w(\cdot)$, se puede generar mediante:

$$u_k = r\left(f\left(NN_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}) + NN_2(\boldsymbol{x_2^k}; \boldsymbol{\theta_2})\right) + f\left(NN_w(\boldsymbol{w^k}; \boldsymbol{\theta_w}) + NN_3(\boldsymbol{x_3^k}; \boldsymbol{\theta_3})\right)\right)$$

donde $\boldsymbol{w^k}$ son las variables de interacción con alelos para el péptido $\boldsymbol{p^k}$ y $\boldsymbol{\theta_w}$ son el conjunto de parámetros determinado para las variables de interacción con alelos.

[00641] La Figura 12 ilustra la generación de una probabilidad de presentación para el péptido $\boldsymbol{p^k}$ asociado a los alelos de MHC *h=2, h*=3 con el uso de modelos de red ilustrativos $NN_2(\cdot)$, $NN_3(\cdot)$ y $NN_w(\cdot)$. Como se muestra en la Figura 12, el modelo de red $NN_2(\cdot)$ recibe las variables de interacción con alelos

$\boldsymbol{x_2^k}$ para el alelo de MHC *h=2* y genera el producto $NN_2(\boldsymbol{x_2^k})$. El modelo de red $NN_w(\cdot)$ recibe las variables de no interacción con alelos $\boldsymbol{w^k}$ para el péptido $\boldsymbol{p^k}$ y genera el producto $NN_w(\boldsymbol{w^k})$. Los productos se combinan y se mapean con la función $f(\cdot)$. El modelo de red $NN_3(\cdot)$ recibe las variables de interacción con alelos $\boldsymbol{x_3^k}$ para el alelo de MHC *h=3* y genera el producto $NN_3(\boldsymbol{x_3^k})$, que se combina nuevamente con el producto $NN_w(\boldsymbol{w^k})$ del mismo modelo de red $NN_w(\cdot)$ y se mapea mediante la función $f(\cdot)$. Ambos productos se combinan para generar la probabilidad de presentación estimada u_k.

[00642] En otra implementación, la probabilidad de presentación implícita por alelo para el alelo de MHC *h* se genera mediante:

$$u_k'^{h} = f\left(g_h\left([\boldsymbol{x_h^k}\, \boldsymbol{w^k}]; \boldsymbol{\theta_h}\right)\right). \quad (22)$$

de modo que la probabilidad de presentación se genere con:

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados}\right) = r\left(\sum_{h=1}^{m} a_h^k \cdot f\left(g_h\left([\boldsymbol{x_h^k}\, \boldsymbol{w^k}]; \boldsymbol{\theta_h}\right)\right)\right).$$

X.C.7. Ejemplo 4: Modelos de segundo orden

[00643] En una implementación, $s(\cdot)$ es una función de segundo orden y la probabilidad de presentación estimada u_k para el péptido $\boldsymbol{p^k}$ se proporciona con:

$$u_k = \Pr\left(\boldsymbol{p^k} \text{ presentados}\right) = \sum_{h=1}^{m} a_h^k \cdot u_k'^{h}(\boldsymbol{\theta}) - \sum_{h=1}^{m}\sum_{j<h} a_h^k \cdot a_j^k \cdot u_k'^{h}(\boldsymbol{\theta}) \cdot u_k'^{j}(\boldsymbol{\theta}) \quad (23)$$

donde los elementos u'^{h}_k son la probabilidad de presentación implícita por alelo para el alelo de MHC *h*. Los valores para el conjunto de parámetros $\boldsymbol{\theta}$ para las probabilidades implícitas por alelo se pueden determinar al minimizar la función de pérdida con respecto a $\boldsymbol{\theta}$, donde *i* es cada ejemplo en el subconjunto *S* de datos de entrenamiento 170 generados a partir de células que expresan los alelos individuales de MHC y/o células que expresan múltiples alelos de MHC. Las probabilidades de presentación implícita por alelo se pueden encontrar en cualquier forma de las que se muestran en las ecuaciones (18), (20) y (22) que se describieron anteriormente.

[00644] En un aspecto, el modelo de ecuación (23) puede implicar que existe una posibilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante dos de los alelos de MHC simultáneamente, en la cual la presentación con dos alelos de HLA es independiente desde el punto de vista estadístico.

[00645] De acuerdo con la ecuación (23), la probabilidad de presentación de que una secuencia peptídica $\boldsymbol{p^k}$ se presente mediante uno o más alelos de MHC *H* se puede generar mediante la combinación de las probabilidades de presentación implícita por alelo y la resta de la probabilidad de que cada par de los alelos de MHC presente simultáneamente el péptido $\boldsymbol{p^k}$ de la suma para generar la probabilidad de presentación de que la secuencia peptídica $\boldsymbol{p^k}$ se presente mediante los alelos de MHC *H*.

[00646] A modo de ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de HLA *h=2, h=3*, entre *m=4* alelos de HLA identificados diferentes con el uso de las funciones de transformación afín $g_h(\cdot)$, se puede generar con:

$$u_k = f\left(\boldsymbol{x_2^k} \cdot \boldsymbol{\theta_2}\right) + f\left(\boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right) - f\left(\boldsymbol{x_2^k} \cdot \boldsymbol{\theta_2}\right) \cdot f\left(\boldsymbol{x_3^k} \cdot \boldsymbol{\theta_3}\right),$$

donde $\boldsymbol{x_2^k}$, $\boldsymbol{x_3^k}$ son las variables de interacción con alelos para los alelos de HLA *h=2, h=3* y $\boldsymbol{\theta_2}$, $\boldsymbol{\theta_3}$ son el conjunto de parámetros determinado para los alelos de HLA *h=2, h=3*.

[00647] A modo de otro ejemplo, la probabilidad de que el péptido $\boldsymbol{p^k}$ se presente mediante los alelos de HLA *h=2, h=3*, entre *m=4* alelos de HLA identificados diferentes con el uso de las funciones de transformación en red $g_h(\cdot)$, $g_w(\cdot)$, se puede generar mediante:

$$u_k = f\left(NN_2(\boldsymbol{x_2^k}; \boldsymbol{\theta_2})\right) + f\left(NN_3(\boldsymbol{x_3^k}; \boldsymbol{\theta_3})\right) - f\left(NN_2(\boldsymbol{x_2^k}; \boldsymbol{\theta_2})\right) \cdot f\left(NN_3(\boldsymbol{x_3^k}; \boldsymbol{\theta_3})\right),$$

donde $NN_2(\cdot)$, $NN_3(\cdot)$ son los modelos de red identificados para los alelos de HLA *h=2, h=3* y $\boldsymbol{\theta_2}$, $\boldsymbol{\theta_3}$ son el conjunto de parámetros determinado para los alelos de HLA *h=2, h=3*.

XI.A Ejemplo 5: Módulo de predicción

[00648] El módulo de predicción 320 recibe los datos de la secuencia y selecciona los neoantígenos candidatos en los datos de la secuencia con el uso de modelos de presentación. Específicamente, los datos de la secuencia pueden ser secuencias de ADN, secuencias de ARN y/o secuencias proteicas extraídas de células de tejidos tumorales de pacientes. El módulo de predicción 320 procesa los datos de la secuencia en múltiples secuencias peptídicas $\boldsymbol{p^k}$ que tienen 8-15 aminoácidos para MHC-I o 6-30 aminoácidos para MHC-II. Por ejemplo, el módulo de predicción 320 puede procesar la secuencia determinada «IEFROEIFJEF (SEQ ID NO: 76)» en tres secuencias peptídicas con 9 aminoácidos «IEFROEIFJ (SEQ ID NO: 77)», «EFROEIFJE (SEQ ID NO: 78),» y «FROEIFJEF (SEQ ID NO: 79)». En una modalidad, el módulo de predicción 320 puede identificar los neoantígenos candidatos que son secuencias peptídicas mutadas al comparar los datos de secuencia extraídos de células de tejidos normales de un paciente con los datos de secuencia extraídos de células de tejidos tumorales del paciente para identificar partes que contengan una o más mutaciones.

[00649] El módulo de presentación 320 aplica uno o más modelos de presentación a las secuencias peptídicas procesadas para estimar las probabilidades de presentación de las secuencias peptídicas. Específicamente, el módulo de predicción 320 puede seleccionar una o más secuencias peptídicas de neoantígenos candidatos que tengan probabilidades de presentarse en moléculas HLA tumorales mediante la aplicación de modelos de presentación a los neoantígenos candidatos. En una implementación, el módulo de presentación 320 selecciona las secuencias de antígenos candidatos que tienen probabilidad de presentación estimadas por encima de un umbral predeterminado. En otra implementación, el modelo de presentación selecciona las secuencias *N* de neoantígenos candidatos que tengan las probabilidades de presentación estimadas más elevadas (donde *N* es en general la

cantidad máxima de epítopos que se pueden suministrar en una vacuna). Una vacuna que incluye los neoantígenos candidatos seleccionados para un paciente determinado puede inyectarse en el paciente para inducir respuestas inmunitarias.

XI.B. Ejemplo 6: Módulo de diseño de casete.

XI.B.1 Resumen

[00650] El módulo de diseño de casete 324 genera una secuencia de casete para vacuna según los péptidos *v* candidatos seleccionados para la inyección en un paciente. Específicamente, para un conjunto de péptidos seleccionados $\boldsymbol{p^k}$, *k=1, 2, …, v* para la inclusión en una vacuna de capacidad *v*, la secuencia de casete se proporciona mediante concatenación de una serie de secuencias terapéuticas de epítopos $\boldsymbol{p'^k}$, *k=1, 2, …, v* que incluye, cada una, la secuencia de un péptido $\boldsymbol{p^k}$ correspondiente. En una modalidad, el módulo de diseño de casete 324 puede concatenar los epítopos directamente contiguos entre sí. Por ejemplo, un casete de vacuna **C** puede representarse como:

$$\mathbf{C} = [\boldsymbol{p'^{t_1}}\ \boldsymbol{p'^{t_2}}\ \cdots\ \boldsymbol{p'^{t_v}}] \quad (24)$$

donde $\boldsymbol{p'^{ti}}$ denota el epítopo número *i* del casete. Por lo tanto, t_i corresponde a un índice *k=1, 2, …, v* para el péptido seleccionado en la posición número *i* del casete. En otra modalidad, el módulo de diseño del casete 324 puede concatenar los epítopos con una o más secuencias enlazadoras opcionales entre los epítopos contiguos. Por ejemplo, un casete para vacuna **C** se puede representar como:

$$\mathbf{C} = [\boldsymbol{p'^{t_1}}\ \boldsymbol{l}_{(\mathbf{t_1},\mathbf{t_2})}\ \boldsymbol{p'^{t_2}}\ \boldsymbol{l}_{(\mathbf{t_2},\mathbf{t_3})} \cdots\ \boldsymbol{l}_{(\mathbf{t}_{v-1},\mathbf{t}_v)}\ \boldsymbol{p'^{t_v}}] \quad (25)$$

donde $\boldsymbol{l_{(ti,tj)}}$ denota una secuencia enlazadora ubicada entre el epítopo número *i* $\boldsymbol{p'^{ti}}$ y el epítopo número *j=i+1* $\boldsymbol{p'^{j=i+1}}$ del casete. El módulo de diseño del casete 324 determina cuál de los epítopos seleccionados $\boldsymbol{p'^k}$, *k=1, 2, …, v* se disponen en posiciones diferentes del casete, así como cualquier secuencia enlazadora que se coloque entre los epítopos. Una secuencia de casete **C** se puede cargar como vacuna según cualquiera de los métodos descritos en la presente memoria descriptiva.

[00651] En una modalidad, el conjunto de epítopos terapéuticos puede generarse según los péptidos seleccionados determinados por el módulo de predicción 320 asociado con las probabilidades de presentación por encima de un umbral predeterminado, donde las probabilidades de presentación se determinan mediante los modelos de presentación. Sin embargo, se aprecia que en otras modalidades, el conjunto de epítopos terapéuticos se puede generar según uno o más de una cantidad de cualesquiera métodos (solos o combinados), por ejemplo, según la afinidad de unión o la afinidad de unión predicha a los alelos de HLA clase I o clase II del paciente, la estabilidad de unión o la estabilidad de unión predicha a los alelos de HLA clase I o clase II del paciente, muestreo aleatorio y similares.

[00652] En una modalidad, los epítopos terapéuticos $\boldsymbol{p'^k}$ pueden corresponder a los propios péptidos seleccionados $\boldsymbol{p^k}$. En otra modalidad, los epítopos terapéuticos $\boldsymbol{p'^k}$ también pueden incluir secuencias flanqueantes de extremo C y/o N, además de los péptidos seleccionados. Por ejemplo, un epítopo $\boldsymbol{p'^k}$

incluido en el casete se puede representar como una secuencia [n^k p^k c^k] donde c^k es una secuencia flanqueante de extremo C una al extremo C del péptido seleccionado p^k y n^k es una secuencia flanqueante de extremo N unida al extremo N del péptido seleccionado p^k. En un ejemplo al que se hace referencia a lo largo del resto de la memoria descriptiva, las secuencias flanqueantes de extremo N y C son secuencias flanqueantes naturales de extremo N y C del epítopo terapéutico para vacunas en el contexto de su proteína fuente. En un ejemplo al que se hace referencia a lo largo del resto de la memoria descriptiva, el epítopo terapéutico p'^k representa un epítopo de longitud fija. En otro ejemplo, el epítopo terapéutico p'^k puede representar un epítopo de longitud variable, en el cual la longitud del epítopo puede variar, por ejemplo, según la longitud de la secuencia flanqueante C o N. Por ejemplo, cada una de la secuencia flanqueante de extremo C c^k y la secuencia flanqueante de extremo N n^k tiene longitudes diversas de 2-5 residuos, lo que produce 16 elecciones posibles para el epítopo p'^k.

[00653] En una modalidad, el módulo de diseño de casete 324 genera secuencias de casete teniendo en cuenta la presentación de epítopos de unión que abarquen la unión entre un par de epítopos terapéuticos en el casete. Los epítopos de unión son secuencias de epítopos no propios novedosos pero irrelevantes que surgen en el casete debido al proceso de concatenación de los epítopos terapéuticos y las secuencias enlazadoras en el casete. Las secuencias novedosas de epítopos de unión son diferentes de los epítopos terapéuticos propios del casete. Un epítopo de unión que abarca los epítopos p'^{ti} y p'^{tj} puede incluir cualquier secuencia de epítopos que se superponga tanto con p'^{ti} como con p'^{tj} que sea diferente a las secuencias de los propios epítopos terapéuticos p'^{ti} y p'^{tj}. Específicamente, cada unión entre el epítopo p'^{ti} y un epítopo contiguo p'^{tj} del casete con o sin una secuencia enlazadora opcional $l^{(ti,tj)}$ se puede asociar a $n_{(ti,tj)}$ epítopos de unión $e_n^{(ti,tj)}$, $n=1, 2, \ldots, n_{(ti,tj)}$. Los epítopos de unión pueden ser secuencias que se superponen al menos parcialmente tanto con el epítopo p'^{ti} como con p'^{tj} o pueden ser secuencias que se superponen al menos parcialmente con secuencias enlazadoras colocadas entre los epítopos p'^{ti} y p'^{tj}. Los epítopos de unión pueden presentarse mediante los MHC clase I, MHC clase II o ambos.

[00654] La Figura 38 muestra dos secuencias de casete ilustrativas, el casete 1 ($\mathbf{C_1}$) y el casete 2 ($\mathbf{C_2}$). Cada casete tiene una capacidad para vacuna de $v=2$ e incluye los epítopos terapéuticos $p'^{t1} = p^1$ = SINFEKL (SEQ ID NO: 80) y $p'^{t2} = p^2$ = LLLLLVVVV (SEQ ID NO: 81) y una secuencia enlazadora $l^{(t1,t2)}$ = AAY entre los dos epítopos. Específicamente, la secuencia de casete $\mathbf{C_1}$ se proporciona mediante [p^1 $l^{(t1,t2)}$ p^2], mientras que la secuencia de casete $\mathbf{C_2}$ se proporciona mediante [p^2 $l^{(t1,t2)}$ p^1]. Los epítopos de unión ilustrativos $e_n^{(1,2)}$ del casete $\mathbf{C_1}$ pueden ser secuencias tal como EKLAAYLLL (SEQ ID NO: 82), KLAAYLLLL (SEQ ID NO: 83) y FEKLAAYL (SEQ ID NO: 84) que abarcan tanto el epítopo p'^1 como p'^2 en el casete y pueden ser secuencias tal como AAYLLLLL (SEQ ID NO: 85) y YLLLLLVVV (SEQ ID NO: 86) que abarcan la secuencia enlazadora y un único epítopo seleccionado en el casete. De manera

similar, los epítopos de unión ilustrativos $\boldsymbol{e}_m^{(2,1)}$ del casete $\mathbf{C_2}$ pueden ser secuencias tal como VVVVAAYSIN (SEQ ID NO: 87), VVVVAAY (SEQ ID NO: 88) y AYSINFEK (SEQ ID NO: 89). Si bien ambos casetes implican el mismo conjunto de secuencias $\boldsymbol{p^1}$, $\boldsymbol{I}^{(c1,c2)}$ y $\boldsymbol{p^2}$, el conjunto de epítopos de unión que se identifica es diferente según la secuencia ordenada de los epítopos terapéuticos dentro del casete.

[00655] En una modalidad, el módulo de diseño de casete 324 genera una secuencia de casete que reduce la probabilidad de que los epítopos de unión se presenten en el paciente. Específicamente, cuando se inyecta el casete en el paciente, los epítopos de unión tienen la capacidad de presentarse mediante los alelos de HLA clase I o de HLA clase II del paciente y estimular respectivamente una respuesta de linfocitos T CD8 o CD4. Tales reacciones son a veces indeseables ya que los linfocitos T que reaccionan a los epítopos de unión no tienen un beneficio terapéutico y pueden disminuir la respuesta inmunitaria a los epítopos terapéuticos seleccionados en el casete mediante competencia de antígenos.[76]

[00656] En una modalidad, el módulo de diseño del casete 324 se repite a través de uno o más casetes candidatos y determina una secuencia de casete para la cual una puntuación de presentación de los epítopos de unión asociados a la secuencia del casete se encuentra por debajo de un umbral numérico. La puntuación de presentación de epítopos de unión es una cantidad asociada a las probabilidades de presentación de los epítopos de unión en el casete y un valor mayor de puntuación de presentación de epítopos de unión indica una probabilidad mayor de que los epítopos de unión del casete se presenten mediante HLA clase I o HLA clase II, o ambos.

[00657] En una modalidad, el módulo de diseño de casete 324 puede determinar una secuencia de casete asociada a la puntuación de presentación de epítopos menor entre las secuencias de casete candidatos. En un ejemplo, la puntuación de presentación para una secuencia de casete determinada **C** se determina según un conjunto de mediciones de distancia $d(\boldsymbol{e}_n^{(ti,tj)}, n=1, 2, \ldots, n_{(ti,tj)}) = d_{(ti,tj)}$, cada una asociada a una unión en el casete **C**. Específicamente, una medición de distancia $d_{(ti,tj)}$ especifica una probabilidad de que se presenten uno o más de los epítopos de unión que abarquen entre el par de epítopos terapéuticos contiguos $\boldsymbol{p}^{\pi ti}$ y $\boldsymbol{p}^{\pi tj}$. La puntuación de presentación de epítopos de unión para el casete **C** se puede determinar entonces mediante la aplicación de una función (por ejemplo, suma, función estadística) al conjunto de mediciones de distancia para el casete C. Matemáticamente, la puntuación de presentación se proporciona mediante:

$$puntuación = h(d_{(t_1,t_2)}, d_{(t_2,t_3)}, \ldots, d_{(t_{v-1},t_v)}) \quad (26)$$

donde $h(\cdot)$ es alguna función que mapea las mediciones de distancia de cada unión a una puntuación. En un ejemplo particular al que se hace referencia a lo largo del resto de la memoria descriptiva, la función $h(\cdot)$ es la suma de las mediciones de distancia del casete.

[00658] El módulo de diseño de casete 324 puede repetirse a lo largo de una o más secuencias de casete candidatos, determinar la puntuación de presentación de epítopos de unión para los casetes candidatos e identificar una secuencia de casete óptima asociada a una puntuación de epítopos de unión por debajo del umbral. En una modalidad en particular a la que se hace referencia a lo largo del resto de la memoria descriptiva, la medición de distancia $d(\cdot)$ para una unión determinada se puede proporcionar mediante la suma de las probabilidades de presentación o la cantidad esperada de epítopos de unión presentados tal como se determina mediante modelos de presentación descritos en las secciones VII y VIII de la memoria descriptiva. Sin embargo, se aprecia que en otras modalidades, la medición de distancia puede derivar de otros factores, solos o combinados con los modelos, como el que se ejemplificó anteriormente, donde estos otros factores pueden incluir la derivación de la medición de distancia de cualquiera de uno o más de los siguientes (solos o combinados): las mediciones de afinidad o estabilidad de unión de HLA o las predicciones para HLA clase I o HLA clase II y un modelo de presentación o inmunogenicidad preparado con datos de espectrometría de masa de HLA o de epítopos de linfocitos T, para HLA clase I o HLA clase II. En una modalidad, la medición de distancia puede combinar información sobre la presentación de HLA clase I y HLA clase II. Por ejemplo, la medición de distancia puede ser el número de epítopos de unión que se predice que se van a unir a cualquiera de los alelos de HLA clase I o HLA clase II del paciente con una afinidad de unión por debajo de un umbral. En otro ejemplo, la medición de distancia puede ser la cantidad esperada de epítopos que se predice que se van a presentar mediante cualquiera de los alelos de HLA clase I o HLA clase II del paciente.

[00659] El módulo de diseño de casete 324 puede verificar adicionalmente una o más secuencias de casetes candidatos para identificar si alguno de los epítopos de unión en las secuencias de casetes candidatos son epítopos propios para un paciente determinado para el cual se está diseñando la vacuna. Para lograr esto, el módulo de diseño de casete 324 verifica los epítopos de unión contra una base de datos conocida, tal como BLAST. En una modalidad, el módulo de diseño de casete se puede configurar para diseñar casetes que eviten los epítopos de unión propios al configurar la medición de distancia $d_{(ti,tj)}$ en un valor muy grande (por ejemplo, 100) para pares de epítopos t_i,t_j donde la concatenación del epítopo t_i al extremo N del epítopo t_j produce la formación de un epítopo de unión propio.

[00660] Nuevamente con referencia al ejemplo en la Figura 38, el módulo de diseño de casete 324 determina (por ejemplo) una medición de distancia $d_{(t1,t2)} = d_{(1,2)} = 0{,}39$ para la unión individual (t_1,t_2) en el casete $\mathbf{C_1}$ proporcionada por la suma de las probabilidades de presentación de todos los epítopos de unión posibles $\boldsymbol{e}_n^{(t1,t2)} = \boldsymbol{e}_n^{(1,2)}$ con longitudes, por ejemplo, de 8 a 15 aminoácidos para MHC clase I o 6-30 aminoácidos para MHC clase II. Dado que no se encuentran presentes otras uniones en el casete $\mathbf{C_1}$, la puntuación de presentación de epítopos de unión, que es una suma de las mediciones de distancia

para el casete $\mathbf{C_1}$, también se proporciona con 0,39. El módulo de diseño de casete 324 también determina una medición de distancia $d_{(t1,t2)} = d_{(2,1)}$ = 0,068 para la unión individual en el casete $\mathbf{C_2}$ que se proporciona mediante la suma de las probabilidades de presentación de todos los epítopos de unión posibles $\boldsymbol{e}_n^{(t1,t2)} = \boldsymbol{e}_n^{(2,1)}$ con longitudes de 8 a 15 para MHC clase I o 9-30 aminoácidos para MHC clase II. En el presente ejemplo, la puntuación de epítopos de unión para el casete $\mathbf{C_2}$ también se proporciona mediante la medición de distancia de la unión individual 0,068. El módulo de diseño de casete 324 produce la secuencia de casete de $\mathbf{C_2}$ como casete óptimo ya que la puntuación de presentación de epítopos de unión es menor que la de la secuencia de casete de $\mathbf{C_1}$.

[00661] En algunos casos, el módulo de diseño de casete 324 puede llevar a cabo un enfoque de fuerza bruta y repetirse a lo largo de todas o la mayoría de las secuencias de casetes candidatos posibles para seleccionar la secuencia con la menor puntuación de presentación de epítopos de unión. Sin embargo, la cantidad de tales casetes candidatos puede ser prohibitivamente grande a medida que aumenta la capacidad de la vacuna *v*. Por ejemplo, para una capacidad de vacuna de *v=20* epítopos, el módulo de diseño de casete 324 tiene que repetir ~10^{18} casetes candidatos posibles para determinar el casete con la menor puntuación de presentación de epítopos de unión. Esta determinación puede resultar pesada desde el punto de vista informático (en términos de recursos de procesamiento informático necesarios) y a veces inabordable, para que el módulo de diseño de casete 324 se complete dentro de una cantidad de tiempo razonable para generar la vacuna para el paciente. Además, contabilizar los posibles epítopos de unión para cada casete candidato puede ser aún más engorroso. Por tanto, el módulo de diseño de casete 324 puede seleccionar una secuencia de casete según formas de repetición mediante una cantidad de secuencias de casetes candidatos que sea significativamente menor que la cantidad de secuencias de casetes candidatos para el enfoque por fuerza bruta.

[00662] En una modalidad, el módulo de diseño de casete 324 genera un subconjunto de casetes candidatos generados aleatoriamente, o al menos en forma pseudoaleatoria, y selecciona el casete candidato asociado a una puntuación de presentación de epítopos de unión menor que un umbral predeterminado como secuencia de casete. Adicionalmente, el módulo de diseño de casete 324 puede seleccionar el casete candidato del subconjunto con la menor puntuación de presentación de epítopos de unión como secuencia de casete. Por ejemplo, el módulo de diseño de casete 324 puede generar un subconjunto de ~1 millón de casetes candidatos para un conjunto de *v=20* epítopos seleccionados y seleccionar el casete candidato con una puntuación mínima de presentación de epítopos de unión. Si bien la generación de un subconjunto de secuencias aleatorias de casetes y la selección de una secuencia de casete con una puntuación escasa de presentación de epítopos de unión del subconjunto puede ser menos que óptima con respecto al enfoque por fuerza bruta, requiere significativamente menos recursos informáticos, lo que hace que su implementación sea posible desde el punto de vista

técnico. Además, llevar a cabo el método por fuerza bruta en oposición a esta técnica más eficiente solo puede producir una mejora menor o incluso insignificante en la puntuación de presentación de epítopos de unión, lo que hace que no valga la pena desde una perspectiva de adjudicación de recursos.

[00663] En otra modalidad, el módulo de diseño de casete 324 determina una configuración de casete mejorada mediante la formulación de la secuencia de epítopos para el casete como un problema del viajante asimétrico (TSP, por su sigla en inglés). Dada una lista de nodos y distancias entre cada par de nodos, el TSP determina una secuencia de nodos asociada a la distancia total menor para visitar cada nodo exactamente una vez y regresar al nodo original. Por ejemplo, dadas las ciudades A, B y C con distancias conocidas entre cada una, la solución del TSP genera una secuencia cerrada de ciudades, para la cual la distancia total recorrida para visitar cada ciudad exactamente una vez es la menor entre las rutas posibles. La versión asimétrica del TSP determina la secuencia óptima de nodos cuando la distancia entre un par de nodos es asimétrica. Por ejemplo, la «distancia» para desplazarse del nodo A al nodo B puede ser diferente de la «distancia» para desplazarse del nodo B al nodo A.

[00664] El módulo de diseño de casete 324 determina una secuencia de casete mejorada al resolver un TSP asimétrico, en el cual cada nodo corresponde a un epítopo terapéutico $\boldsymbol{p'^{k}}$. La distancia desde un nodo correspondiente al epítopo $\boldsymbol{p'^{k}}$ a otro nodo correspondiente al epítopo $\boldsymbol{p'^{m}}$ se proporciona mediante la medición de distancia de epítopos de unión $d_{(k,m)}$, mientras que la distancia desde el nodo correspondiente al epítopo $\boldsymbol{p'^{m}}$ al nodo correspondiente al epítopo $\boldsymbol{p'^{k}}$ se proporciona mediante la medición de distancia $d_{(m,k)}$ que puede ser diferente a la medición de distancia $d_{(k,m)}$. Al resolver un casete óptimo mejorado con el uso de un TSP asimétrico, el módulo de diseño de casete 324 puede encontrar una secuencia de casete que genere una puntuación de presentación menor en las uniones entre los epítopos del casete. La solución del TSP asimétrico indica una secuencia de epítopos terapéuticos que corresponden al orden en el cual debería concatenarse los epítopos en un casete para minimizar la puntuación de presentación de epítopos de unión en las uniones del casete. Específicamente, dado un conjunto de epítopos terapéuticos $k=1, 2, \ldots, v$, el módulo de diseño de casete 324 determina las mediciones de distancia $d_{(k,m)}$, $k,m = 1, 2, \ldots, v$ para cada par ordenado posible de epítopos terapéuticos en el casete. En otras palabras, para un par determinado k, m de epítopos, se determina tanto la medición de distancia $d_{(k,m)}$ para la concatenación del epítopo terapéutico $\boldsymbol{p'^{m}}$ luego del epítopo $\boldsymbol{p'^{k}}$ como la medición de distancia $d_{(m,k)}$ para la concatenación del epítopo terapéutico $\boldsymbol{p'^{k}}$ luego del epítopo $\boldsymbol{p'^{m}}$, ya que estas mediciones de distancia pueden ser diferentes entre sí.

[00665] En una modalidad, el módulo de diseño de casete 324 resuelve el TSP asimétrico mediante un problema de programación lineal con enteros. Específicamente, el módulo de diseño de casete 324 genera una matriz de vía $(v+1) \times (v+1)$ **P** determinada por lo siguiente:

$\mathbf{P} = \begin{bmatrix} 0 & \mathbf{0}^{1\times v} \\ \mathbf{0}^{v\times 1} & \mathbf{D} \end{bmatrix}$ (26).

La matriz *v* x *v* **D** es una matriz de distancia asimétrica, donde cada elemento D(*k, m*), *k=1, 2, …, v; m=1, 2, …, v* corresponde a la medición de distancia para una unión del epítopo ***p'k*** al epítopo ***p'm***. Las filas *k = 2, …, v* de **P** corresponden a nodos de los epítopos originales, mientras que la fila 1 y la columna 1 corresponden a un «nodo fantasma» que se encuentra en la distancia cero de todos los otros nodos. La adición del «nodo fantasma» a la matriz codifica la noción de que el casete para la vacuna es lineal en lugar de circular, de modo que no hay unión entre el primer y el último epítopo. En otras palabras, la secuencia no es circular y el primer epítopo no se supone concatenado luego del último epítopo en la secuencia. Sea x_{km} una variable binaria cuyo valor es 1 si hay una vía dirigida (es decir, una unión de un epítopo con otro en el casete) donde el epítopo ***p'k*** se concatena con el extremo N del epítopo ***p'm*** y 0 de otro modo. Además, sea *E* el conjunto de todos los epítopos terapéuticos para vacuna *v* y sea $S \subset E$ un subconjunto de epítopos. Para cualquiera de tal subconjunto *S*, sea *out*(*S*) la cantidad de uniones de un epítopo con otro $x_{km} = 1$ donde *k* es un epítopo en *S* y *m* es un epítopo en *E**S*. Dada una matriz de vía **P** conocida, el módulo de diseño de casete 324 encuentra una matriz de vía **X** que resuelve el siguiente problema de programación lineal de enteros:

$$\min_{x} \sum_{k=1}^{v+1} \sum_{k\neq m, m=1}^{v+1} \mathrm{P}_{km} \cdot x_{km} \quad (27)$$

en el cual P_{km} denota el elemento P(*k,m*) de la matriz de vía **P**, sujeto a las siguientes limitaciones:

$$\sum_{k=1}^{v+1} x_{km} = 1\,, m = 1, 2, \ldots, v+1$$

$$\sum_{m=1}^{v+1} x_{km} = 1\,, k = 1, 2, \ldots, v+1$$

$$x_{kk} = 0, k = 1, 2, \ldots, v+1$$

$$\mathrm{out(S)} \geq 1, S \subset E, 2 \leq |S| \leq |V|/2$$

Las primeras dos limitaciones garantizan que cada epítopo aparezca exactamente una vez en el casete. La última limitación asegura que el casete se encuentre conectado. En otras palabras, el casete codificado por *x* es una secuencia de proteína lineal conectada.

[00666] Las soluciones para x_{km}, *k,m = 1, 2, …, v+1* en el problema de programación lineal de enteros de la ecuación (27) indica la secuencia cerrada de nodos y nodos fantasma que se pueden usar para inferir una o más secuencias de epítopos terapéuticos para el casete que disminuyen la puntuación de

presentación de los epítopos de unión. Específicamente, un valor de x_{km} = 1 indica que una «vía» existe del nodo *k* al nodo *m* o, en otras palabras, que el epítopo terapéutico $\boldsymbol{p'^m}$ debería concatenarse luego del epítopo terapéutico $\boldsymbol{p'^k}$ en la secuencia de casete mejorada. Una solución de x_{km} = 0 indica que no existe tal vía o, en otras palabras, que el epítopo terapéutico $\boldsymbol{p'^m}$ no debería concatenarse luego del epítopo terapéutico $\boldsymbol{p'^k}$ en la secuencia de casete mejorada. En conjunto, los valores de x_{km} en el problema de programación de enteros de la ecuación (27) representa una secuencia de nodos y el nodo fantasma, en la cual la vía entra y sale de cada nodo exactamente una vez. Por ejemplo, los valores de $x_{fantasma,1}=1$, $x_{13}=1$, $x_{32}=1$ *y* $x_{2,fantasma}=1$ (0 de otro modo) pueden indicar una secuencia *fantasma→1→3→2→fantasma* de nodos y nodos fantasma.

[00667] Una vez que se resolvió la secuencia, se eliminan los nodos fantasma de la secuencia para generar una secuencia refinada solamente con los nodos originales correspondientes a los epítopos terapéuticos en el casete. La secuencia refinada indica el orden en el cual deberían concatenarse los epítopos seleccionados en el casete para mejorar la puntuación de presentación. Por ejemplo, siguiendo el ejemplo del párrafo anterior, es posible eliminar el nodo fantasma para generar una secuencia refinada *1→3→2*. La secuencia refinada indica un modo posible de concatenar epítopos en el casete, a saber $\boldsymbol{p^1 \rightarrow p^3 \rightarrow p^2}$.

[00668] En una modalidad, cuando los epítopos terapéuticos $\boldsymbol{p'^k}$ son epítopos de longitud variable, el módulo de diseño de casete 324 determina las mediciones de distancia candidatas correspondientes a las diferentes longitudes de los epítopos terapéuticos $\boldsymbol{p'^k}$ y $\boldsymbol{p'^m}$ e identifica la medición de distancia $d_{(k,m)}$ como la menor medición de distancia candidata. Por ejemplo, cada uno de los epítopos $\boldsymbol{p'^k}$=[$\boldsymbol{n^k\ p^k\ c^k}$] y $\boldsymbol{p'^m}$=[$\boldsymbol{n^m\ p^m\ c^m}$] puede incluir una secuencia flanqueante de extremo N y C correspondiente que puede variar (en una modalidad) de 2-5 aminoácidos. Por lo tanto, la unión entre los epítopos $\boldsymbol{p'^k}$ y $\boldsymbol{p'^m}$ se asocia a 16 conjuntos diferentes de epítopos de unión según 4 valores de longitud posibles de $\boldsymbol{n^k}$ y los 4 valores de longitud posibles de $\boldsymbol{c^m}$ que se ubican en la unión. El módulo de diseño de casete 324 puede determinar las mediciones de distancia candidatas para cada conjunto de epítopos de unión y determinar la medición de distancia $d_{(k,m)}$ como el valor menor. El módulo de diseño de casete 324 puede construir luego la matriz de vía **P** y resolver el problema de programación lineal de enteros en la ecuación (27) para determinar la secuencia del casete.

[00669] En comparación con el enfoque de muestreo aleatorio, la resolución de la secuencia del casete con el problema de programación de enteros requiere la determinación de las mediciones de distancia *v* x (*v-1*), que cada una corresponde a un par de epítopos terapéuticos en la vacuna. Una secuencia de casete determinada mediante este enfoque puede producir una secuencia con una presentación de epítopos de unión significativamente menor mientras que posiblemente requiera significativamente menos recursos informáticos que el enfoque de muestreo aleatorio, particularmente

cuando la cantidad de secuencias del casete candidatas generadas es grande.

XI.B.2. Comparación de la presentación de epítopos de unión para secuencias de casete generadas mediante muestreo aleatorio en oposición a TSP asimétrico

[00670] Se generaron dos secuencias de casete que incluían v=20 de epítopos terapéuticos mediante muestreo aleatorio de 1.000.000 permutaciones (secuencia de casete $\mathbf{C_1}$) y mediante resolución del problema de programación lineal de enteros en la ecuación (27) (secuencia de casete $\mathbf{C_2}$). Las mediciones de distancia, y por tanto la puntuación de presentación, se determinó según el modelo de presentación descrito en la ecuación (14), en el cual f es la función sigmoide, $\boldsymbol{x_h^i}$ es la secuencia de péptidos $\boldsymbol{p^i}$, $\mathbf{g_h}(\cdot)$ es la función de red neuronal, **w** incluye la secuencia flanqueante, las transcripciones logarítmicas por millón de kilobases (TPM, por su sigla en inglés) del péptido $\boldsymbol{p^i}$, la antigenicidad de la proteína del péptido $\boldsymbol{p^i}$ y la muestra de identificación de origen del péptido $\boldsymbol{p^i}$ y $\mathbf{g_w}(\cdot)$ de la secuencia flanqueante y TPM logarítmicas son respectivamente funciones de redes neuronales. Cada una de las funciones de redes neuronales para $\mathbf{g_h}(\cdot)$ incluyó un nodo de salida de un perceptrón de múltiples capas (MLP, por su sigla en inglés) con una capa escondida con dimensiones de entrada de datos 231 (11 residuos x 21 caracteres por residuo, inclusive caracteres del panel), ancho 256, activaciones de unidad lineal rectificada (ReLU, por su sigla en inglés) en la capa escondida, activaciones lineales en la capa de salida y un nodo de salida por alelo de HLA en el conjunto de datos de entrenamiento. La función de red neuronal para la secuencia flanqueante fue un MLP con una capa escondida con una dimensión de entrada de datos de 210 (5 residuos de la secuencia flanqueante de extremo N + 5 residuos de la secuencia flanqueante de extremo C x 21 caracteres por residuo, inclusive los caracteres del panel), ancho 32, activaciones de ReLU en la capa escondida y activación lineal en la capa de salida. La función de red neuronal para la TPM de logaritmo de ARN fue un MLP con una capa escondida con una dimensión de entrada de datos de 1, ancho 16, activaciones ReLU en la capa escondida y activación lineal en la capa de salida. Los modelos de presentación se construyeron para alelos de HLA HLA-A*02:04, HLA-A*02:07, HLA-B*40:01, HLA-B*40:02, HLA-C*16:02 y HLA-C*16:04. Se comparó la puntuación de presentación que indica la cantidad esperada de epítopos de unión presentados de las dos secuencias de casetes. Los resultados demostraron que la puntuación de presentación para la secuencia del casete generada por la resolución de la ecuación de (27) se asoció a una mejora de ~4 veces respecto a la puntuación de presentación para la secuencia del casete generada por el muestreo aleatorio.

[00671] Específicamente, los v=20 epítopos se proporcionaron mediante:

$\boldsymbol{p^{*1}}$ = YNYSYWISIFAHTMWYNIWHVQWNK (SEQ ID NO: 90)

$\boldsymbol{p^{*2}}$ = IEALPYVFLQDQFELRLLKGEQGNN (SEQ ID NO: 91)

$\boldsymbol{p^{*3}}$ = DSEETNTNYLHYCHFHWTWAQQTTV (SEQ ID NO: 92)

p'^{4} = GMLSQYELKDCSLGFSWNDPAKYLR (SEQ ID NO: 93)

p'^{5} = VRIDKFLMYVWYSAPFSAYPLYQDA (SEQ ID NO: 94)

p'^{6} = CVHIYNNYPRMLGIPFSVMVSGFAM (SEQ ID NO: 95)

p'^{7} = FTFKGNIWIEMAGQFERTWNYPLSL (SEQ ID NO: 96)

p'^{8} = ANDDTPDFRKCYIEDHSFRFSQTMN (SEQ ID NO: 97)

p'^{9} = AAQYIACMVNRQMTIVYHLTRWGMK (SEQ ID NO: 98)

p'^{10} = KYLKEFTQLLTFVDCYMWITFCGPD (SEQ ID NO: 99)

p'^{11} = AMHYRTDIHGYWIEYRQVDNQMWNT (SEQ ID NO: 100)

p'^{12} = THVNEHQLEAVYRFHQVHCRFPYEN (SEQ ID NO: 101)

p'^{13} = QTFSECLFFHCLKVWNNVKYAKSLK (SEQ ID NO: 102)

p'^{14} = SFSSWHYKESHIALLMSPKKNHNNT (SEQ ID NO: 103)

p'^{15} = ILDGIMSRWEKVCTRQTRYSYCQCA (SEQ ID NO: 104)

p'^{16} = YRAAQMSKWPNKYFDFPEFMAYMPI (SEQ ID NO: 105)

p'^{17} = PRPGMPCQHHNTHGLNDRQAFDDFV (SEQ ID NO: 106)

p'^{18} = HNIISDETEVWEQAPHITWVYMWCR (SEQ ID NO: 107)

p'^{19} = AYSWPVVPMKWIPYRALCANHPPGT (SEQ ID NO: 108)

p'^{20} = HVMPHVAMNICNWYEFLYRISHIGR(SEQ ID NO: 109).

En el primer ejemplo, se generaron en forma aleatoria 1.000.000 secuencias de casete candidatas diferentes con los 20 epítopos terapéuticos. La puntuación de presentación se generó para cada una de las secuencias de casete candidatas. La secuencia de casete candidata identificada como aquella con la puntuación de presentación menor fue:

$\mathbf{C_1}$ =

THVNEHQLEAVYRFHQVHCRFPYENAMHYQMWNTYRAAQMSKWPNKYFDFPEFMAYMPICVHIYNN
YPRMLGIPFSVMVSGFAMAYSWPVVPMKWIPYRALCANHPPGTANDDTPDFRKCYIEDHSFRFSQTM
NIEALPYVFLQDQFELRLLKGEQGNNDSEETNTNYLHYCHFHWTWAQQTTVILDGIMSRWEKVCTRQT
RYSYCQCAFTFKGNIWIEMAGQFERTWNYPLSLSFSSWHYKESHIALLMSPKKNHNNTQTFSECLFFH
CLKVWNNVKYAKSLKHVMPHVAMNICNWYEFLYRISHIGRHNIISDETEVWEQAPHITWVYMWCRVRI
DKFLMYVWYSAPFSAYPLYQDAKYLKEFTQLLTFVDCYMWITFCGPDAAQYIACMVNRQMTIVYHLTR
WGMKYNYSYWISIFAHTMWYNIWHVQWNKGMLSQYELKDCSLGFSWNDPAKYLRPRPGMPCQHHN
THGLNDRQAFDDFV (SEQ ID NO: 111)

con una puntuación de presentación de 6,1 de la cantidad esperada de epítopos de unión presentados. La puntuación de presentación de mediana de las 1.000.000 secuencias aleatorias fue 18,3. El

experimento muestra que la cantidad esperada de epítopos de unión presentados se puede reducir significativamente mediante la identificación de una secuencia de casete entre los casetes muestreados en forma aleatoria.

[00672] En el segundo ejemplo, se identificó una secuencia de casete **C_2** mediante la resolución del problema de programación lineal de enteros en la ecuación (27). Específicamente, se determinó la medición de distancia de cada unión posible entre un par de epítopos terapéuticos. Se usaron las mediciones de distancia para resolver la solución del problema de programación de enteros. La secuencia de casete identificada mediante este enfoque fue:

C_2 =

IEALPYVFLQDQFELRLLKGEQGNNILDGIMSRWEKVCTRQTRYSYCQCAHVMPHVAMNICNWYEFLY
RISHIGRTHVNEHQLEAVYRFHQVHCRFPYENFTFKGNIWIEMAGQFERTWNYPLSLAMHYQMWNTS
FSSWHYKESHIALLMSPKKNHNNTVRIDKFLMYVWYSAPFSAYPLYQDAQTFSECLFFHCLKVWNNVK
YAKSLKYRAAQMSKWPNKYFDFPEFMAYMPIAYSWPVVPMKWIPYRALCANHPPGTCVHIYNNYPRM
LGIPFSVMVSGFAMHNIISDETEVWEQAPHITWVYMWCRAAQYIACMVNRQMTIVYHLTRWGMKYNY
SYWISIFAHTMWYNIWHVQWNKGMLSQYELKDCSLGFSWNDPAKYLRKYLKEFTQLLTFVDCYMWIT
FCGPDANDDTPDFRKCYIEDHSFRFSQTMNDSEETNTNYLHYCHFHWTWAQQTTVPRPGMPCQHHN
THGLNDRQAFDDFV

con una puntuación de presentación de 1,7. La puntuación de presentación de la secuencia de casete **C_2** mostró una mejora de ~4 veces respecto a la puntuación de presentación de la secuencia de casete **C_1** y una mejora de ~11 veces respecto a la puntuación de presentación de mediana de los 1.000.000 casetes candidatos generados en forma aleatoria. El tiempo de ejecución para el casete generado **C_1** fue de 20 segundos en un único hilo de una CPU Intel Xeon E5-2650 de 2,30 GHz. El tiempo de ejecución para el casete generado **C_2** fue de 1 segundo en un único hilo de la misma CPU. Por tanto, en el presente ejemplo, la secuencia de casete identificada mediante la resolución del problema de programación de enteros de la ecuación (27) produce una solución ~4 veces mejor con un costo informático 20 veces menor.

[00673] Los resultados muestran que el problema de programación de enteros posiblemente puede proporcionar una secuencia de casete con una cantidad menor de epítopos de unión presentados que una identificada a partir de muestreo aleatorio, posiblemente con menos recursos informáticos.

XI.B.3. Comparación de la presentación de epítopos de unión para la selección de secuencias de casete generadas mediante MHCflurry y el modelo de presentación

[00674] En el presente ejemplo, se generaron secuencias de casete que incluyen *v*=20 epítopos terapéuticos se seleccionaron según la secuenciación de exoma tumoral/normal, la secuenciación de

transcriptomas tumorales y tipificación de HLA de una muestra de cáncer de pulmón mediante el muestreo aleatorio de 1.000.000 permutaciones y mediante la resolución del problema de programación lineal de enteros en la ecuación (27). Se determinaron las mediciones de distancia, y por tanto la puntuación de presentación, según la cantidad de epítopos de unión predicha por MHCflurry, un predictor de afinidad de unión de HLA y péptidos, para unir los HLA del paciente con una afinidad menor que una diversidad de umbrales (por ejemplo, 50-1000nM, o mayor, o menor). En este ejemplo, se seleccionaron las 20 mutaciones somáticas no sinónimas seleccionadas como epítopos terapéuticos de entre 98 mutaciones somáticas identificadas en la muestra tumoral mediante clasificación de las mutaciones de acuerdo con el modelo de presentación en la Sección XI.B que antecede. Sin embargo, se aprecia que en otras modalidades, los epítopos terapéuticos se pueden selecciona según otros criterios; tal como aquellos basados en la estabilidad, o combinaciones de criterios tales como puntuación de presentación, afinidad y otros. Además, se aprecia que los criterios usados para priorizar epítopos terapéuticos para su inclusión en la vacuna no tienen por qué ser los mismos criterios usados para la determinación de la medición de distancia D(*k, m*) usada en el módulo de diseño de casete 324.

[00675] Los alelos de HLA clase I del paciente fueron HLA-A*01:01, HLA-A*03:01, HLA-B*07:02, HLA-B*35:03, HLA-C*07:02, HLA-C*14:02.

[00676] Específicamente en este ejemplo, los v=20 epítopos terapéuticos fueron

SSTPYLYYGTSSVSYQFPMVPGGDR (SEQ ID NO: 112)
EMAGKIDLLRDSYIFQLFWREAAEP (SEQ ID NO: 113)
ALKQRTWQALAHKYNSQPSVSLRDF (SEQ ID NO: 114)
VSSHSSQATKDSAVGLKYSASTPVR (SEQ ID NO: 115)
KEAIDAWAPYLPEYIDHVISPGVTS (SEQ ID NO: 116)
SPVITAPPSSPVFDTSDIRKEPMNI (SEQ ID NO: 117)
PAEVAEQYSEKLVYMPHTFFIGDHA (SEQ ID NO: 118)
MADLDKLNIHSIIQRLLEVRGS (SEQ ID NO: 119)
AAAYNEKSGRITLLSLLFQKVFAQI (SEQ ID NO: 120)
KIEEVRDAMENEIRTQLRRQAAAHT (SEQ ID NO: 121)
DRGHYVLCDFGSTTNKFQNPQTEGV (SEQ ID NO: 122)
QVDNRKAEAEEAIKRLSYISQKVSD (SEQ ID NO: 123)
CLSDAGVRKMTAAVRVMKRGLENLT (SEQ ID NO: 124)
LPPRSLPSDPFSQVPASPQSQSSSQ (SEQ ID NO: 125)
ELVLEDLQDGDVKMGGSFRGAFSNS (SEQ ID NO: 126)
VTMDGVREEDLASFSLRKRWESEPH (SEQ ID NO: 127)

IVGVMFFERAFDEGADAIYDHINEG (SEQ ID NO: 128)

TVTPTPTPTGTQSPTPTPITTTTTV (SEQ ID NO: 129)

QEEMPPRPCGGHTSSSLPKSHLEPS(SEQ ID NO: 130)

PNIQAVLLPKKTDSHHKAKGK (SEQ ID NO: 131)

[00677] Los resultados de este ejemplo en la tabla a continuación comparan la cantidad de epítopos de unión predicha por MHCflurry para unir los HLA del paciente con una afinidad menor que el valor en la columna de umbral (donde nM significa nanoMolar) tal como se halló mediante tres métodos ilustrativos. Para el primer método, el casete óptimo hallado mediante la formulación del problema del viajante (ATSP) descrita anteriormente fue con 1s de tiempo de ejecución. Para el segundo método, el casete óptimo tal como se determinó mediante la adopción del mejor casete se halló luego de1 millón de muestras aleatorias. Para el tercer método, la cantidad mediana de epítopos de unión se halló en 1 millón de muestras aleatorias.

Umbral (nM)	ATSP # Unión Epítopos de unión	Muestreo aleatorio # Epítopos de unión con unión	Mediana # Epítopos de unión con unión
50	0	0	3
100	0	0	7
150	0	1	12
500	15	26	55
1000	68	91	131

[00678] Los resultados de este ejemplo ilustran que se puede utilizar cualquiera de una cantidad de criterios para identificar si un diseño de casete determinado cumple o no los requisitos de diseño. Específicamente, tal como se demostró con ejemplos previos, la secuencia de casete seleccionada de muchas candidatas puede especificarse mediante la secuencia de casete que tiene una puntuación mínima de presentación de epítopos de unión o al menos tal puntuación es menor que un umbral identificado. Este ejemplo representa que se pueden utilizar otros criterios, tal como la afinidad de unión, para especificar si un diseño de casete determinado cumple o no los requisitos de diseño. Para este criterio, se puede establecer un umbral de afinidad de unión (por ejemplo, 50-1000, o mayor o menor) que especifique que la secuencia de diseño del casete debería tener menos que alguna cantidad umbral de epítopos de unión que supere el umbral (por ejemplo, 0) y se puede utilizar cualesquiera de una cantidad de métodos (por ejemplo, los métodos uno a tres ilustrados en la tabla) para identificar si una secuencia de casete candidata determinada cumple estos requisitos. Estos métodos ilustrativos ilustran

además que, según el método utilizado, es posible que los umbrales deban establecerse en forma diferente. Es posible contemplar otros criterios, tal como aquellos que se basan en la estabilidad, o combinaciones de criterios, tal como la puntuación de presentación, la afinidad y otros.

[00679] En otro ejemplo, se generaron los mismos casetes con el mismo tipo de HLA y 20 epítopos terapéuticos de etapas anteriores en esta sección (XI.C), pero en lugar de usar mediciones de distancias basadas en la predicción de afinidad de unión, la medición de distancia para los epítopos m, k fue la cantidad de péptidos que abarcan la unión de m a k que se predice que se van a presentar mediante los alelos de HLA clase I del paciente con probabilidad de presentación superior a una serie de umbrales (entre una probabilidad de 0,005 y 0,5, o más, o menos), donde las probabilidad de presentación se determinaron mediante el modelo de presentación en la Sección XI.B que antecede. El presente ejemplo ilustra además la amplitud de criterios que se pueden considerar al identificar si una secuencia de casete candidata cumple con los requisitos de diseño para su uso en la vacuna.

Umbral (probabilidad)	ATSP # Epítopos de unión	Muestreo aleatorio # Epítopos de unión	Mediana # Epítopos de unión
0,005	58	79	118
0,01	39	59	93
0,05	7	33	47
0,1	5	14	35
0,2	1	8	25
0,5	0	2	14

[00680] Los ejemplos que anteceden identificaron que los criterios para la determinación de si una secuencia de casete candidata puede variar según la implementación. Cada uno de estos ejemplos ilustró que el conteo de la cantidad de epítopos de unión que supera o es insuficiente respecto al criterio puede ser un conteo utilizado en la determinación de si la secuencia de casete candidata cumple ese criterio. Por ejemplo, si el criterio es una cantidad de epítopos que cumple o supera un umbral de afinidad de unión por HLA, que la secuencia de casete candidata tenga más o menos que esa cantidad puede determinar si la secuencia de casete candidata cumple con los criterios para su uso como el casete seleccionado para la vacuna. De manera similar si el criterio es la cantidad de epítopos de unión que supera un umbral de probabilidad de presentación.

[00681] Sin embargo, en otras modalidades, los cálculos que no sean de conteos se pueden llevar a cabo para determinar si una secuencia de casete candidata cumple con los criterios de diseño. Por

ejemplo, en lugar del conteo de epítopos que supera/es escaso respecto a algún umbral, se puede determinar en su lugar qué proporción de epítopos de unión supera o es escasa respecto al umbral, por ejemplo, si el X % superior de epítopos de unión tiene una probabilidad de presentación que supera cierto umbral Y, o si X % de epítopos de unión tiene una afinidad de unión a HLA menor o mayor que Z nM. Estos son meramente ejemplos, en general los criterios pueden estar basados en cualquier atributo de epítopos de unión individuales o estadística derivada de las acumulaciones de algunos o todos los epítopos de unión. Aquí, X puede ser en general cualquier número entre 0 y 100 % (por ejemplo, 75 % o menos) e Y puede ser cualquier valor entre 0 y 1, y Z puede ser cualquier número adecuado para el criterio en cuestión. Estos valores se pueden determinar en forma empírica y depende de los modelos y criterios usados, así como de la calidad de los datos de entrenamiento usados.

[00682] Como tal, en determinados aspectos, los epítopos de unión con mayores probabilidades de presentación se pueden eliminar; los epítopos de unión con pocas probabilidades de presentación pueden conservarse; los epítopos de unión que se unen potentemente, es decir, los epítopos de unión con una afinidad de unión menor que 1000nM o 500nM o algún otro umbral, pueden eliminarse; y/o los epítopos con unión frágil, es decir, epítopos de unión con una afinidad de unión que supera los 1000nM o 500nM o algún otro umbral, pueden conservarse.

[00683] Si bien los ejemplos que anteceden identificaron secuencias candidatas con el uso de una implementación de un modelo de presentación descrito anteriormente, estos principios aplican igualmente a una implementación en la cual los epítopos para la disposición en las secuencias de casete se identifiquen también según otros tipos de modelos, tal como aquellos basados en la afinidad, la estabilidad y otros.

XII. Ejemplo 7: Resultados de experimentación que muestran el rendimiento del modelo de presentación ilustrativo

[00684] La validez de los diversos modelos de presentación descritos anteriormente fue evaluada según los datos de prueba *T* que fueron subconjuntos de datos de entrenamiento 170 que no se usaron para el entrenamiento de los modelos de presentación o un conjunto de datos aparte de los datos de entrenamiento 170 que tiene variables y estructuras de datos similares a los datos de entrenamiento 170.

[00685] Una medición pertinente que indica el rendimiento de los modelos de presentación es:

$$\text{Valor predictivo positivo (PPV)} = P(y_{i\in T} = 1 \mid u_{i\in T} \geq t) = \frac{\sum_{i\in T} \mathbb{1}(y_i = 1,\, u_i \geq t)}{\sum_{i\in T} \mathbb{1}(u_i \geq t)}$$

que indica la relación de la cantidad de ejemplos de péptidos que se predijo correctamente que se presentarían en los alelos de HLA asociados respecto a la cantidad de ejemplos de péptidos que se predijo que se presentarían en los alelos de HLA. En una implementación, se predijo que un péptido $\boldsymbol{p^i}$

en los datos de prueba *T* se presentaría en uno o más alelos de HLA asociados si la estimación de probabilidad correspondiente u_i es mayor o igual que un valor de umbral determinado *t.* Otra medición pertinente que indica el rendimiento de los modelos de presentación es:

$$\text{Memoria} = P(u_{i\in T} \geq t |\, y_{i\in T} = 1) = \frac{\sum_{i\in T} \mathbb{1}(y_i = 1,\, u_i \geq t)}{\sum_{i\in T} \mathbb{1}(y_i = 1)}$$

que indica la relación de la cantidad de ejemplos de péptidos que se predijo correctamente que se presentarían en los alelos de HLA asociados respecto a la cantidad conocida de ejemplos de péptidos que se presentarían en los alelos de HLA. Otra medición pertinente que indica el rendimiento de los modelos de presentación es el área bajo la curva (AUC, por su sigla en inglés) de la característica operativa del receptor (ROC, por su sigla en inglés). La ROC grafica la memoria en función de la tasa de falsos positivos (FPR, por su sigla en inglés), que se proporciona mediante:

$$\text{FPR} = P(u_{i\in T} \geq t |\, y_{i\in T} = 0) = \frac{\sum_{i\in T} \mathbb{1}(y_i = 0,\, u_i \geq t)}{\sum_{i\in T} \mathbb{1}(y_i = 0)}.$$

XII.A. Comparación del rendimiento del modelo de presentación en datos de espectrometría de masas en oposición al modelo de vanguardia

[00686] La Figura 13A compara los resultados de rendimiento de un modelo de presentación ilustrativo, tal como se presenta en la presente, y los modelos de vanguardia para la predicción de la presentación de péptidos en datos de espectrometría de masa para múltiples alelos. Los resultados demuestran que el modelo de presentación ilustrativo tuvo un rendimiento significativamente mejor en la predicción de presentación de péptidos que los modelos de vanguardia basados en predicciones de afinidad y la estabilidad.

[00687] Específicamente, el modelo de presentación ilustrativo que se muestra en la FIG. 13A como «MS» fue el máximo del modelo de presentación por alelos que se muestra en la ecuación (12), con el uso de la función de dependencia afín $g_h(\cdot)$ y la función expit $f(\cdot)$. Se entrenó el modelo de presentación ilustrativo basándose en un subconjunto de los datos de espectrometría de masas del alelo individual HLA-A*02:01 del conjunto de datos IEDB (conjunto de datos «D1») (pueden encontrarse los datos en http://www.iedb.org/doc/mhc_ligand_full.zip) y un subconjunto de espectrometría de masas del alelo individual HLA-B*07:02 del conjunto de datos IEDB (conjunto de datos «D2») (pueden encontrarse los datos en http://www.iedb.org/doc/mhc_ligand_full.zip). Todos los péptidos de la proteína fuente que contienen los péptidos presentados en el conjunto de prueba fueron eliminados de los datos de entrenamiento de modo que el modelo de presentación ilustrativo no pudiera simplemente memorizar las secuencias de los antígenos presentados.

[00688] El modelo que se muestra en la Figura 13A como «Afinidad» fue un modelo similar al modelo actual de vanguardia que predice la presentación de péptidos según las predicciones de afinidad

NETMHCpan. Se proporciona la implementación de NETMHCpan con mayor detalle en http://www.cbs.dtu.dk/services/NetMHCpan/. El modelo que se muestra en la Figura 13A como «Estabilidad» fue un modelo similar al modelo actual de vanguardia que predice la presentación de péptidos según las predicciones de estabilidad NETMHCstab. La implementación de NETMHCstab se proporciona con mayor detalle en http://www.cbs.dtu.dk/services/NetMHCstab-1.0/. Los datos de prueba que son un subconjunto de los datos de espectrometría de masas de la línea celular JY de múltiples alelos HLA-A*02:01 y HLA-B*07:02 del conjunto de datos de Bassani-Sternberg (conjunto de datos «D3») (datos que se pueden encontrar en www.ebi.ac.uk/pride/archive/projects/PXD000394). Las barras de error (tal como se indican en líneas continuas) muestran 95 % de intervalos de confianza.

[00689] Tal como se muestra en los resultados de la Figura 13A, el modelo de presentación ilustrativo entrenado con los datos de la espectrometría de masas tuvo un valor PPV significativamente elevado a 10 % de tasa de memoria con respecto a los modelos de vanguardia que predijeron una presentación de péptidos basada en predicciones de afinidad de unión al MHC o predicciones de estabilidad de unión al MHC. Específicamente, el modelo de presentación ilustrativo tuvo aproximadamente un PPV 14 % mayor que el modelo basado en las predicciones de afinidad y tuvo aproximadamente un PPV 12 % mayor que el modelo basado en las predicciones de estabilidad.

[00690] Estos resultados demuestran que el modelo de presentación ilustrativo tuvo un rendimiento significativamente mejor que los modelos de vanguardia que predicen la presentación de péptidos basándose en las predicciones de afinidad de unión al MHC o de estabilidad de unión al MHC, aunque el modelo de presentación ilustrativo no estaba entrenado basándose en las secuencias de proteínas que contenían los péptidos presentados.

XII.B. Comparación del rendimiento del modelo de presentación en datos de epítopos de linfocitos T en oposición a los modelos de vanguardia

[00691] La Figura 13B compara los resultados de rendimiento de otro modelo de presentación ilustrativo, tal como se presenta en la presente, y modelos de vanguardia para la predicción de presentación de péptidos en los datos de epítopos de linfocitos T. Los datos de epítopos de linfocitos T contienen secuencias de péptidos que se presentaron mediante alelos de MHC en la superficie celular y fueron reconocidos por los linfocitos T. Los resultados demostraron que si bien el modelo de presentación ilustrativo se entrena basándose en los datos de espectrometría de masas, el modelo de presentación ilustrativo tuvo un rendimiento significativamente mejor en la predicción de epítopos de linfocitos T que los modelos de vanguardia basados en predicciones de afinidad y estabilidad. En otras palabras, los resultados de la Figura 13B indicaron no solamente que el modelo de presentación ilustrativo tuvo mejor rendimiento que los modelos de vanguardia en la predicción de presentación de péptidos en los datos de prueba de espectrometría de masas, sino que el modelo de presentación

ilustrativo también tuvo un rendimiento significativamente mejor que los modelos de vanguardia en la predicción de epítopos que realmente fueron reconocidos por los linfocitos T. Esto indica que la diversidad de modelos de presentación tal como se presentan en la presente puede mejorar la identificación de antígenos que posiblemente induzcan respuestas inmunogénicas en el sistema inmunitario.

[00692] Específicamente, el modelo de presentación ilustrativo que se muestra en la FIG. 13B como «MS» fue el modelo de presentación por alelo que se muestra en la ecuación (2), con el uso de la función de transformación afín $g_h(\cdot)$ y la función expit $f(\cdot)$ que se entrenó según un subconjunto del conjunto de datos D1. Todos los péptidos de la proteína fuente que contienen los péptidos presentados en el conjunto de prueba se eliminaron de los datos de entrenamiento de modo que el modelo de presentación no pudiera simplemente memorizar las secuencias de antígenos presentados.

[00693] Cada uno de los modelos se aplicó a los datos de prueba que son un subconjunto de datos de espectrometría de masas en los datos de epítopos de linfocitos T HLA-A*02:01 (conjunto de datos «D4») (pueden encontrarse los datos en www.iedb.org/doc/tcell full v3.zip). El modelo que se muestra en la Figura 13B como «Afinidad» fue un modelo similar al modelo actual de vanguardia que predice la presentación de péptidos basándose en las predicciones de afinidad NETMHCpan y el modelo que se muestra en la Figura 13B como «Estabilidad» fue un modelo similar al modelo actual de vanguardia que predice la presentación de péptidos según las predicciones de estabilidad NETMHCstab. Las barras de error (tal como se indican en líneas continuas) muestran 95 % de intervalos de confianza.

[00694] Tal como se muestra en los resultados de la Figura 13A, el modelo de presentación por alelos entrenado con los datos de la espectrometría de masas tuvo un valor PPV significativamente más elevado a una tasa de 10 % de memoria que los modelos de vanguardia que predicen la presentación de péptidos basándose en las predicciones de afinidad de unión al MHC o de estabilidad de unión al MHC, aunque el modelo de presentación no se entrenó basándose en las secuencias de proteínas que contenían péptidos presentados. Específicamente, el modelo de presentación por alelos tuvo aproximadamente un PPV 9% mayor que el modelo basado en las predicciones de afinidad y tuvo aproximadamente un PPV 8% mayor que el modelo basado en las predicciones de estabilidad.

[00695] Estos resultados demostraron que el modelo de presentación ilustrativo entrenado con los datos de la espectrometría de masas tuvo un rendimiento significativamente mejor que los modelos de vanguardia en la predicción de epítopos que fueron reconocidos por los linfocitos T.

XII.C. Comparación de diferentes rendimientos del modelo de presentación en los datos de espectrometría de masas

[00696] La Figura 13C compara los resultados de rendimiento para un modelo ilustrativo de función de sumas (ecuación (13)), un modelo ilustrativo de suma de funciones (ecuación (19)) y un modelo

ilustrativo de segundo orden (ecuación (23)) para la predicción de la presentación de péptidos en datos de espectrometría de masas de múltiples alelos. Los resultados demostraron que el modelo de suma de funciones y el modelo de segundo orden tuvieron un mejor rendimiento que el modelo de función de sumas. Esto se debe a que el modelo de función de sumas implica que los alelos en una configuración de múltiples alelos pueden interferir unos con otros para la presentación de péptidos, cuando en realidad, la presentación de péptidos es independiente en forma eficaz.

[00697] Específicamente, el modelo de presentación ilustrativo marcado como «sigmoide de sumas» en la Figura 13C fue el modelo de función de sumas con una función de dependencia de red $g_h(\cdot)$, la función de identidad $f(\cdot)$ y la función expit $r(\cdot)$. El modelo ilustrativo marcado como «suma de sigmoides» fue el modelo de suma de funciones en la ecuación (19) con una función de dependencia de red $g_h(\cdot)$, la función expit $f(\cdot)$ y la función de identidad $r(\cdot)$. El modelo ilustrativo marcado como «tangente hiperbólica» fue el modelo de suma de funciones en la ecuación (19) con una función de dependencia de red $g_h(\cdot)$, la función expit $f(\cdot)$ y la función de tangente hiperbólica $r(\cdot)$. El modelo ilustrativo marcado como «segundo orden» fue el modelo de segundo orden en la ecuación (23) con una forma de probabilidad de presentación implícita por alelo que se muestra en la ecuación (18) con una función de dependencia de red $g_h(\cdot)$ y la función expit $f(\cdot)$. Cada modelo se entrenó basándose en un subconjunto del conjunto de datos D1, D2 y D3. Se aplicaron los modelos de presentación ilustrativos a datos de prueba que son un subconjunto aleatorio del conjunto de datos D3 que no se superpuso con los datos de entrenamiento.

[00698] Como se muestra en la Figura 13C, la primera columna hace referencia al AUC de la ROC cuando se aplicó cada modelo de presentación al conjunto de prueba, la segunda columna hacer referencia al valor de pérdida de probabilidad logarítmica negativa y la tercera columna hace referencia al PPV a una tasa de 10 % de memoria. Como se muestra en la Figura 13C, el rendimiento de los modelos de presentación «suma de sigmoides», «tangente hiperbólica» y «segundo orden» empataron aproximadamente en aproximadamente 15-16 % de PPV a una tasa de 10 % de memoria, mientras que el rendimiento del modelo de «sigmoides de sumas» fue ligeramente menor a 11 %.

[00699] Tal como se trató anteriormente en la sección X.C.4., los resultados muestran que los modelos de presentación «suma de sigmoides», «tangente hiperbólica» y «segundo orden» tienen valores elevados de PPV en comparación con el modelo de «sigmoides de suma» ya que los modelos justificaron correctamente cómo se presentaron independientemente los péptidos con cada alelo de MHC en una configuración de múltiples alelos.

XII.D. Comparación del rendimiento del modelo de presentación con y sin entrenamiento de los datos de espectrometría de masas de alelos individuales

[00700] La Figura 13D compara los resultados de rendimiento para dos modelos de presentación ilustrativos que se entrenaron con y sin datos de espectrometría de masas de alelos individuales en la

predicción de presentación de péptidos para datos de espectrometría de masas de múltiples alelos. Los resultados indicaron que los modelos de presentación ilustrativos que se entrenan sin los datos de alelos individuales logran un rendimiento semejante al de los modelos de presentación ilustrativos entrenados con datos de alelos individuales.

[00701] El modelo ilustrativo «con datos de alelos individuales A2/B7» fue el modelo de presentación de «suma de sigmoides» en la ecuación (19) con una función de dependencia de red $g_h(\cdot)$, la función expit $f(\cdot)$ y la función de identidad $r(\cdot)$. El modelo se entrenó basándose en un subconjunto del conjunto de datos D3 y datos de espectrometría de masas de alelos individuales para una diversidad de alelos de MHC de la base de datos IEDB (se pueden encontrar datos en: http://www.iedb.org/doc/mhc_ligand_full.zip). El modelo ilustrativo «sin datos de alelos individuales A2/B7» fue el mismo modelo, pero entrenado basándose en un subconjunto del conjunto de datos D3 de múltiples alelos sin los datos de espectrometría de masas de alelos individuales para los alelos HLA-A*02:01 y HLA-B*07:02, pero con los datos de espectrometría de masas de alelos individuales para otros alelos. Dentro de los datos de entrenamiento de múltiples alelos, la línea celular HCC1937 expresó HLA-B*07:02 pero no HLA-A*02:01 y la línea celular HCT116 expresó HLA-A*02:01 pero no HLA-B*07:02. Los modelos de presentación ilustrativos se aplicaron a datos de prueba que fueron un subconjunto aleatorio del conjunto de datos D3 y no se superpusieron con los datos de entrenamiento.

[00702] Como se muestra en la Figura 13D, las predicciones basadas en las probabilidades de presentación implícita por alelo para el alelo de MHC HLA-A*02:01 tuvieron un rendimiento significativamente mejor en los datos de prueba de alelos individuales para el alelo de MHC HLA-A*02:01 que para el alelo de MHC HLA-B*07:02. Se muestran resultados similares para el alelo de MHC HLA-B*07:02.

[00703] Estos resultados indican que las probabilidades de presentación implícita por alelo del modelo de presentación pueden predecir y distinguir correctamente los motivos de unión a los alelos individuales de MHC, aunque la asociación directa entre los péptidos y cada uno de los alelos individuales de MHC no se conociera en los datos de entrenamiento.

XII.E. Comparación del rendimiento de predicción por alelos sin entrenamiento con los datos de espectrometría de masas de alelos individuales

[00704] La Figura 13E muestra el rendimiento para los modelos ilustrativos «sin datos de alelos individuales A2/B7» y «con datos de alelos individuales A2/B7» que se muestran en la Figura 13D en los datos de espectrometría de masas de alelos individuales HLA-A*02:01 y HLA-B*07:02 que se ofrecen en el análisis que se muestra en la Figura 13D. Los resultados indican que si bien el modelo de presentación ilustrativo se entrena sin los datos de espectrometría de masas de alelos individuales para estos dos alelos, el modelo puede aprender los motivos de unión para cada alelo de MHC.

[00705] La columna «Correlación» hace referencia a la correlación entre las etiquetas reales que indican si el péptido se presentó en el alelo correspondiente en los datos de prueba y la etiqueta para predicción. Como se muestra en la Figura 13E, «modelo A2 que predice B7» indica el rendimiento del modelo cuando la presentación del péptido se predice para los datos de HLA-B*07:02 de alelo único en función del estimado de probabilidad de presentación por alelo para el alelo de MHC HLA-A*02:01. De manera similar, «modelo A2 que predice A2» indica el rendimiento del modelo cuando la presentación del péptido se predice para HLA-A*02:01 de alelo único en función del estimado de probabilidad de presentación por alelo para el alelo de MHC HLA-A*02:01. El «modelo B7que predice B7» indica el rendimiento del modelo cuando la presentación del péptido se predice para los datos de HLA-B*07:02 de alelo único en función del estimado de probabilidad de presentación por alelo para el alelo de MHC HLA-B*07:02. El «modelo B7 que predice A2» indica el rendimiento del modelo cuando la presentación del péptido se predice para HLA-A*02:01 de alelo único en función del estimado de probabilidad de presentación por alelo para el alelo de MHC HLA-B*07:02.

[00706] Como se muestra en la Figura 13E, la capacidad predictiva de probabilidades implícitas por alelo para un alelo de HLA es significativamente superior para el alelo previsto y significativamente inferior para el otro alelo de HLA. De forma similar, los resultados mostrados en la Figura 13D, los modelos de presentación de ejemplo aprendidos correctamente para diferenciar la presentación de péptidos de alelos individuales HLA-A*02:01 y HLA-B*07:02, incluso a pesar de que no se encontró la asociación directa entre la presentación del péptido y estos alelos en múltiples datos de entrenamiento de múltiples alelos.

XII.F. Residuos de anclaje de aparición frecuente en motivos de anclaje canónicos conocidos de coincidencias de predicción por alelo

[00707] La Figura 13F muestra los residuos de anclaje comunes en las posiciones 2 y 9 entre los nonámeros previstos por el modelo de ejemplo de «datos de alelo único sin A2/B7» mostrado en la Figura 13D. Se previó que los péptidos se presentaron si la posibilidad estimada fue sobre 5 %. Los resultados mostraron que los residuos de anclaje más comunes en los péptidos identificados para la presentación en los alelos de MHC HLA-A*02:01 y HLA-B*07:02, coincidieron previamente con motivos de anclaje conocidos para estos alelos de MHC. Esto indica que los modelos de presentación de ejemplo aprendieron correctamente la unión de péptidos en función de aminoácidos de las secuencias de péptidos, como se esperó.

[00708] Como se muestra en la Figura 13F, los aminoácidos L/M en la posición 2 y los aminoácidos V/L en la posición 9 son conocidos por ser motivos de residuos de anclaje canónicos (como se muestra en la Tabla 4 de https://link.springer.com/article/10.1186/1745-7580-4-2) para HLA-A*02:01, y el aminoácido P en la posición 2 y los aminoácidos L/V en la posición 9 son conocidos por ser motivos de

residuos de anclaje canónicos para HLA-B*07:02. Los motivos de residuos de anclaje más comunes en las posiciones 2 y 9 para los péptidos identificados el modelo coincidió con los motivos de residuo de anclaje canónicos conocidos para ambos alelos de HLA.

XII.G. Comparación de los rendimientos del modelo de presentación con y sin variables de no interacción con alelos

[00709] La Figura 13G compara los resultados del rendimiento entre el modelo de presentación de ejemplo que incorporó las secuencias flanqueantes de extremo C y N como variables de interacción con alelos y un modelo de presentación de ejemplo que incorporó secuencias flanqueantes de extremo C y N como variables de no interacción con alelos. Los resultados mostraron que incorporar las secuencias flanqueantes de extremo C y N como variables de no interacción de alelos mejoró significativamente el rendimiento del modelo. Más específicamente, es ventajoso identificar características adecuadas para la presentación de péptidos que son comunes a través de diferentes alelos de MHC, y modelarlas de forma tal que la potencia estadística para estas variables de no interacción de alelos se comparta a través de los alelos de MHC para mejorar el rendimiento del modelo de presentación.

[00710] El ejemplo del modelo «de interacción con alelos» fue el modelo de suma de funciones que utiliza la forma de probabilidades de presentación por alelo implícitas en la ecuación (22) que incorporó las secuencias flanqueantes de extremo C y N como variables de interacción con alelos, con una función de dependencia de la red $g_h(\cdot)$ y la función expit $f(\cdot)$. El ejemplo del modelo «de no interacción con alelos» fue el modelo de suma de funciones mostrado en la ecuación (21) que incorporó las secuencias flanqueantes de extremo C y N como variables de no interacción con alelos, con una función de dependencia de la red $g_h(\cdot)$ y la función expit $f(\cdot)$. Las variables de no interacción con alelos fueron modeladas a través de una función de dependencia de la red $g_w(\cdot)$ diferente. Ambos modelos fueron entrenados en un subconjunto de datos D3 y los datos de espectrometría de masas de alelo único para una variedad de alelos de MHC a partir de la base de datos de IEDB (los datos pueden encontrarse en http://www.iedb.org/doc/mhc_ligand_full.zip). Cada uno de los modelos de presentación fue aplicado a un conjunto de datos de prueba que es un subconjunto aleatorio del conjunto de datos D3 que no se superpuso con los datos de entrenamiento.

[00711] Como se muestra en la Figura 13G, que incorpora las secuencias flanqueantes de extremo C y N en el modelo de presentación de ejemplo como variables de no interacción con alelos logró una mejora de aproximadamente 3 % en el valor de PPV con respecto a su moldeado como variables de interacción con alelos. Esto se debe a que, en general, el modelo de presentación de ejemplo «de no interacción con alelos» fue capaz de compartir la potencia estadística de las variables de no interacción con alelos a través de los alelos de MHC mediante el moldeo del efecto con una función de dependencia de la red diferente con una adición muy pequeña en la potencia computacional.

XII.H. Dependencia entre los péptidos presentados y la cuantificación de ARNm

[00712] La Figura 13H muestra la dependencia entre la la abundancia de ARNm y la frecuencia de los péptidos en una célula tumoral, según se determina mediante espectrometría de masas. Los resultados mostraron que existe una dependencia fuerte entre la expresión de ARNm y la presentación de péptidos.

[00713] Específicamente, el eje horizontal en la Figura 13H indica la expresión de ARNm en función de cuartiles de transcripción por millón (TPM, por sus siglas en inglés). El eje vertical en la figura 13H indica la fracción de epítopos presentados a partir de genes en cuartiles de expresión de ARNm correspondientes. Cada línea sólida es una gráfica relacionada con las dos mediciones a partir de una muestra tumoral que se relaciona con los datos de espectrometría de masa correspondientes y con las mediciones de expresión de ARNm. Como se muestra en la Figura 13H, existe una correlación positiva fuerte entre la expresión de ARNm y la fracción de péptidos en el gen correspondiente. Específicamente, los péptidos de genes en el cuartil superior de la expresión de ARN tienen una probabilidad de presentarse 20 veces mayor que el cuartil inferior. Además, se presentaron esencialmente 0 péptidos de genes que no se detectaron a través de ARN.

[00714] Los resultados indican que el rendimiento del modelo de presentación puede mejorar ampliamente mediante la incorporación de las mediciones de cuantificación de ARNm, ya que estas mediciones son fuertemente predictivas de la presentación de péptidos.

XII.I. Comparación del rendimiento del modelo de presentación con la incorporación de los datos de cuantificación de ARN

[00715] La Figura 13I muestra el rendimiento de dos modelos de presentación de ejemplo, uno de los cuales se entrena en función de los datos de células tumorales de la espectrometría de masas, otro de los cuales incorpora datos de cuantificación de ARNm y datos de células tumorales de la espectrometría de masas. Como se espera a partir de la Figura 13H, los resultados indican que existe una mejora significativa en el rendimiento mediante la incorporación de las mediciones de cuantificación de ARNm en el modelo de presentación de ejemplo, ya que la expresión del ARNm es un fuerte indicador de la presentación de péptidos.

[00716] «MHCflurry + filtro de ARN» fue un modelo similar al modelo del estado técnica actual que predice la presentación de péptidos en función de las predicciones de afinidad. Se implementó mediante el uso de MHCflurry junto con un filtro de expresión génica estándar que eliminó todos los péptidos de las proteínas con mediciones de cuantificación de ARNm que fueron menos de 3,2 FPKM. La implementación de MHCflurry se proporciona en detalle en https://github.com/hammerlab/mhcflurry/ y en http://biorxiv.org/content/early/2016/05/22/054775. El modelo «Modelo de ejemplo, sin ARN» fue el modelo de presentación de ejemplo «suma de sigmoides» mostrado en la ecuación (21) con la función

de dependencia de la red $g_h(\cdot)$, la función de dependencia de la red $g_w(\cdot)$, y la función expit $f(\cdot)$. El modelo «Modelo de ejemplo, sin ARN» incorporó las secuencias flanqueantes del extremo C como variables de no interacción con alelos a través de la función de dependencia de la red $g_w(\cdot)$.

[00717] El modelo «Modelo de ejemplo, con ARN» fue el modelo de presentación «suma de sigmoides» mostrado en la ecuación (19) con la función de dependencia de la red $g_h(\cdot)$, la función de dependencia de la red $g_w(\cdot)$ en la ecuación (10) que incorpora los datos de cuantificación de ARNm a través de una función log y la función expit $f(\cdot)$. El modelo «Modelo de ejemplo, con ARN» incorporó las secuencias flanqueantes de extremo C como variables de no interacción con alelos a través de las funciones de dependencia de la red $g_w(\cdot)$ e incorporó las mediciones de cuantificación de ARNm a través de la función log.

[00718] Cada modelo fue entrenado en una combinación de datos de espectrometría de masas con alelos únicos del conjunto de datos de IEDB, 7 líneas celulares de los datos de espectrometría de masas con múltiples alelos del conjunto de datos Bassani-Sternberg y 20 muestras tumorales de espectrometría de masas. Cada modelo se aplicó a un conjunto de prueba que incluye 5000 proteínas retenidas de las 7 muestras tumorales que constituyeron 9830 péptidos presentados de un total de 52.156.840 péptidos.

[00719] Como se muestra en las primeras dos gráficas de barras de la Figura 13I, el modelo «Ejemplo de modelo, sin ARN» tiene un valor de PPV a 20 % de memoria de 21 % mientras que el modelo del estado de la técnica es aproximadamente 3 %. Esto indica una mejora del rendimiento inicial de 18 % en el valor de PPV, incluso cuando la incorporación de las mediciones de cuantificación de ARNm. Como se muestra en la tercera gráfica de barras de la Figura 13I, el modelo «Modelo de ejemplo, con ARN» que incorpora los datos de cuantificación de ARNm en el modelo de presentación muestra un valor de PPV de aproximadamente 30 % que es casi un aumento del 10 % en el rendimiento en comparación con el modelo de presentación de ejemplo sin mediciones de cuantificación de ARNm.

[00720] Por lo tanto, los resultados indican que como se espera los hallazgos en la Figura 13H, la expresión de ARNm es de hecho un predictor fuerte de la predicción de péptidos que permite una mejora significativa en el rendimiento de un modelo de presentación con muy poca adición de complejidad computacional.

XII.J. Ejemplo de parámetros determinados para el alelo de MHC HLA-C*16:04

[00721] La Figura 13J compara la probabilidad de la presentación de péptidos para diferentes longitudes de péptidos entre los resultados generados por el modelo de presentación «Modelo de ejemplo, con ARN» descrito en referencia a la Figura 13I y resultados predichos mediante modelos de la técnica previa que no representan la longitud del péptido cuando se predice la presentación del péptido. Los resultados indican que el modelo de presentación de ejemplo «Modelo de ejemplo, con ARN» de la Figura 13I capturó una variación en las probabilidades a través de los péptidos de diferentes

longitudes.

[00722] El eje horizontal denotó muestras de péptidos con longitudes 8, 9, 10 y 11. El eje vertical denotó la probabilidad de la presentación de péptidos condicionada por las longitudes del péptido. La gráfica «Verdad (detos de prueba ciegos)» mostraron la proporción de péptidos presentados de acuerdo con la longitud del péptido en una muestra del conjunto de prueba de muestra. La probabilidad de presentación varió con la longitud del péptido. Por ejemplo, como se muestra en la Figura 13J, un péptido 10mer con motivos de anclaje HLA-A2 L/V canónicos fue aproximadamente 3 veces menos probable de presentarse que un 9mer con los mismos residuos de anclaje. La gráfica «Modelos que ignoran longitud» indicó mediciones previstas si los modelos del estado de la técnica que ignoran la longitud del péptido se aplicaran al mismo conjunto de datos de prueba para la predicción de la presentación. Estos modelos pueden ser versiones NetMHC antes de la versión 4.0, versiones NetMHCpan antes de la versión 3.0 y MHCflurry, que no toman en cuenta variaciones en la presentación de péptidos de acuerdo con la longitud del péptido. Como se muestra en la Figura 13J, la proporción de los péptidos presentados será constante a través de diferentes valores de la longitud de péptidos, que indica que estos modelos no podrían capturar la variación en la presentación de péptidos de acuerdo con la longitud. La gráfica «Modelo ilustrativo, con ARN» indica mediciones generadas a partir del modelo de presentación «Modelo ilustrativo, con ARN». Como se muestra en la Figura 13J, las mediciones generadas por el modelo «Modelo ilustrativo, con ARN» sigue estrechamente aquellas mostradas en la «Verdad (datos de prueba ciegos)» y representaron correctamente diferentes grados de presentación de péptidos para las longitudes 8, 9, 10 y 11.

[00723] Por lo tanto, los resultados muestran que los modelos de presentación de ejemplo como se presentan en la presente generaron predicciones mejoradas no solamente para péptidos 9mer, sino también para péptidos de otras longitudes entre 8-15, mientras que representa hasta 40 % de los péptidos presentados en alelos de clase I de HLA.

XII.K. Ejemplo de parámetros determinados para el alelo de MHC HLA-C*16:04

[00724] Lo que sigue a continuación muestra un conjunto de parámetros determinados para una variación del modelo de presentación por alelo (ecuación 2)) para el alelo de MHC HLA-C*16:04 observados por *h*:

$$u_k = \text{expit}\left(\text{relu}\left(\boldsymbol{x_h^k} \cdot \boldsymbol{W_h^1} + \boldsymbol{b_h^1}\right) \cdot \boldsymbol{W_h^2} + b_h^2\right),$$

donde relu(·) es la función de unidad lineal rectificada (RELU) y $\boldsymbol{W_h^1}$, $\boldsymbol{b_h^1}$, $\boldsymbol{W_h^2}$, y b_h^2 son el conjunto de parámetros $\boldsymbol{\theta}$ determinado para el modelo. Las variables de interacción con alelos $\boldsymbol{x_h^k}$ consiste en secuencias de péptidos. Las dimensiones de $\boldsymbol{W_h^1}$ son (231 x 256), las dimensiones de $\boldsymbol{b_h^1}$ (1 x 256), las dimensiones de $\boldsymbol{W_h^2}$ son (256 x 1), y $\boldsymbol{b_h^2}$ es un escalar. Para fines demostrativos, los valores para $\boldsymbol{b_h^1}$,

b_h^2, $\boldsymbol{W_h^1}$, y $\boldsymbol{W_h^2}$ se describen en detalle en la publicación de PCT WO2017106638, incorporada en la presente mediante referencia para todas las instrucciones en esta.

XII.L. MHC II Ejemplo 1

[00725] Se describen métodos para determinar neoantígenos del MHC clase II en mayor detalle en la solicitud internacional PCT/US2018/028438, que se incorpora a la presente mediante esta referencia por toda su descripción.

[00726] La FIG. 13K es un histograma de longitudes de péptidos eluidos de alelos del MHC clase II en células tumorales humanas y linfocitos infiltrantes de tumores (TIL) mediante el uso de espectrometría de masas. Específicamente, se llevó a cabo una peptidómica de espectrometría de masas sobre los alelos homocigotos HLA-DRB1*12:01 (“Conjunto de datos 1”) y las muestras multialélicas HLA-DRB1*12:01, HLA-DRB1*10:01 (“Conjunto de datos 2”). Los resultados muestran que las longitudes de péptidos eluidos de alelos del MHC clase II se encuentran dentro del intervalo de 6-30 aminoácidos. La distribución de frecuencia que se muestra en la FIG. 13K es similar a la de las longitudes de péptidos eluidos de alelos del MHC clase II con técnicas de espectrometría de masas del estado de la técnica, como se muestra en la FIG. 1C de la referencia 91.

[00727] La FIG. 13L ilustra la dependencia entre la cuantificación de ARNm y los péptidos presentados por residuo para el Conjunto de datos 1 y el Conjunto de datos 2. Los resultados mostraron que existe una dependencia fuerte entre la expresión de ARNm y la presentación de péptidos para los alelos del MHC clase II.

[00728] Específicamente, el eje horizontal en la FIG. 13B indica la expresión de ARNm en función de transcripciones $\log_{10}$ por millón (TPM) de cuartiles. El eje vertical en la FIG. 13L indica la presentación de péptidos por residuo como múltiplo de la del intervalo más bajo correspondiente a una expresión de ARNm entre $10^{-2} < \log_{10}\text{TPM} < 10^{-1}$. Una línea continua es una gráfica que relaciona la cuantificación de ARNm y la presentación de péptidos para el Conjunto de datos 1 y otra es para el Conjunto de datos 2. Tal como se muestra en la FIG. 13L, existe una correlación positiva fuerte entre la expresión de ARNm y la presentación de péptidos por residuo en el gen correspondiente. Específicamente, los péptidos de genes en el intervalo de $10^1 < \log_{10}\text{TPM} < 10^2$ de la expresión de ARN tienen una probabilidad de presentarse 5 veces mayor que el intervalo inferior.

[00729] Los resultados indican que el rendimiento del modelo de presentación puede mejorar ampliamente mediante la incorporación de las mediciones de cuantificación de ARNm, ya que estas mediciones son fuertemente predictivas de la presentación de péptidos.

[00730] La FIG. 13M compara los resultados de rendimiento, por ejemplo, modelos de presentación entrenados y evaluados mediante el uso del Conjunto de datos 1 y el Conjunto de datos 2. Para cada conjunto de características del modelo de los ejemplos de modelos de presentación, la FIG. 13M

representa un valor de PPV con una memoria de 10 % cuando las características del conjunto de características del modelo se clasifican como características de interacción de alelos y, de manera alternativa, cuando las características del conjunto de características del modelo se clasifican como variables de características de no interacción de alelos. Como se observa en la FIG. 13M, para cada conjunto de características del modelo de los ejemplos de modelos de presentación, se muestra en el lado izquierdo un valor de PPV con una memoria de 10 % que se identificó cuando las características del conjunto de características del modelo se clasificaron como características de interacción de alelos, y se muestra en el lado derecho un valor de PPV con una memoria del 10 % que se identificó cuando las características del conjunto de características del modelo se clasificaron como características de no clasificación de alelos. Cabe destacar que la característica de la secuencia de péptidos siempre se clasificó como una característica de interacción de alelos a los fines de la FIG. 13M. Los resultados mostraron que los modelos de presentación alcanzaron un valor de PPV con una memoria de 10 % dentro del intervalo de 14 % hasta 29 %, que es significativamente (aproximadamente 500 veces) mayor que el PPV para una predicción aleatoria.

[00731] Se consideraron para este experimento secuencias de péptidos con longitudes de 9-20. Los datos se dividieron en conjuntos de entrenamiento, validación y evaluación. Los bloques de péptidos de bloques de 50 residuos del Conjunto de datos 1 y el Conjunto de datos 2 se asignaron a los conjuntos de entrenamiento y evaluación. Se retiraron los péptidos que se duplicar en cualquier parte del proteoma, para garantizar que ninguna secuencia de péptidos apareciera en el conjunto de entrenamiento y de evaluación al mismo tiempo. La prevalencia de la presentación de péptidos en el conjunto de entrenamiento y evaluación aumentó 50 veces al retirar los péptidos que no se presentaron. Esto es debido a que el Conjunto de datos 1 y el Conjunto de datos 2 son de muestras tumorales humanas en las que solamente una fracción de las células son alelos de HLA clase II, lo que da como resultado rendimientos de péptidos que fueron aproximadamente 10 veces menores que en las muestras puras de alelos de HLA clase II, que aún se subestima debido a una sensibilidad imperfecta de la espectrometría de masas. El conjunto de entrenamiento contuvo 1064 péptidos presentados y 3.810.070 péptidos no presentados. El conjunto de evaluación contuvo 314 péptidos presentados y 807.400 péptidos no presentados.

[00732] El modelo ilustrativo 1 fue el modelo de suma de funciones en la ecuación (22) con una función de dependencia de red $g_h(\cdot)$, la función expit $f(\cdot)$ y la función de identidad $r(\cdot)$. La función de dependencia de red $g_h(\cdot)$ se estructuró como un perceptrón de múltiples capas (MLP) con 256 nodos ocultos y activaciones de unidad lineal rectificada (ReLU). Además de la secuencia de péptidos, las variables de interacción de alelos $\boldsymbol{w}$ contenían la secuencia de flanqueo del extremo C y el extremo N codificada por one-hot, un índice que indica la variable categórica del gen de origen $G=\text{gen}(\boldsymbol{p}^i)$ del

péptido $\boldsymbol{p}^i$, y una variable que indica la medición de cuantificación del ARNm. El modelo ilustrativo 2 fue idéntico al modelo ilustrativo 1, salvo que se omitió la secuencia de flanqueo del extremo C y extremo N de las variables de interacción de alelos. El modelo ilustrativo 3 fue idéntico al modelo ilustrativo 1, salvo que se omitió el índice del gen de origen de las variables de interacción de alelos. El modelo ilustrativo 4 fue idéntico al modelo ilustrativo 1, salvo que se omitió la medición de cuantificación de ARNm de las variables de interacción de alelos.

[00733] El modelo ilustrativo 5 fue el modelo de suma de funciones en la ecuación (20) con una función de dependencia de red $g_h(\cdot)$, la función expit $f(\cdot)$, la función de identidad $r(\cdot)$ y la función de dependencia $g_w(\cdot)$ de la ecuación (12). La función de dependencia $g_w(\cdot)$ también incluyó un modelo de red que tomó la medición de cuantificación de ARNm como entrada, estructurada como un MLP con 16 nodos ocultos y activaciones de ReLU, y un modelo de red que tomó la secuencia de flanqueo C como entrada, estructurado como un MLP con 32 nodos ocultos y activaciones de ReLU. La función de dependencia de red $g_h(\cdot)$ se estructuró como un perceptrón de múltiples capas con 256 nodos ocultos y activaciones de unidad lineal rectificada (ReLU). El modelo ilustrativo 6 fue idéntico al modelo ilustrativo 5, salvo que se omitió el modelo de red de la secuencia de flanqueo del extremo C y extremo N. El modelo ilustrativo 7 fue idéntico al modelo ilustrativo 5, salvo que se omitió el índice del gen de origen de las variables de no interacción de alelos. El modelo ilustrativo 8 fue idéntico al modelo ilustrativo 5, salvo que se omitió el modelo de red de la medición de cuantificación de ARNm.

[00734] La prevalencia de los péptidos presentados en el conjunto de evaluación fue de aproximadamente 1/2400 y, por lo tanto, el PPV de una predicción aleatoria también sería de aproximadamente 1/2400 = 0,00042. Tal como se muestra en la FIG. 13M, el modelo de presentación con mejor rendimiento alcanzó un valor de PPV de aproximadamente 29 %, que es alrededor de 500 veces mejor que el valor de PPV de una predicción aleatoria.

XII.M. MHC II Ejemplo 2

[00735] La FIG. 13N es un histograma que representa la cantidad de péptidos secuenciados mediante el uso de espectrometría de masas para cada muestra de un total de 39 muestras que comprende moléculas de HLA clase II. Además, para cada muestra de la pluralidad de muestras, el histograma que se muestra en la FIG. 13N representa la cantidad de péptidos secuenciados mediante el uso de espectrometría de masas con distintos umbrales del valor q. Específicamente, para cada muestra de la pluralidad de muestras, la FIG. 13N representa la cantidad de péptidos secuenciados mediante el uso de espectrometría de masas con un valor q menor que 0,01, con un valor q menor que 0,05 y con un valor q menor que 0,2.

[00736] Como se indicó anteriormente, cada una de las 39 muestras de la FIG. 13N comprendía moléculas de HLA clase II. Más específicamente, cada una de las 39 muestras de la FIG. 13N

comprendía moléculas de HLA-DR. La molécula de HLA-DR es un tipo de molécula de HLA clase II. Aun más específicamente, cada una de las 39 muestras de la FIG. 13N comprendía moléculas de HLA-DRB1, moléculas de HLA-DRB3, moléculas de HLA-DRB4 y/o moléculas de HLA-DRB5. La molécula de HLA-DRB1, la molécula de HLA-DRB3, la molécula de HLA-DRB4 y la molécula de HLA-DRB5 son tipos de la molécula de HLA-DR.

[00737] Si bien este experimento en particular se llevó a cabo mediante el uso de muestras que comprenden moléculas de HLA-DR, y particularmente moléculas de HLA-DRB1, moléculas de HLA-DRB3, moléculas de HLA-DRB4 y moléculas de HLA-DRB5, en modalidades alternativas, este experimento puede llevarse a cabo con muestras que comprenden uno o más de cualquier tipo de moléculas de HLA clase II. Por ejemplo, en modalidades alternativas, pueden llevarse a cabo experimentos idénticos con muestras que comprenden moléculas de HLA-DP y/o HLA-DQ. Los expertos en la técnica están familiarizados con esta capacidad de modelar cualquier tipo de moléculas del MHC clase II mediante el uso de las mismas técnicas y aun así alcanzar resultados confiables. Por ejemplo, Jensen, Kamilla Kjaergaard, et ál.[76] es un ejemplo de un documento científico reciente que utiliza métodos idénticos para modelar la afinidad de unión tanto para moléculas de HLA-DR como para moléculas de HLA-DQ y HLA-DP. Por lo tanto, un experto en la técnica entendería que los experimentos y modelos descritos en la presente podrían usarse para modelar de forma separada o simultánea no solo moléculas de HLA-DR, sino cualquier otra molécula del MHC clase II, y producir aun resultados confiables.

[00738] Para secuenciar los péptidos de cada una de las 39 muestras totales, se llevó a cabo espectrometría de masas para cada muestra. El espectro de masas resultante para la muestra se buscó después con Comet y se clasificó con Percolator para secuenciar los péptidos. Después, se identificó la cantidad de péptidos secuenciados en la muestra para múltiples umbrales distintos de valor q de Percolator. Específicamente, para la muestra, se determinó la cantidad de péptidos secuenciados con un valor q de Percolator menor que 0,01, con un valor q de Percolator menor que 0,05 y con un valor q de Percolator menor que 0,2.

[00739] Para cada una de las 39 muestras, la cantidad de péptidos secuenciados con cada uno de los umbrales distintos de valor q de Percolator se representa en la FIG. 13N. Por ejemplo, como se muestra en la FIG. 13N, para la primera muestra, se secuenciaron aproximadamente 4000 péptidos con un valor q menor que 0,2 mediante espectrometría de masas, se secuenciaron aproximadamente 2800 péptidos con un valor q menor que 0,05 mediante espectrometría de masas, y se secuenciaron aproximadamente 2300 péptidos con un valor q menor que 0,01 mediante espectrometría de masas.

[00740] En general, la FIG. 13N demuestra la capacidad de utilizar espectrometría de masas para secuenciar una gran cantidad de péptidos de muestras que contienen moléculas del MHC clase II, con

valores q bajos. En otras palabras, los datos representados en la FIG. 13N demuestran la capacidad de secuenciar de manera confiable péptidos que las moléculas del MHC clase II pueden presentar, mediante el uso de espectrometría de masas.

[00741] La FIG. 13O es un histograma que representa la cantidad de muestras en las que se identificó un alelo particular de molécula del MHC clase II. Más específicamente, para las 39 muestras totales que comprenden moléculas de HLA clase II, la FIG. 13O representa la cantidad de muestras en las que se identificaron determinados alelos de moléculas del MHC clase II.

[00742] Como se trató anteriormente con referencia a la FIG. 13N, cada una de las 39 muestras de la FIG. 13N comprendía moléculas de HLA-DRB1, moléculas de HLA-DRB3, moléculas de HLA-DRB4 y/o moléculas de HLA-DRB5. Por lo tanto, la FIG. 13O representa la cantidad de muestras en las que se identificaron determinados alelos de las moléculas de HLA-DRB1, HLA-DRB3, HLA-DRB4 y HLA-DRB5. Para identificar los alelos de HLA presentes en una muestra, se lleva a cabo un tipado de DR de HLA clase II. Después, para identificar la cantidad de muestras en las que se identificó un alelo particular de HLA, simplemente se suma la cantidad de muestras en las que se identificó el alelo HLA mediante el uso de tipificación de DR de HLA clase II. Por ejemplo, como se muestra en la FIG. 13O, 19 de las 39 muestras totales contenían el alelo de la molécula de HLA clase II HLA-DRB4*01:03. Dicho de otra forma, 19 de las 39 muestras totales contenían el alelo HLA-DRB4*01:03 para la molécula de HLA-DRB4. En general, la FIG. 13O representa la capacidad de identificar una amplia gama de alelos de moléculas del MHC clase II de las 39 muestras que comprenden moléculas de HLA clase II.

[00743] La FIG. 13P es un histograma que representa la proporción de péptidos presentados por las moléculas del MHC clase II en las 39 muestras totales, para cada longitud de péptido de una gama de longitudes de péptidos. Para determinar la longitud de cada péptido en cada una de las 39 muestras totales, se secuenció cada péptido mediante el uso de espectrometría de masas como se trató anteriormente con referencia a la FIG. 13N, y después simplemente se cuantificó la cantidad de residuos en el péptido secuenciado.

[00744] Como se indicó anteriormente, las moléculas del MHC clase II típicamente presentan péptidos con longitudes de 9-20 aminoácidos. Por consiguiente, la FIG. 13P representa la proporción de péptidos presentados por las moléculas del MHC clase II en las 39 muestras para cada longitud de péptido de 9-20 aminoácidos, inclusive. Por ejemplo, como se muestra en la FIG. 13P, aproximadamente 22 % de los péptidos presentados por las moléculas del MHC clase II en las 39 muestras comprende una longitud de 14 aminoácidos.

[00745] En función de los datos que se muestran en la FIG. 13P, se identificaron que las longitudes modales para los péptidos presentados por las moléculas del MHC clase II en las 39 muestras fueron de 14 y 15 aminoácidos. Estas longitudes modales identificadas para los péptidos presentados por las

moléculas del MHC clase II en las 39 muestras coinciden con informes anteriores de longitudes modales para los péptidos presentados por las moléculas del MHC clase II. Además, también congruente con informes anteriores, los datos de la FIG. 13P indican que más del 60 % de los péptidos presentados por las moléculas del MHC clase II de las 39 muestras comprende longitudes que no son de 14 y 15 aminoácidos. Dicho de otra forma, la FIG. 13P indica que, si bien los péptidos presentados por las moléculas del MHC clase II más frecuentemente tienen una longitud de 14 o 15 aminoácidos, una gran proporción de los péptidos presentados por las moléculas del MHC clase II *no* tienen una longitud de 14 o 15 aminoácidos. Por consiguiente, es incorrecto asumir que los péptidos de todas las longitudes tienen probabilidades iguales de ser presentados por las moléculas del MHC clase II, o que las moléculas del MHC clase II solo presentan péptidos que comprenden una longitud de 14 o 15 aminoácidos. Como se trata en mayor detalle más adelante con referencia a la FIG. 13T, estas suposiciones erróneas se usan actualmente en muchos modelos de vanguardia para predecir la presentación de péptidos por parte de moléculas del MHC clase II y, por lo tanto, las probabilidades de presentación previstas por estos modelos a menudo no son confiables.

[00746] La FIG. 13Q es una gráfica lineal que representa la relación entre la expresión génica y la prevalencia de la presentación del producto de expresión génica por parte de una molécula del MHC clase II, para los genes presentes en las 39 muestras. Más específicamente, la FIG. 13Q representa la relación entre la expresión génica y la proporción de residuos resultantes de la expresión génica que conforman el extremo N de un péptido presentado por una molécula del MHC clase II. Para cuantificar la expresión génica en cada una de las 39 muestras totales, se lleva a cabo una secuenciación de ARN del ARN incluido en cada muestra. En la FIG. 13Q, la expresión génica se mide a través de secuenciación de ARN en unidades de transcripciones por millón (TPM). Para identificar la prevalencia de la presentación de los productos de expresión génica para cada una de las 39 muestras, se llevó a cabo para cada muestra una identificación de datos de peptidómica de DR de HLA clase II.

[00747] Como se ilustra en la FIG. 13Q, para las 39 muestras, existe una correlación fuerte entre el nivel de expresión génica y la presentación de los residuos del producto génico expresado por una molécula del MHC clase II. Específicamente, tal como se muestra en la FIG. 13Q, es 100 veces menos probable que los péptidos resultantes de la expresión de los genes menos expresados se presenten en una molécula del MHC clase II, que los péptidos resultantes de la expresión de los genes más expresados. En términos más simples, los productos de los genes más expresados se presentan más frecuentemente en moléculas del MHC clase II.

[00748] Las FIGS. 13H-J son gráficas lineales que comparan el rendimiento de diversos modelos de presentación en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por al menos una de las moléculas del MHC clase II presentes en el

conjunto de datos de evaluación. Tal como se muestra en las FIGS. 13H-J, el rendimiento de un modelo en la predicción de la probabilidad de que un péptido se presente en al menos una de las moléculas del MHC clase II presentes en el conjunto de datos de evaluación se determina mediante la identificación de una relación de la tasa de verdadero positivo con respecto a la tasa de falso positivo para cada predicción realizada por el modelo. Estas relaciones identificadas para un modelo dado pueden visualizarse como una curva de ROC (característica de rendimiento de diagnóstico), en una gráfica lineal donde un eje x cuantifica una tasa de falso positivo y un eje x cuantifica una tasa de verdadero positivo. Se usa un área bajo la curva (AUC) para cuantificar el rendimiento del modelo. Específicamente, un modelo con un AUC mayor tiene un rendimiento mayor (es decir, mayor precisión) con respecto a un modelo con un AUC mayor. En las FIGS. 13H-J, la línea punteada negra con una pendiente de 1 (es decir, una relación de tasa de verdadero positivo con respecto a tasa de falso positivo de 1) representa la curva esperada para adivinar de forma aleatoria las probabilidades de presentación de péptidos. El AUC de la línea punteada es 0,5. Las curvas de ROC y la métrica de AUC se tratan en gran detalle con respecto a la parte superior de la Sección XII que antecede.

[00749] La FIG. 13R es una gráfica lineal que compara el rendimiento de cinco modelos de presentación ilustrativos en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por una molécula del MHC clase II, dados distintos conjuntos de variables de interacción de alelos y de no interacción de alelos. Dicho de otra forma, la FIG. 13R cuantifica la importancia relativa de diversas variables de interacción de alelos y de no interacción de alelos de la predicción de la probabilidad de que un péptido sea presentado por una molécula del MHC clase II.

[00750] La arquitectura del modelo de cada modelo de presentación ilustrativo de los cinco modelos de presentación ilustrativos utilizada para generar las curvas de ROC de la gráfica lineal de la FIG. 13R comprendió un conjunto de cinco modelos de suma de sigmoides. Cada modelo de suma de sigmoides del conjunto se configuró para modelar la presentación de péptidos para hasta cuatro alelos de HLA-DR únicos por muestra. Además, cada modelo de suma de sigmoides del conjunto se configuró para realizar predicciones de probabilidad de presentación de péptidos en función de las siguientes variables de interacción de alelos y de no interacción de alelos: secuencia de péptidos, secuencia flanqueante, expresión de ARN en unidades de TPM, identificador de genes e identificador de muestras. El componente de interacción de alelos de cada modelo de suma de sigmoides del conjunto de MLP con una capa escondida con activaciones de ReLu como 256 unidades escondidas.

[00751] Antes de utilizar los modelos ilustrativos para predecir la probabilidad de que los péptidos de un conjunto de datos de evaluación de péptidos sean presentados por una molécula del MHC clase II, se entrenaron y validaron los modelos ilustrativos. Para entrenar, validar y finalmente evaluar los

modelos ilustrativos, los datos descritos anteriormente para las 39 muestras se dividieron en conjuntos de datos de entrenamiento, validación y evaluación.

[00752] Para garantizar que ningún péptido haya aparecido en más de uno de los conjuntos de datos de entrenamiento, validación y evaluación, se llevó a cabo el siguiente procedimiento. Primero, se eliminaron todos los péptidos de las 39 muestras totales que aparecían en más de una ubicación en el proteoma. Entonces, los péptidos de las 39 muestras totales se dividieron en bloques de 10 péptidos adyacentes. Cada bloque de los péptidos de las 39 muestras totales se asignó de forma única al conjunto de datos de entrenamiento, el conjunto de datos de validación y el conjunto de datos de evaluación. De esta manera, no aparece ningún péptido en más de un conjunto de datos de los conjuntos de datos de entrenamiento, validación y evaluación.

[00753] De los 28.081.944 péptidos en las 39 muestras totales, el conjunto de datos comprendía 21.077 péptidos presentados por moléculas del MHC clase II de 38 de las 39 muestras totales. Los 21.077 péptidos incluidos en el conjunto de datos de entrenamiento tenían entre 9 y 20 aminoácidos de longitud, inclusive. Los modelos ilustrativos utilizados para generar las curvas de ROC de la FIG. 13R se entrenaron con el conjunto de datos de entrenamiento mediante el uso de un optimizador ADAM y detención temprana.

[00754] El conjunto de datos de validación consistía en 2346 péptidos presentados por moléculas del MHC clase II de las mismas 38 muestras utilizadas en el conjunto de datos de entrenamiento. El conjunto de validación se utilizó únicamente para la detención temprana.

[00755] El conjunto de datos de evaluación comprendía péptidos presentados por moléculas del MHC clase II que se identificaron a partir de una muestra tumoral mediante el uso de espectrometría de masas. Específicamente, el conjunto de datos de evaluación comprendía 203 péptidos presentados por moléculas del MHC clase II, específicamente moléculas HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB4*01:03 y HLA-DRB5*01:01, que se identificaron a partir de la muestra tumoral. Los péptidos incluidos en el conjunto de datos de evaluación se mantuvieron fuera del conjunto de datos de entrenamiento descrito anteriormente.

[00756] Como se indicó anteriormente, la FIG. 13R cuantifica la importancia relativa de diversas variables de interacción de alelos y variables de no interacción de alelos de la predicción de la probabilidad de que un péptido sea presentado por una molécula del MHC clase II. También como se indicó anteriormente, los modelos ilustrativos utilizados para generar las curvas de ROC de la gráfica lineal de la FIG. 13R se configuró para realizar predicciones de probabilidad de presentación de péptidos en función de las siguientes variables de interacción de alelos y de no interacción de alelos: secuencia de péptidos, secuencia flanqueante, expresión de ARN en unidades de TPM, identificador de genes e identificador de muestras. Para cuantificar la importancia relativa de cuatro de estas cinco variables

(secuencia de péptidos, secuencia flanqueante, expresión de ARN e identificador génico) para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II, cada modelo ilustrativo de los cinco modelos ilustrativos descritos anteriormente se evaluó mediante el uso de datos del conjunto de datos de evaluación, con una combinación diferente de las cuatro variables. Específicamente, para cada péptido del conjunto de datos de evaluación, un modelo ilustrativo 1 generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia de péptidos, una secuencia flanqueante, un identificador génico y un identificador de muestra, pero no del ARN. De manera similar, para cada péptido del conjunto de datos de evaluación, un modelo ilustrativo 2 generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia de péptidos, expresión de ARN, un identificador génico y un identificador de muestra, pero no de una secuencia flanqueante. De manera similar, para cada péptido del conjunto de datos de evaluación, un modelo ilustrativo 3 generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia flanqueante, expresión de ARN, un identificador génico y un identificador de muestra, pero no de una secuencia de péptidos. De manera similar, para cada péptido del conjunto de datos de evaluación, un modelo ilustrativo 4 generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia flanqueante, expresión de ARN, una secuencia de péptidos y un identificador de muestra, pero no de un identificador génico. Finalmente, para cada péptido del conjunto de datos de evaluación, un modelo ilustrativo 5 generó predicciones de la probabilidad de presentación de péptidos en función de las cinco variables de secuencia flanqueante, expresión de ARN, secuencia de péptidos, identificador de muestra e identificador génico.

[00757] El rendimiento de cada uno de estos cinco modelos ilustrativos se muestra en la gráfica lineal de la FIG. 13R. Específicamente, cada uno de los cinco modelos ilustrativos se asocia con una curva de ROC que representa una relación de verdadero positivo con respecto a la tasa de falso positivo para cada predicción hecha por el modelo. Por ejemplo, la FIG. 13R representa una curva para el modelo ilustrativo 1 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia de péptidos, una secuencia flanqueante, un identificador génico y un identificador de muestra, pero no del ARN. La FIG. 13R representa una curva para el modelo ilustrativo 2 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia de péptidos, expresión de ARN, un identificador génico y un identificador de muestra, pero no de una secuencia flanqueante. La FIG. 13R también representa una curva para el modelo ilustrativo 3 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia flanqueante, expresión de ARN, un identificador génico y un identificador de muestra, pero no de una secuencia de péptidos. La FIG. 13R también representa una curva para el modelo ilustrativo 4 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia flanqueante, expresión de ARN,

una secuencia de péptidos y un identificador de muestra, pero no de un identificador génico. Y, finalmente, la FIG. 13R representa una curva para el modelo ilustrativo 5 que generó predicciones de la probabilidad de presentación de péptidos en función de las cinco variables de secuencia flanqueante, expresión de ARN, secuencia de péptidos, identificador de muestra e identificador génico.

[00758] Como se indicó anteriormente, el rendimiento de un modelo para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II se cuantifica mediante la identificación de un AUC para una curva de ROC que representa una relación de la tasa de verdadero positivo con respecto a la tasa de falso positivo para cada predicción realizada por el modelo. Un modelo con un AUC mayor tiene un rendimiento mayor (es decir, mayor precisión) con respecto a un modelo con un AUC menor. Tal como se muestra en la FIG. 13R, la curva para el modelo ilustrativo 5 que generó predicciones de la probabilidad de presentación de péptidos en función de las cinco variables de secuencia flanqueante, expresión de ARN, secuencia de péptidos, identificador de muestra e identificador génico, logró la mayor AUC de 0,98. Por lo tanto, el modelo ilustrativo 5 que utilizó las cinco variables para generar predicciones de presentación de péptidos logró el mejor rendimiento. La curva para el modelo ilustrativo 2 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia de péptidos, expresión de ARN, un identificador génico y un identificador de muestra, pero no de una secuencia flanqueante, logró la segunda mayor AUC de 0,97. Por lo tanto, la secuencia flanqueante puede identificarse como la variable menos importante para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II. La curva para el modelo ilustrativo 4 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia flanqueante, expresión de ARN, una secuencia de péptidos y un identificador de muestra, pero no de un identificador génico, logró la tercera mayor AUC de 0,96. Por lo tanto, el identificador génico puede identificarse como la segunda variable menos importante para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II. La curva para el modelo ilustrativo 3 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia flanqueante, expresión de ARN, un identificador génico y un identificador de muestra, pero no de una secuencia de péptidos, logró la menor AUC de 0,88. Por lo tanto, la secuencia de péptidos puede identificarse como la variable más importante para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II. La curva para el modelo ilustrativo 1 que generó predicciones de la probabilidad de presentación de péptidos en función de una secuencia de péptidos, una secuencia flanqueante, un identificador génico y un identificador de muestra, pero no del ARN, logró la segunda menor AUC de 0,95. Por lo tanto, la expresión de ARN puede identificarse como la segunda variable más importante para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II.

[00759] La FIG. 13S es una gráfica lineal que compara el rendimiento de cuatro modelos de presentación diferentes en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por una molécula de MHC clase II.

[00760] El primer modelo evaluado en la FIG. 13S se denomina en la presente como un "modelo completo de no interacción". El modelo completo de no interacción es una modalidad de los modelos de representación descritos anteriormente en los cuales las variables de no interacción de alelos $\boldsymbol{w^k}$ y las variables de interacción de alelos $\boldsymbol{x_h^k}$ se ingresan en funciones de dependencia separadas, tal como, por ejemplo, una red neuronal, y, entonces, las salidas de estas funciones de dependencia separadas se suman. Específicamente, el modelo completo de no interacción es una modalidad de los modelos de representación descritos anteriormente en los cuales las variables de no interacción de alelos $\boldsymbol{w^k}$ se ingresan en una función de dependencia $\boldsymbol{g_w}$, las variables de interacción de alelos $\boldsymbol{x_h^k}$ se ingresan en una función de dependencia separada $\boldsymbol{g_h}$, y las salidas de la función de dependencia $\boldsymbol{g_w}$ y la función de dependencia $\boldsymbol{g_h}$ se suman entre sí. Por lo tanto, en algunas modalidades, el modelo completo de no interacción determina la probabilidad de presentación de péptidos mediante el uso de la ecuación 8, como se mostró anteriormente. Además, las modalidades del modelo completo de no interacción en las cuales las variables de no interacción de alelos $\boldsymbol{w^k}$ se ingresan en una función de dependencia $\boldsymbol{g_w}$, las variables de interacción de alelos $\boldsymbol{x_h^k}$ se ingresan en una función de dependencia separada $\boldsymbol{g_h}$, y las salidas de la función de dependencia $\boldsymbol{g_w}$ y la función de dependencia $\boldsymbol{g_h}$ se suman, se discuten en detalle anteriormente con respecto a la parte superior de la Sección X.B.2., la parte inferior de la Sección X.B.3., la parte superior de la Sección X.C.3. y la parte superior de la Sección X.C.6.

[00761] El segundo modelo evaluado en la FIG. 13S se denomina en la presente como un "modelo completo de interacción". El modelo completo de interacción es una modalidad de los modelos de representación descritos anteriormente en los cuales las variables de no interacción de alelos $\boldsymbol{w^k}$ se concatenan directamente con las variables de interacción de alelos $\boldsymbol{x_h^k}$ antes de ingresarse en una función de dependencia separada, tal como, por ejemplo, una red neuronal. Por lo tanto, en algunas modalidades, el modelo completo de interacción determina la probabilidad de presentación de péptidos mediante el uso de la ecuación 9, como se mostró anteriormente. Además, las modalidades del modelo completo de interacción en el cual las variables de no interacción de alelos $\boldsymbol{w^k}$ se concatenan con las variables de interacción de alelos $\boldsymbol{x_h^k}$ antes de que las variables se ingresen en una función de dependencia se discuten en detalle anteriormente con respecto a la parte inferior de la Sección X.B.2., la parte inferior de la Sección X.C.2. y la parte inferior de la Sección X.C.5.

[00762] El tercer modelo evaluado en la FIG. 13S se denomina en la presente como un "modelo de CNN". El modelo de CNN comprende una red neuronal convolucional y es similar al modelo completo de no interacción descrito anteriormente. Sin embargo, las capas de la red neuronal convolucional del

modelo de CNN difieren de las capas de la red neuronal del modelo completo de no interacción. Específicamente, la capa de entrada de la red neuronal convolucional del modelo de CNN acepta una cadena de péptidos de 20-mer y posteriormente incorpora la cadena de péptidos de 20-mer como un tensor (n, 20, 21). Las siguientes capas de la red neuronal convolucional del modelo de CNN comprenden una capa de filtro convolucional 1-D con tamaño 5 y stride de 1, una agrupación de muestreo máxima global, una capa de dropout con p=0,2 y finalmente una capa densa de 34 nodos con una activación de ReLu.

[00763] El cuarto y último modelo evaluado en la FIG. 13S se denomina en la presente “modelo de LSTM”. El modelo de LSTM comprende una red neuronal de memoria a largo plazo. La capa de entrada de la red neuronal de memoria a largo plazo del modelo de LSTM acepta una cadena de péptidos de 20-mer y posteriormente incorpora la cadena de péptidos de 20-mer como un tensor (n, 20, 21). Las siguientes capas de la red neuronal de memoria a largo plazo del modelo de LSTM comprenden una capa de memoria a largo plazo de 128 nodos, una capa de dropout con p=0,2 y finalmente una capa densa de 34 nodos con una activación de ReLu.

[00764] Antes de usar cada uno de los cuatro modelos de la FIG. 13S para predecir la probabilidad de que los péptidos del conjunto de datos de evaluación de péptidos se presenten en una molécula de MHC clase II, los modelos se entrenaron mediante el uso del conjunto de datos de entrenamiento de 38 muestras descrito anteriormente y se validaron mediante el uso del conjunto de datos de validación descrito anteriormente. Después de este entrenamiento y validación de los modelos, cada uno de los cuatro modelos se evaluó con el conjunto de datos de evaluación de la muestra 39 retenido descrito anteriormente. Específicamente, para cada uno de los cuatro modelos, se ingresó en el modelo cada péptido del conjunto de datos de evaluación, y el modelo produjo posteriormente una probabilidad de presentación para el péptido.

[00765] El rendimiento de cada uno de los cuatro modelos se representa en la gráfica lineal de la FIG. 13S. Específicamente, cada uno de los cuatro modelos se asocia con una curva de ROC que representa una relación de la tasa de verdadero positivo con respecto a la tasa de falso positivo para cada predicción realizada por el modelo. Por ejemplo, la FIG. 13S ilustra una curva de ROC para el modelo de CNN, una curva de ROC para el modelo de interacción completa, una curva de ROC para el modelo de LSTM y una curva de ROC para el modelo de no interacción completa.

[00766] Como se indicó anteriormente, el rendimiento de un modelo para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II se cuantifica mediante la identificación de un AUC para una curva de ROC que representa una relación de la tasa de verdadero positivo con respecto a la tasa de falso positivo para cada predicción realizada por el modelo. Un modelo con un AUC mayor tiene un rendimiento mayor (es decir, mayor precisión) con respecto a un modelo con un AUC

menor. Tal como se muestra en la FIG. 13S, la curva del modelo de interacción completa tuvo el AUC máxima de 0,982. Por lo tanto, el modelo de interacción completa tuvo el mejor rendimiento. La curva del modelo de no interacción completa tuvo la segunda AUC máxima de 0,977. Por lo tanto, el modelo de no interacción completa tuvo el segundo mejor rendimiento. La curva del modelo de CNN tuvo el AUC más baja de 0,947. Por lo tanto, el modelo de CNN tuvo el peor rendimiento. La curva del modelo de LSTM tuvo la segunda AUC más baja de 0,952. Por lo tanto, el modelo de LSTM tuvo el segundo peor rendimiento. Sin embargo, cabe destacar que todos los modelos evaluados en la FIG. 13S tienen un AUC mayor que 0,9. Por consiguiente, a pesar de la varianza en la arquitectura entre estos, todos los modelos evaluados en la FIG. 13S pueden tener predicciones relativamente precisas de presentación de péptidos.

[00767] La FIG. 13T es una gráfica lineal que compara el rendimiento de dos modelos ilustrativos de primera clase de la técnica previa dados dos criterios distintos y dos modelos ilustrativos de presentación dados dos conjuntos distintos de variables de interacción de alelos y de no interacción de alelos en la predicción de la probabilidad de que los péptidos de un conjunto de datos de evaluación sean presentados por una molécula de MHC clase II. Específicamente, la FIG. 13T es una gráfica lineal que compara el rendimiento de un modelo ilustrativo de primera clase de la técnica previa que utiliza una afinidad de unión mínima prevista de NetMHCII 2.3 como criterio para generar predicciones (modelo ilustrativo 1), un modelo ilustrativo de primera clase de la técnica previa que utiliza un rango de unión mínimo previsto de NetMHCII 2.3 como criterio para generar predicciones (modelo ilustrativo 2), un modelo ilustrativo de presentación que genera predicciones de probabilidad de presentación de péptidos en función del tipo de molécula de MHC clase II y secuencia de péptidos (modelo ilustrativo 4), y un modelo ilustrativo de presentación que genera predicciones de la probabilidad de presentación de péptidos en función del tipo de molécula de MHC clase II, secuencia de péptidos, expresión de ARN, identificador de genes y secuencia flanqueante (modelo ilustrativo 3).

[00768] El modelo de primera clase de técnica previa usado como modelo ilustrativo 1 y modelo ilustrativo 2 en la FIG. 13T es el modelo de NetMHCII 2.3. El modelo de NetMHCII 2.3 genera predicciones de probabilidad de presentación de péptidos en función de la secuencia de péptidos y tipo de molécula de MHC clase II. El modelo de NetMHCII 2.3 se evaluó mediante el uso del sitio web de NetMHCII 2.3 (www.cbs.dtu.dk/services/NetMHCII/, PMID 29315598)[76].

[00769] Como se indicó anteriormente, el modelo de NetMHCII 2.3 se evaluó de acuerdo con dos criterios distintos. Específicamente, el modelo ilustrativo 1 generó predicciones de probabilidad de presentación de péptidos de acuerdo con la afinidad de unión mínima prevista por NetMHCII 2.3, y el modelo ilustrativo 2 generó predicciones de probabilidad de presentación de péptidos de acuerdo con el rango de unión mínimo previsto por NetMHCII 2.3.

[00770] El modelo de presentación utilizado como modelo ilustrativo 3 y modelo ilustrativo 4 es una modalidad del modelo de presentación descrito en la presente que se entrena mediante el uso de datos obtenidos mediante espectrometría de masas. Como se indicó anteriormente, el modelo de presentación generó predicciones de probabilidad de presentación de péptidos en función de dos conjuntos distintos de variables de interacción de alelos y de no interacción de alelos. Específicamente, el modelo ilustrativo 4 generó predicciones de probabilidad de presentación de péptidos en función de la secuencia de péptidos y tipo de molécula de MHC clase II (la misma variable utilizada por el modelo de NetMHCII 2.3), y el modelo ilustrativo 3 generó predicciones de probabilidad de presentación de péptidos en función del tipo de molécula de MHC clase II, secuencia de péptidos, expresión de ARN, identificador de genes y secuencia flanqueante.

[00771] Antes de usar los modelos ilustrativos de la FIG. 13T para predecir la probabilidad de que los péptidos del conjunto de datos de evaluación de péptidos sean presentados por una molécula de MHC clase II, los modelos se entrenaron y validaron. El modelo de NetMHCII 2.3 (modelo ilustrativo 1 y modelo ilustrativo 2) se entrenó y se validó con sus propios conjuntos de datos de entrenamiento y validación en función de ensayos de afinidad de unión a péptidos-HLA depositados en la base de datos de epítopos inmunitarios (IEDB, www.iedb.org). Se sabe que el conjunto de datos de entrenamiento utilizado para entrenar el modelo de NetMHCII 2.3 comprende casi exclusivamente péptidos de 15-mer. Por otro lado, los modelos ilustrativos 3 y 4 se entrenaron mediante el uso del conjunto de datos de entrenamiento descrito anteriormente con respecto a la FIG. 13R y se validaron con el conjunto de datos de validación descrito anteriormente con respecto a la FIG. 13R.

[00772] Después del entrenamiento y validación de los modelos, cada uno de los modelos se evaluó con un conjunto de datos de evaluación. Como se indicó anteriormente, el modelo de NetMHCII 2.3 se entrena en un conjunto de datos que comprende casi exclusivamente péptidos de 15-mer, lo que significa que NetMHCII 3.2 no tiene la capacidad de darle una prioridad distinta a péptidos de distintos pesos, mediante lo cual se reduce el rendimiento de predicción de NetMHCII 3.2 sobre los datos de espectrometría de masas de presentación de HLA clase II que contienen péptidos de todas las longitudes. Por lo tanto, para proporcionar una comparación justa entre los modelos no afectados por la longitud variable de los péptidos, el conjunto de datos de evaluación incluyó exclusivamente péptidos de 15-mer. Específicamente, el conjunto de datos de evaluación comprendió 933 péptidos de 15-mer. 40 de los 933 péptidos del conjunto de datos de evaluación fueron por moléculas de MHC clase II—específicamente por las moléculas de HLA-DRB1*07:01, HLA-DRB1*15:01, HLA-DRB4*01:03 y HLA-DRB5*01:01. Los péptidos incluidos en el conjunto de datos de entrenamiento se retuvieron de los conjuntos de datos de entrenamiento descritos anteriormente.

[00773] Para evaluar los modelos ilustrativos mediante el uso del conjunto de datos de evaluación,

para cada uno de los modelos ilustrativos, para cada uno de los 933 péptidos del conjunto de datos de evaluación, el modelo generó una predicción de la probabilidad de presentación para el péptido. Específicamente, para cada péptido del conjunto de datos de evaluación, el modelo ilustrativo 1 generó un puntaje de presentación para el péptido con las moléculas de MHC clase II mediante el uso de la secuencia de péptidos y tipos de molécula de MHC clase II, al clasificar el péptido según la afinidad de unión mínima prevista por NetMHCII 2.3 entre los cuatro alelos DR de MHC clase II del conjunto de datos de evaluación. De forma similar, para cada péptido del conjunto de datos de evaluación, el modelo ilustrativo 2 generó un puntaje de presentación para el péptido con las moléculas de MHC clase II mediante el uso de la secuencia de péptidos y tipos de molécula de MHC clase II, al clasificar el péptido según el rango de unión mínimo previsto por NetMHCII 2.3 (es decir, la afinidad de unión normalizada por cuantil) entre los cuatro alelos DR de MHC clase II del conjunto de datos de evaluación. Para cada péptido del conjunto de datos de evaluación, el modelo ilustrativo 4 generó una probabilidad de presentación para el péptido con las moléculas de MHC clase II en función de la secuencia de péptidos y tipo de molécula de MHC clase II. De forma similar, para cada péptido del conjunto de datos de evaluación, el modelo ilustrativo 3 generó una probabilidad de presentación para el péptido con las moléculas de MHC clase II en función del tipo de molécula de MHC clase II, secuencia de péptidos, expresión de ARN, identificador de genes y secuencia flanqueante.

[00774] El rendimiento de cada uno de los cuatro modelos ilustrativos se representa en la gráfica lineal de la FIG. 13T. Específicamente, cada uno de los cuatro modelos ilustrativos se asocia con una curva de ROC que representa una relación de la tasa de verdadero positivo con respecto a la tasa de falso positivo para cada predicción realizada por el modelo. Por ejemplo, la FIG. 13T representa una curva de ROC para el modelo ilustrativo 1 que utilizó la afinidad de unión mínima prevista con NetMHCII 2.3 para generar predicciones, una curva de ROC para el modelo ilustrativo 2 que utilizó un rango de unión mínimo previsto con NetMHCII 2.3 para generar predicciones, una de ROC para el modelo ilustrativo 4 que generó probabilidades de presentación de péptidos en función de la secuencia de péptidos y tipo de molécula de MHC clase II y una curva de ROC para el modelo ilustrativo 3 que generó probabilidades de presentación de péptidos en función del tipo de molécula de MHC clase II, secuencia de péptidos, expresión de ARN, identificador de genes y secuencia flanqueante.

[00775] Como se indicó anteriormente, el rendimiento de un modelo para predecir la probabilidad de que un péptido sea presentado por una molécula del MHC clase II se cuantifica mediante la identificación de un AUC para una curva de ROC que representa una relación de la tasa de verdadero positivo con respecto a la tasa de falso positivo para cada predicción realizada por el modelo. Un modelo con un AUC mayor tiene un rendimiento mayor (es decir, mayor precisión) con respecto a un modelo con un AUC menor. Tal como se muestra en la FIG. 13T, la curva para el modelo ilustrativo 3 que generó

probabilidades de presentación de péptidos basándose en el tipo de molécula de MHC clase II, secuencia de péptidos, expresión de ARN, identificador génico y secuencia flanqueante, logró la mayor AUC de 0,95. Por lo tanto, el modelo ilustrativo 3 que generó probabilidades de presentación de péptidos basándose en el tipo de molécula de MHC clase II, secuencia de péptidos, expresión de ARN, identificador génico y secuencia flanqueante, logró el mejor rendimiento. La curva para el modelo ilustrativo 4 que generó probabilidades de presentación de péptidos basándose en el tipo de molécula de MHC clase II y secuencia de péptidos, logró la segunda mayor AUC de 0,91. Por lo tanto, el modelo ilustrativo 4 que generó probabilidades de presentación de péptidos basándose en el tipo de molécula de MHC clase II y secuencia de péptidos, logró el segundo mejor rendimiento. La curva para el modelo ilustrativo 1 que utilizó la afinidad de unión mínima prevista de NetMHCII 2.3 para generar predicciones logró la menor AUC de 0,75. Por lo tanto, el modelo ilustrativo 1 que utilizó la afinidad de unión mínima prevista de NetMHCII 2.3 para generar predicciones logró el peor rendimiento. La curva para el modelo ilustrativo 2 que utilizó el rango de unión mínimo previsto de NetMHCII 2.3 para generar predicciones logró la segunda menor AUC de 0,76. Por lo tanto, el modelo ilustrativo 2 que utilizó el rango de unión mínimo previsto de NetMHCII 2.3 para generar predicciones logró el segundo peor rendimiento.

[00776] Tal como se muestra en la FIG. 13T, la discrepancia en rendimiento entre los modelos ilustrativos 1 y 2 y los modelos ilustrativos 3 y 4 es grande. Específicamente, el rendimiento del modelo de NetMHCII 2.3 (que utiliza ya sea el criterio de afinidad de unión mínima prevista de NetMHCII 2.3 o el rango de unión mínimo previsto de NetMHCII 2.3) es casi un 25 % inferior que el rendimiento del modelo de presentación descrito en la presente (que genera las probabilidades de presentación de péptidos basándose ya sea en el tipo de molécula de MHC clase II y la secuencia de péptidos o en el tipo de molécula de MHC clase II, la secuencia de péptidos, expresión de ARN, identificador génico y secuencia flanqueante). Por lo tanto, la FIG. 13T demuestra que los modelos de presentación descritos en la presente pueden lograr predicciones significativamente más precisas que el modelo de primera clase de la técnica previa, el modelo de NetMHCII 2.3.

[00777] Además, como se discutió anteriormente, el modelo de NetMHCII 2.3 se entrena en un conjunto de datos de entrenamiento que comprende casi exclusivamente péptidos de 15-mer. Como resultado de esto, el modelo de NetMHCII 2.3 no se entrena para aprender qué longitudes de péptidos son más probables que se presenten por las moléculas del MHC clase II. Por lo tanto, el modelo de NetMHCII 2.3 no pondera sus predicciones en la probabilidad de presentación de péptidos por las moléculas del MHC clase II de acuerdo con la longitud del péptido. En otras palabras, el modelo de NetMHCII 2.3 no modifica sus predicciones de la probabilidad de presentación de péptidos por las moléculas del MHC clase II para péptidos que tienen longitudes por fuera de la longitud de péptido modal de 15 aminoácidos. Como resultado de esto, el modelo de NetMHCII 2.3 predice una probabilidad

excesiva de presentación de péptidos con longitudes mayores o menores que 15 aminoácidos.

[00778] Por otro lado, los modelos de presentación descritos en la presente se entrenan mediante el uso de datos de péptidos obtenidos a través de espectrometría de masas y, por lo tanto, se pueden entrenar en conjuntos de datos de entrenamiento que comprenden péptidos de todas las longitudes diferentes. Como resultado de esto, los modelos de presentación descritos en la presente pueden aprender qué longitudes de péptidos son más probables que se presenten por las moléculas del MHC clase II. Por lo tanto, los modelos de presentación descritos en la presente pueden ponderar las predicciones en la probabilidad de presentación de péptidos por las moléculas del MHC clase II de acuerdo con la longitud del péptido. En otras palabras, los modelos de presentación descritos en la presente pueden modificar sus predicciones de la probabilidad de presentación de péptidos por las moléculas del MHC clase II para péptidos que tienen longitudes por fuera de la longitud de péptido modal de 15 aminoácidos. Como resultado de esto, los modelos de presentación descritos en la presente pueden lograr predicciones significativamente más precisas para péptidos con longitudes mayores o menores que 15 aminoácidos, que el modelo de primera clase de la técnica previa, el modelo de NetMHCII 2.3. Esta es una ventaja del uso de los modelos de presentación descritos en la presente para predecir la probabilidad de presentación de péptidos por las moléculas del MHC clase II.

XII.N. Ejemplo de parámetros determinados para alelos de MHC II

[00779] A continuación, se muestra un conjunto de parámetros determinados para una variación del modelo de presentación de múltiples alelos (ecuación (16)) que genera probabilidades de presentación implícita por alelo para los alelos del MHC clase II HLA-DRB1*12:01 and HLA-DRB1*10:01:

$$\boldsymbol{u} = \mathrm{expit}\left(\mathrm{relu}\left(\mathbf{X} \cdot \mathbf{W^1} + \mathbf{b^1}\right) \cdot \mathbf{W^2} + \mathbf{b^2}\right),$$

donde relu(·) es la función de unidad lineal rectificada (RELU) y $\mathbf{W^1}$, $\mathbf{b^1}$, $\mathbf{W^2}$ y $\mathbf{b^2}$ son el conjunto de parámetros $\boldsymbol{\theta}$ determinado para el modelo. Las variables de interacción de alelos **X** están contenidas en una matriz de 1 x 399) que consiste en 1 fila de secuencias de péptidos con relleno central y codificadas por one-hot por péptido de entrada. Las dimensiones de $\mathbf{W^1}$ son (399 x 256), las dimensiones de $\mathbf{b^1}$ (1 x 256), las dimensiones de $\mathbf{W^2}$ son (256 x 2), y $\mathbf{b^2}$ son (1 x 2). La primera columna de la entrada indica la probabilidad implícita de presentación por alelo para la secuencia de péptidos mediante el alelo HLA-DRB1*12:01 y la segunda columna de la salida indica la implícita por alelo para la secuencia de péptidos mediante el alelo HLA-DRB1*10:01. Con fines demostrativos, los valores para $\mathbf{W^1}$, $\mathbf{b^1}$, $\mathbf{W^2}$ y $\mathbf{b^2}$ se describen en detalle en la solicitud internacional PCT/US2018/028438, que se incorpora a la presente mediante esta referencia por toda su descripción.

XIII. Computadora de ejemplo

[00780] La Figura 14 ilustra una computadora de ejemplo 1400 para la implementación de las entidades mostradas en las Figuras 1 y 3. La computadora 1400 incluye al menos un procesador 1402 acoplado a un conjunto de chips 1404. El conjunto de chips 1404 incluye un núcleo controlador de memoria 1420 y un núcleo controlador de entrada/salida (I/O) 1422. Una memoria 1406 y un adaptador de gráficos 1412 se acopla al núcleo controlador de memoria 1420 y una pantalla 1418 se acopla al adaptador de gráficos 1412. Un dispositivo de almacenamiento 1408, un dispositivo de entrada 1414 y un adaptador de red 1416 se acoplan al núcleo controlador I/O 1422. Otras modalidades de la computadora 1400 tienen diferentes arquitecturas.

[00781] El dispositivo de almacenamiento 1408 es un medio de almacenamiento legible por computadora no transitorio tal como un disco duro, un disco compacto memoria de solo lectura (CD-ROM), DVD o un dispositivo de memoria en estado sólido. La memoria 1406 mantiene las instrucciones y datos utilizados por el procesador 1402. La interfaz de entrada 1414 es una interfaz de pantalla táctil, un ratón, una bola de seguimiento u otro tipo de dispositivo de señalamiento, un teclado o alguna combinación de estos y se utiliza para ingresar datos en la computadora 1400. En algunas modalidades, la computadora 1400 puede configurarse para recibir la entrada (por ejemplo, comandos) desde la interfaz de entrada 1414 mediante gestos del usuario. El adaptador de gráficos 1412 muestra imágenes y otra información en la pantalla 1418. El adaptador de red 1416 acopla la computadora 1400 a una o más redes informáticas.

[00782] La computadora 1400 adaptada para ejecutar módulos de programas informáticos para proporcionar la funcionalidad descrita en la presente. Tal como se usa en la presente, el término «módulo» se refiere a la lógica del programa informático utilizada para proporcionar la funcionalidad especificada. Por lo tanto, un módulo se puede implementar en hardware, firmware y/o software. En una modalidad, los módulos de programa se almacenan en el dispositivo de almacenamiento 1408, se cargan en la memoria 1406 y se ejecutan con el procesador 1402.

[00783] Los tipos de computadoras 1400 utilizados por las entidades de la Figura 1 pueden variar dependiendo de la modalidad y el poder de procesamiento exigido por la entidad. Por ejemplo, el sistema de identificación de la presentación 160 puede ejecutarse en una computadora única 1400 o computadoras múltiples 1400 comunicarse entre sí a través de la red tales como torres de servidores. Las computadoras 1400 pueden no tener algunos de los componentes descritos anteriormente, tales como los adaptadores gráficos 1412 y pantallas 1418.

XIV. Ejemplo de vector de administración del neoantígeno

[00784] A continuación figuran ejemplos de modalidades específicas para llevar a cabo la presente invención. Los ejemplos se proporcionan solamente con fines ilustrativos y no pretenden limitar el alcance de la presente invención de ninguna forma. Se han llevado a cabo esfuerzos para asegurar la

exactitud de los números usados (por ejemplo, cantidades, temperaturas, etc.), pero se deben tener en cuenta algunos errores y desviaciones experimentales.

[00785] La práctica de la presente invención empleará, a menos que se indique lo contrario, métodos convencionales de química de proteínas, bioquímica, técnicas de ADN recombinante y farmacología, dentro de técnica. Tales técnicas se explican en su totalidad en la bibliografía. Véase, por ejemplo, T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et ál., *Molecular Cloning: A Laboratory Manual* (2.ª edición, 1989); *Methods In Enzymology* (S. Colowick y N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pensilvania: Mack Publishing Company, 1990); Carey y Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vol. A y B(1992).

XIV.A. Diseño del casete del neoantígeno

[00786] A través de la vacunación, pueden administrarse neoantígenos específicos al tumor (TSNA) restringidos a múltiples MHC de clase I que estimulan la o las respuestas inmunitarias celulares correspondientes. En un ejemplo, se diseña un casete de vacuna para codificar múltiples epítopos como un producto génico único donde los epítopos se integraron dentro de su secuencia de péptidos circundantes naturales o separadas por secuencias de enlace no naturales. Se identificaron varios parámetros de diseño que podrían impactar potencialmente el procesamiento y presentación del antígeno y, por lo tanto, la magnitud y amplitud de las respuestas de linfocitos T CD8 específicos de TSNA. En el presente ejemplo, se diseñaron diversos casetes modelo y se construyeron para evaluar: (1) si las respuestas de linfocitos T potentes podrían generarse para múltiples epítopos incorporados en un casete de expresión único; (2) qué hace un enlazador óptimo colocado entre los TSNA dentro del casete de expresión que conduce a un procesamiento y presentación óptima de todos los epítopos; (3) si la posición relativa de los epítopos dentro del casete tienen un efecto en las respuestas de linfocitos T; (4) si la cantidad de epítopos dentro de un casete influencia la magnitud o calidad de las respuestas del linfocito T a epítopos individuales; (5) si la adición de las secuencias de direccionamiento celular mejora las respuestas de linfocitos T.

[00787] Se desarrollaron dos lecturas para evaluar la presentación de antígenos y respuestas de linfocitos T específicas para los epítopos marcadores dentro de los casetes modelo: (1) un análisis basado en células in vitro que permite la evaluación de la presentación del antígeno como determinada por la activación de linfocitos T indicadores especialmente diseñados (Aarnoudse et ál., 2002; Nagai et ál., 2012); y (2) un ensayo in vivo que utilizó ratones transgénicos HLA-A2 (Vitiello et ál., 1991) para evaluar la inmunogenicidad luego de la vacunación de epítopos derivados del casete de origen humano por sus respuestas de linfocitos T específicos del epítopo correspondientes (Cornet et ál., 2006; Depla

et ál., 2008; Ishioka et ál., 1999).

XIV.B. Evaluación del diseño del casete del neoantígeno

XIV.B.1. Métodos y materiales

Diseño y clonación del TCR y el casete

[00788] Los TCR seleccionados reconocen péptidos NLVPMVATV (SEQ ID NO: 132) (PDB# 5D2N), CLGGLLTMV (SEQ ID NO: 133) (PDB#3REV), GILGFVFTL (SEQ ID NO: 134) (PDB#1OGA) LLFGYPVYV (SEQ ID NO: 135) (PDB#1AO7) cuando se presentan mediante A*0201. Se construyeron vectores de transferencia que contienen subunidades de TCR unidas al péptido 2A (beta seguido de alfa), los IRES de EMCV y subunidades CD8 unidas a 2A (beta seguida por alfa y por el gen de resistencia de puromicina). Las secuencias de marco abierto se optimizaron por codones y se sintetizaron mediante GeneArt.

Generación de líneas celulares para estudios de procesamiento y presentación del epítopo in vitro

[00789] Se adquirieron péptidos de ProImmune o Genscript diluidos hasta 10 mg/mL con tris(2-carboxietil)fosfina (TCEP) con 10 mM en agua/DMSO (2:8, v/v). El medio de cultivo celular y complementos, a menos que se establezca lo contrario, se obtuvieron de Gibco. El suero bovino fetal inactivado con calor (GBShi) se obtuvo de Seradigm. QUANTI-Luc Substrate, Zeocin, y Puromycin se obtuvieron de InvivoGen. Las células NFAT Jurkat-Lucia (InvivoGen) se mantuvieron en RPMI 1640 complementado con FBShi al 10 %, piruvato de sodio y Zeocina 100 µg/mL. Una vez que se transducen, estas células recibieron adicionalmente puromicina 3 µg/mL. Las células T2 (ATCC CRL-1992) se cultivaron en medio Iscove (IMDM) más FBShi al 20 %. Las células U-87 MG (ATCC HTB-14) se mantuvieron en un medio MEM Eagles complementado con FBShi al 10 %.

[00790] Las células NFAT Jurkat-Lucia contienen una construcción indicadora Lucia que puede inducirse mediante NFAT. El gen Lucia, cuando se activa mediante el acoplamiento del receptor de linfocitos T (TCR) provoca la segregación de una luciferasa que utiliza coelenterazina en el medio de cultivo. Esta luciferasa puede medirse mediante el uso del reactivo de detección de luciferasa QUANTI-Luc. Las células Jurkat-Lucia se transdujeron con lentivirus para expresar TCR específicas de antígenos. Se obtuvo el vector de transferencia de lentivirus derivado de VIH de GeneCopoeia y los plásmidos de soporte de lentivirus que expresan VSV-G (pCMV-VsvG), Rev (pRSV-Rev) y Gag-pol (pCgpV) se obtuvieron de Cell Design Labs.

[00791] Se prepararon lentivirus mediante transfección de matraces T75 con confluencia de 50-80 % de células HEK293 con Lipofectamina 2000 (Thermo Fisher), con el uso de 40 µl de lipofectamina y 20 µg de la mezcla de ADN (4:2:1:1 en peso del plásmido de transferencia:pCgpV:pRSV-Rev:pCMV-VsvG). 8-10 mL del medio con virus se concentraron mediante el uso del sistema Lenti-X (Clontech), y el virus

se resuspendió en 100-200 µl de medio fresco. Se utilizó este volumen para superponer un volumen igual de células Jurkat-Lucia (se utilizaron células 5x10E4-1x10E6 en diferentes experimentos). Luego del cultivo en 0,3 µg/ml de medio con puromicina, las células se clasificaron para obtener capacidad de clonación. Estos clones de TCR de Jurkat-Lucia se evaluaron para determinar la actividad y selectividad mediante el uso de células T2 cargadas con péptidos.

Ensayo de procesamiento y presentación de epítopos in vitro

[00792] Los linfocitos T2 se utilizan normalmente para examinar el reconocimiento del antígeno mediante TCR. Los linfocitos T2 no tienen un transportador de péptidos para el procesamiento de antígenos (con deficiencia de TAP) y no pueden cargar péptidos endógenos en el retículo endoplasmático para la presentación en el MHC. Sin embargo, los linfocitos T2 pueden cargarse fácilmente con péptidos exógenos. Los cinco péptidos marcadores (NLVPMVATV (SEQ ID NO: 132), CLGGLLTMV (SEQ ID NO: 133), GLCTLVAML (SEQ ID NO: 136), LLFGYPVYV (SEQ ID NO: 135), GILGFVFTL (SEQ ID NO: 134)) y dos péptidos irrelevantes (WLSLLVPFV (SEQ ID NO: 137), FLLTRICT (SEQ ID NO: 138)) se cargaron en linfocitos T2. Brevemente, los linfocitos T2 se contaron y diluyeron hasta 1x106 células/mL con IMDM más FBShi al 1 %. Se agregaron péptidos hasta obtener 10 µg de péptido/1x106 células. Las células se incubaron luego a 37 °C durante 90 minutos. Se lavaron las células dos veces con IMDM más FBShi al 20 % diluidas hasta 5x10E5 células/mL y se colocaron 100 µL en placas de 96 pocillos Costar para cultivo de tejidos. Se contaron y se diluyeron clones de TCR Jurkat-Lucia hasta 5x10E5 células/mL en RPMI 1640 más FBShi al 10 % y se agregaron 100 µL a las células T2. Se incubaron placas durante la noche 37 ºC, CO2 al 5 %. Las placas se centrifugaron a 400 g durante 3 minutos y se retiraron 20 µL del sobrenadante a una placa Greiner de fondo plano blanca. Se preparó sustrato QUANTI-Luc de acuerdo con las instrucciones y se agregaron 50 µL/pocillo. Se leyó la expresión de luciferasa en un dispositivo molecular SpectraMax iE3x.

[00793] Para evaluar la presentación del epítopo del marcador mediante los casetes adenovirales, se utilizaron células U-87 MG como células que presentan antígenos sustitutos (APC) y se transdujeron con los vectores adenovirales. Se cosecharon células U-87 MG y se colocaron en placas en medio de cultivo como 5x10E5 células/100 µl en una placa para cultivo de tejido Costar de 96 pocillos. Las placas se incubaron durante aproximadamente 2 horas a 37 ºC. Se diluyeron casetes adenovirales con MEM mas FBShi al 10 % hasta una MOI de 100, 50, 10, 5, 1 y 0 y se agregaron las células U-87 MG como 5 µl/pocillo. Las placas se incubaron nuevamente durante aproximadamente 2 horas a 37 ºC. Se contaron y se diluyeron clones de TCR Jurkat-Lucia hasta 5x10E5 células/mL en RPMI más FBShi al 10 % y se agregaron a las células U-87 MG como 100 µL/pocillo. Las placas se incubaron nuevamente durante aproximadamente 24 horas a 37 ºC, CO2 al 5 %. Las placas se centrifugaron a 400 g durante 3 minutos y se retiraron 20 µL del sobrenadante a una placa Greiner de fondo plano blanca. Se preparó sustrato

QUANTI-Luc de acuerdo con las instrucciones y se agregaron 50 µL/pocillo. Se leyó la expresión de luciferasa en un dispositivo molecular SpectraMax iE3x.

Cepas de ratón para estudios de inmunogenicidad

[00794] Se obtuvieron ratones HLA-A2.1 (HLA-A2 Tg) transgénico de Taconic Labs, Inc. Estos ratones portan un transgen que consiste en una molécula de clase I quimérica que consiste en el líder HLA-A2.1, dominios α1 y α2 y los dominios H2-Kb α3 murino, transmembrana y citoplasmático (Vitiello et ál., 1991). Los ratones utilizados para estos estudios fueron progenie de la primera generación (F1) de hembras BALB/cAnNTac de tipo salvaje y machos HLA-A2.1 Tg homocigotos en la referencia C57Bl/6).

Inmunizaciones del vector de adenovirus (Ad5v)

[00795] Los ratones HLA-A2 Tg se inmunizaron con partículas $1x10^{10}$ a $1x10^{6}$ virales de vectores adenovirales mediante inyección intramuscular bilateral en el tibial anterior. Las respuestas inmunitarias se midieron 12 días después de la inmunización.

Aislamiento de linfocitos

[00796] Se aislaron linfocitos de bazos y nodos linfáticos recientemente recolectados de ratones inmunizados. Los tejidos se disociaron en RPMI con suero bovino fetal al 10 % con penicilina y estreptomicina (RPMI completo) mediante el uso del disociador de tejido GentleMACS de acuerdo con las instrucciones del fabricante.

Análisis immunospot ligado a enzimas (ELISPOT) ex vivo

[00797] El análisis ELISPOT se llevó a cabo de acuerdo con los lineamientos generales de armonización ELISPOT (Janetzki et ál., 2015) con el kit ELISpotPLUS de IFNg de ratón (MABTECH). Se incubaron $1x10^{5}$ esplenocitos con 10 uM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNg de 96 pocillos. Se desarrollaron manchas mediante el uso de fosfatasa alcalina. Se marca una duración de 10 minutos para la reacción y se inactiva la reacción al colocar la placa bajo agua del grifo. Se contaron las manchas mediante el uso de un AID vSpot Reader Spectrum. Para el análisis ELISPOT, los pocillos con una saturación > 50 % se registraron como «cantidad excesiva para contar». Las muestras con una desviación de los pocillos replicados > 10 % se excluyeron del análisis. Luego, se corrigieron los conteos de manchas con respecto a la confluencia de pocillos mediante el uso de la fórmula: conteo de manchas + 2 x (conteo de manchas x % de confluencia /[100 % - % de confluencia]). Se corrigió el fondo negativo mediante la eliminación de los conteos de manchas en los pocillos de estimulación peptídica de caracter negativo de los pocillos estimulados con antígenos. Finalmente, los pocillos etiquetados como «cantidad excesiva para contar» se ajustaron al valor corregido observado más alto, con redondeo hasta el múltiplo de cien más cercano.

Tinción de citosina intracelular (ICS) ex vivo y análisis de citometría de flujo

[00798] Los linfocitos aislados recientemente a una densidad de 2-5x10^6 células/mL se incubaron con 10 uM de los péptidos indicados durante 2 horas. Luego de dos horas, se agregó brefeldin A a una concentración de 5 ug/ml y las células se incubaron con un estimulante durante 4 horas adicionales. Luego de la estimulación, se etiquetaron células viables con un tinte de viabilidad ajustable eFluor780 de acuerdo con el protocolo del fabricante y se tiñeron con APC anti-CD8 (clon 53-6.7, BioLegend) a una dilución 1:400. Se utilizó anti-IFNg PE (clon XMG1.2, BioLegend) a 1:100 para la tinción intracelular. Las muestras se recolectaron en un citómetro de flujo Attune NxT (Thermo Scientific). Los datos de la citometría de flujo se graficaron y se analizaron mediante el uso de FlowJo. Para analizar el grado de respuesta específica al antígeno, se calculó tanto el porcentaje de IFNg+ de las células CD8+ como la cantidad de células IFNg+ total/1x10^6 de células vivas en respuesta a cada estimulante de péptidos.

XIV.B.2. Evaluación in vitro de diseños del casete de neoantígenos

[00799] Como ejemplo de la evaluación del diseño del casete de neoantígenos, se desarrolló un ensayo basado en células in vitro para evaluar si los epítopos humanos seleccionados dentro de los casetes de vacunas modelo fueron expresados, procesados y presentados por células que presentan antígenos (Figura 15). Luego del reconocimiento, los linfocitos T indicadores que fueron diseñados para expresar uno de cinco TCR específicos para combinaciones de péptido-HLA ampliamente caracterizadas se activan y traslocan el factor nuclear de linfocitos T activados (NFAT, por sus siglas en inglés) en el núcleo lo que conduce a la activación transcripcional de un gen indicador de luciferasa. La estimulación antigénica de las líneas de linfocitos T CD8 se cuantificó mediante bioluminiscencia.

[00800] Se modificaron las líneas indicadoras Jurkat-Lucia individuales mediante transducción lentiviral con una construcción de expresión que incluye una cadena de TCR beta y de TCR alfa específica para el antígeno separada por una secuencia de omisión ribosómica P2A para asegurar las cantidades equimolares del producto traducido (Banu et ál., 2014) La adición de un segundo elemento CD8 beta-P2A-CD8 alfa a la construcción lentiviral proporcionó la expresión del correceptor de CD8, que no se encuentra en la línea celular indicadora original, ya que el CD8 en la superficie celular es crucial para que la afinidad de unión se dirija a moléculas pMHC y potencie la señalización a través del acoplamiento de su cola citoplasmática (Lyons et ál., 2006; Yachi et ál., 2006).

[00801] Luego de la transducción lentiviral, los indicadores Jurkat-Lucia se expandieron en una selección de puromicina, sometida a clasificación de células activadas por fluorescencia (FACS, por sus siglas en inglés) y se evaluaron las poblaciones monoclonales para determinar la expresión de luciferasa. Esto proporcionó líneas celulares indicadoras transducidas de forma estable para los antígenos de péptidos específicos 1, 2, 4 y 5 con respuestas celulares funcionales. (Tabla 2).

Tabla 2: Desarrollo de un ensayo de activación de linfocitos T in vitro. El reconocimiento de linfocitos T

específicos de péptidos mediante inducción de luciferasa indica el procesamiento y presentación eficaz de los antígenos del casete de vacuna.

	Diseño del casete corto
Epítopo	**AAY**
1	24,5 ± 0,5
2	11,3 ± 0,4
3*	n/a
4	26,1 ± 3,1
5	46,3 ± 1,9

* Linfocito T indicador para el epítopo 3 no generado aún

[00802] En otro ejemplo, una serie de casetes cortos, todos los epítopos se incorporaron en la misma posición (Figura 16A) y solamente los enlazadores que separan los epítopos restringidos HLA-A*0201 (Figura 16B) fueron variados. Los linfocitos T indicadores se mezclaron individualmente con células que presentan antígenos (APC) U-87 que se infectaron con construcciones adenovirales que expresan casetes cortos, y la expresión de luciferasa se midió con respecto a controles no infectados. Los cuatro antígenos en los casetes modelo fueron reconocidos por linfocitos T indicadores correspondientes, lo que demuestra un procesamiento y una presentación eficaz de múltiples antígenos. La magnitud de las respuestas de linfocitos T siguió tendencias ampliamente similares para los enlazadores naturales y AAY. Los antígenos liberados del casete a base del enlazador RR muestra inducciones de luciferasa inferiores (Tabla 3). El enlazador de DPP, diseñado para interrumpir el procesamiento del antígeno, produjo un casete de vacuna que condujo a una presentación del epítopo baja (Tabla 3).

Tabla 3: Evaluación de las secuencias de enlace en casetes cortos. La inducción de luciferasa en el ensayo de activación de linfocitos T in vitro indicó que, aparte del casete basado en DPP, todos los enlazadores facilitaron una liberación eficaz de los antígenos del casete. El epítopo de linfocitos T solamente (sin enlazador) = 9AA, enlazador natural un solo lado = 17AA, enlazador natural ambos lados = 25AA, enlazadores no naturales = AAY, RR, DPP

	Diseños del casete corto					
Epítopo	**9AA**	**17AA**	**25AA**	**AAY**	**RR**	**DPP**
1	33,6 ± 0,9	42,8 ± 2,1	42,3 ± 2,3	24,5 ± 0,5	21,7 ± 0,9	0,9 ± 0,1
2	12,0 ± 0,9	10,3 ± 0,6	14,6 ± 04	11,3 ± 0,4	8,5 ± 0,3	1,1 ± 0,2

3*	n/a	n/a	n/a	n/a	n/a	n/a
4	26,6 ± 2,5	16,1 ± 0,6	16,6 ± 0,8	26,1 ± 3,1	12,5 ± 0,8	1,3 ± 0,2
5	29,7 ± 0,6	21,2 ± 0,7	24,3 ± 1,4	46,3 ± 1,9	19,7 ± 0,4	1,3 ± 0,1

* Linfocito T indicador para el epítopo 3 no generado aún

[00803] En otro ejemplo, se construyó una serie de casetes cortos que, aparte de epítopos humanos y de ratón, contuvieron secuencias de direccionamiento tales como ubiquitina (Ub), péptidos de señal de MHC e Ig-kappa (SP) y/o motivos transmembrana (TM) de MHC posicionados ya sea en el extremo N o C del casete. (Figura 17). Cuando se administran APC U-87 mediante un vector adenoviral, los linfocitos T indicadores demostraron nuevamente un procesamiento y presentación eficaces de múltiples antígenos derivados del casete. Sin embargo, la magnitud de respuestas de linfocitos T no se vio sustancialmente afectada por las diversas características de direccionamiento (Tabla 4).

Tabla 4: Evaluación de secuencias de direccionamiento celular agregadas a casetes de vacuna modelo. El empleo del ensayo de activación de linfocitos T in vitro demostró que los cuatro epítopos marcadores restringidos HLA-A*0201 se liberan de forma eficaz de los casetes modelo y las secuencias de direccionamiento no mejoraron sustancialmente el reconocimiento y activación de linfocitos T.

	Diseños del casete corto									
Epítopo	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
1	32,5 ± 1,5	31,8 ± 0,8	29,1 ± 1,2	29,1 ± 1,1	28,4 ± 0,7	20,4 ± 0,5	35,0 ± 1,3	30,3 ± 2,0	22,5 ± 0,9	38,1 ± 1,6
2	6,1 ± 0,2	6,3 ± 0,2	7,6 ± 0,4	7,0 ± 0,5	5,9 ± 0,2	3,7 ± 0,2	7,6 ± 0,4	5,4 ± 0,3	6,2 ± 0,4	6,4 ± 0,3
3*	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
4	12,3 ± 1,1	14,1 ± 0,7	12,2 ± 0,8	13,7 ± 1,0	11,7 ± 0,8	10,6 ± 0,4	11,0 ± 0,6	7,6 ± 0,6	16,1 ± 0,5	8,7 ± 0,5
5	44,4 ± 2,8	53,6 ± 1,6	49,9 ± 3,3	50,5 ± 2,8	41,7 ± 2,8	36,1 ± 1,1	46,5 ± 2,1	31,4 ± 0,6	75,4 ± 1,6	35,7 ± 2,2

* Linfocito T indicador para el epítopo 3 no generado aún

XIV.B.3. Evaluación in vivo de diseños del casete de neoantígenos

[00804] A modo de ejemplo adicional de la evaluación del diseño del casete de neoantígenos, los casetes de vacuna se diseñaron para contener 5 epítopos de MHC clase I humanos ampliamente

caracterizados conocidos por estimular los linfocitos T CD8 de una forma restringida a HLA-A*02:01 (Figura 16A, 17, 19A). Para la evaluación de su inmunogenicidad in vivo, los casetes vacuna que contienen estos epítopos marcadores se incorporaron en vectores adenovirales y se utilizaron para infectar ratones transgénicos HLA-A2 (Figura 18). Este modelo de ratón porta un transgen que consiste en parte en HLA-A*0201 y H2-Kb de ratón, por lo que codifica una molécula de MHC clase I quimérica que consiste en los dominios HLA-A2.1 líder, α1 y α2 ligados al dominio α3, transmembrana y citoplasmático de H2-Kb (Vitiello et ál., 1991). La molécula quimérica permite una presentación del antígeno restringida a HLA-A*02:01 mientras se mantiene la interacción emparejada con las especies del correceptor CD8 con el dominio α3 en el MHC.

[00805] Para los casetes cortos, todos los epítopos marcadores generaron una respuesta de linfocitos T, como se determina mediante ELISPOT de IFN-gamma, que fue aproximadamente 10-50 veces más potente de lo que se describe comúnmente (Cornet et ál., 2006; Depla et ál., 2008; Ishioka et ál., 1999). De todos los enlazadores evaluados, el concatémero de secuencias 25mer, cada uno con un epítopo mínimo flanqueado por sus secuencias de aminoácidos naturales, generaron la respuesta de linfocitos T más grande y amplia (Tabla 5). La tinción de citosinas intracelular (ICS) y el análisis de citometría de flujo reveló que las respuestas de linfocitos T específicos del antígeno se derivan de linfocitos T CD8.

Tabla 5: Evaluación in vivo de las secuencias de enlace en casetes cortos. Los datos de ELISPOT indicaron que los ratones transgénicos HLA-A2, 17 días luego de la infección con partículas virales de adenovirus 1e11, generaron una respuesta de linfocitos T a todos los epítopos restringidos de MHC clase I en el casete.

	Diseños del casete corto					
Epítopo	**9AA**	**17AA**	**25AA**	**AAY**	**RR**	**DPP**
1	2020+/-583	2505+/-1281	6844+/-956	1489+/-762	1675+/-690	1781+/-774
2	4472+/-755	3792+/-1319	7629+/-996	3851+/-1748	4726+/-1715	5868+/-1427
3	5830+/-315	3629+/-862	7253+/-491	4813+/-1761	6779+/-1033	7328+/-1700
4	5536+/-375	2446+/-955	2961+/-1487	4230+/-1759	6518+/-909	7222+/-1824
5	8800+/-0	7943+/-821	8423+/-442	8312+/-696	8800+/-0	1836+/-328

[00806] En otro ejemplo, se construyó una serie de casetes de vacuna y se incorporó en vectores adenovirales que, junto con los 5 epítopos marcadores originales, contuvieron 16 epítopos HLA-A*02:01, A*03:01 y B*44:05 adicionales con reactividad a linfocitos T CD8 conocida (Figura 19A, B). El tamaño de estos casetes largos imitó estrechamente el diseño del casete clínico final y solamente varió la posición de los epítopos con respecto uno del otro. Las respuestas de linfocitos T CD8 fue comparable en magnitud y amplitud para tanto casetes de vacuna largos como cortos, lo que demuestra que (a) la adición de más epítopos no tuvo sustancialmente un impacto en la magnitud de la respuesta inmunitaria al conjunto original de epítopos y (b) la posición de un epítopo en un casete no influenció sustancialmente la respuesta consiguiente de linfocitos T a este (Tabla 6).

Tabla 6: Evaluación in vivo del impacto de la posición del epítopo en casetes largos. Los datos de ELISPOT indicaron que los ratones transgénicos HLA-A2, 17 días luego de la infección con partículas virales de adenovirus 5e10, generaron una respuesta de linfocitos T comparable en magnitud para tanto casetes de vacuna largos como cortos.

	Diseños del casete largo		
Epítopo	**Estándar**	**Desalineado**	**Corto**
1	863+/-1080	804+/-1113	1871+/-2859
2	6425+/-1594	28+/-62	5390+/-1357
3*	23+/-30	36+/-18	0+/-48
4	2224+/-1074	2727+/-644	2637+/-1673
5	7952+/-297	8100+/-0	8100+/-0

* Se sospecha que un error técnico causó una ausencia de respuesta de linfocitos T

XIV.B.4. Diseño del casete de neoantígenos para estudios de inmunogenicidad y toxicológicos

[00807] En resumen, los hallazgos de las evaluaciones del casete modelo (Figura 16-19, Tablas 2-6) demostró que, para los casetes de vacuna modelo, se logró una inmunogenicidad potente cuando se empleó un enfoque de «tira de cuentas» que codifica alrededor de 20 epítopos en el contexto de un vector a base de adenovirus. Los epítopos se ensamblaron mejor mediante la concatenación de secuencias 25mer, cada una integra un epítopo de linfocito T CD8 mínimo (por ejemplo, 9 residuos de aminoácidos) que fueron flanqueados en ambos lados por su secuencia de péptidos natural circundante (por ejemplo, 8 residuos de aminoácidos en cada lado). Como se utiliza en la presente, una secuencia flanqueante «natural» o «nativa» hace referencia a la secuencia flanqueante en el extremo N y/o C de un epítopo dado en el contexto natural de ese epítopo dentro de su proteína de fuente. Por ejemplo, el epítopo de MHC I HCMV pp65 NLVPMVATV (SEQ ID NO: 132) está flanqueado en su extremo 5' por la secuencia 5' WQAGILAR (SEQ ID NO: 139) en su 3' por la secuencia 3' natural QGQNLKYQ (SEQ ID

NO: 140), por lo que genera el péptido 25mer WQAGILARNLVPMVATVQGQNLKYQ (SEQ ID NO: 141) encontrado dentro de la proteína fuente HCMV pp65. La secuencia natural también puede hacer referencia a una secuencia de nucleótidos que codifica un epítopo flanqueado por una o más secuencias flanqueantes naturales. Cada secuencia 25mer se conecta directamente a la siguiente secuencia 25mer. En los casos donde el epítopo de linfocito T CD8 mínimo es mayor que o menor que 9 aminoácidos, la longitud del péptido flanqueante puede ajustarse de forma tal que la longitud total sea aún una secuencia de péptidos 25mer. Por ejemplo, un epítopo de linfocito T CD8 de 10 aminoácidos puede estar flanqueada por una secuencia de 8 aminoácidos y una de 7 aminoácidos. El concatémero fue seguido por dos epítopos de MHC clase II universales que fueron incluidos para estimular los linfocitos T CD4 auxiliares y mejorar la inmunogenicidad in vivo general de los antígenos del casete de vacuna. (Alexander et ál., 1994; Panina-Bordignon et ál., 1989). Los epítopos de clase II se enlazaron al final de los epítopos de clase I mediante un enlazador de aminoácidos GPGPG (SEQ ID NO:56). Los dos epítopos de clase II también se enlazaron entre sí mediante un enlazador de aminoácidos GPGPG (SEQ ID NO: 56), así como también flanqueados en el extremo C por un enlazador de aminoácido GPGPG (SEQ ID NO: 56). Ni la posición ni la cantidad de epítopos parecieron tener un impacto sustancial en el reconocimiento o la respuesta de linfocitos T. Las secuencias de direccionamiento tampoco parecieron afectar sustancialmente la inmunogenicidad de los antígenos derivados del casete.

[00808] A modo de ejemplo adicional, en función de datos in vitro e in vivo obtenidos con casetes modelo (Figura 16-19, Tabla 2-6), se generó un diseño de casete que alterna epítopos caracterizados ampliamente de linfocitos T conocidos por ser inmunogénicos en primates no humanos (NHP), ratones y seres humanos. Los 20 epítopos, todos integrados en sus secuencias 25mer naturales, son seguidos de dos epítopos de MHC clase II universales que estuvieron presentes en todos los casetes modelo evaluados (Figura 20). Este diseño de casete se utilizó para estudiar la inmunogenicidad, así como también los estudios de farmacología y toxicología en múltiples especies.

XV. Vector de administración del casete de neoantígenos de ChAd

XV.A. Construcción del vector de administración del casete de neoantígenos de ChAd

[00809] En un ejemplo, se diseñó adenovirus de chimpancé (ChAd) para constituir un vector de administración para casetes de neoantígenos. En un ejemplo adicional, se sintetizó un vector ChAdV68 de longitud completa en función de AC_000011.1 (secuencia 2 de la patente US 6083716) con las secuencias E1 (nt 457 a 3014) y E3 (nt 27.816- 31.332) eliminadas. Los genes indicadores bajo control del promotor/potenciador de CMV se insertaron en lugar de las secuencias E1 eliminadas. La transfección de este clon en células HEK293 no proporcionó un virus infeccioso. Para confirmar la secuencia del virus C68 de tipo salvaje, se obtuvo un aislado VR-594 del ATCC, se sometió a pasaje y luego se sometió a secuenciación independiente (SEQ ID NO:10). Cuando se compara la secuencia

AC_000011.1 con la secuencia de ATCC VR-594 (SEQ ID NO:10) del virus ChAdV68 de tipo salvaje, se identificaron 6 diferencias de nucleótidos. En un ejemplo, se generó un vector ChAdV68 modificado en función de AC_000011.1, con los nucleótidos de ATCC VR-594 correspondientes sustituidos en cinco posiciones (ChAdV68.5WTnt SEQ ID NO:1).

[00810] En otro ejemplo, se generó un vector ChAdV68 modificado en función de AC_000011.1 con las secuencias E1 (nt 577 a 3403) y E3 eliminadas y los nucleótidos ATCC VR-594 correspondientes sustituidos en cuatro posiciones. Una GFP indicadora (ChAdV68.4WTnt.GFP; SEQ ID NO:11) o casete de neoantígeno modelo (ChAdV68.4WTnt.MAG25mer; SEQ ID NO:12) bajo control del promotor/potenciador de CMV se insertó en lugar de las secuencias E1 eliminadas.

[00811] En otro ejemplo, se generó un vector ChAdV68 modificado en función de AC_000011.1 con las secuencias E1 (nt 577 a 3403) y E3 (nt 27.125- 31.825) eliminadas y los nucleótidos ATCC VR-594 correspondientes sustituidos en cinco posiciones. Una GFP indicadora (ChAdV68.5WTnt.GFP; SEQ ID NO:13) o casete de neoantígeno modelo (ChAdV68.5WTnt.MAG25mer; SEQ ID NO:2) bajo control del promotor/potenciador de CMV se insertó en lugar de las secuencias E1 eliminadas.

Secuencia de ChAdVC68 de cadena completa “ChAdV68.5WTnt” (SEQ ID NO:1); secuencia AC_000011.1 con nucleótidos de ATCC VR-594 correspondientes sustituidos en cinco posiciones.
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTG GGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGT TTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAAC GAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTT TCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAAT CTGAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGA TTACGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGT GTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCC ACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAA GATGAGGCACCTGAGAGACCTGCCCGATGAGAAAATCATCATCGCTTCCGGGAACGAGATTCT GGAACTGGTGGTAAATGCCATGATGGGCGACGACCCTCCGGAGCCCCCCACCCCATTTGAGAC ACCTTCGCTGCACGATTTGTATGATCTGGAGGTGGATGTGCCCGAGGACGATCCCAATGAGGA GGCGGTAAATGATTTTTTTAGCGATGCCGCGCTGCTAGCTGCCGAGGAGGCTTCGAGCTCTAG CTCAGACAGCGACTCTTCACTGCATACCCCTAGACCCGGCAGAGGTGAGAAAAAGATCCCCGA GCTTAAAGGGGAAGAGATGGACTTGCGCTGCTATGAGGAATGCTTGCCCCCGAGCGATGATGA

GGACGAGCAGGCGATCCAGAACGCAGCGAGCCAGGGAGTGCAAGCCGCCAGCGAGAGCTTTG
CGCTGGACTGCCCGCCTCTGCCCGGACACGGCTGTAAGTCTTGTGAATTTCATCGCATGAATAC
TGGAGATAAAGCTGTGTTGTGTGCACTTTGCTATATGAGAGCTTACAACCATTGTGTTTACAGTA
AGTGTGATTAAGTTGAACTTTAGAGGGAGGCAGAGAGCAGGGTGACTGGGCGATGACTGGTTT
ATTTATGTATATATGTTCTTTATATAGGTCCCGTCTCTGACGCAGATGATGAGACCCCCACTACA
AAGTCCACTTCGTCACCCCCAGAAATTGGCACATCTCCACCTGAGAATATTGTTAGACCAGTTCC
TGTTAGAGCCACTGGGAGGAGAGCAGCTGTGGAATGTTTGGATGACTTGCTACAGGGTGGGGT
TGAACCTTTGGACTTGTGTACCCGGAAACGCCCCAGGCACTAAGTGCCACACATGTGTGTTTAC
TTGAGGTGATGTCAGTATTTATAGGGTGTGGAGTGCAATAAAAAATGTGTTGACTTTAAGTGCGT
GGTTTATGACTCAGGGGTGGGGACTGTGAGTATATAAGCAGGTGCAGACCTGTGTGGTTAGCT
CAGAGCGGCATGGAGATTTGGACGGTCTTGGAAGACTTTCACAAGACTAGACAGCTGCTAGAG
AACGCCTCGAACGGAGTCTCTTACCTGTGGAGATTCTGCTTCGGTGGCGACCTAGCTAGGCTA
GTCTACAGGGCCAAACAGGATTATAGTGAACAATTTGAGGTTATTTTGAGAGAGTGTTCTGGTCT
TTTTGACGCTCTTAACTTGGGCCATCAGTCTCACTTTAACCAGAGGATTTCGAGAGCCCTTGATT
TTACTACTCCTGGCAGAACCACTGCAGCAGTAGCCTTTTTTGCTTTTATTCTTGACAAATGGAGT
CAAGAAACCCATTTCAGCAGGGATTACCAGCTGGATTTCTTAGCAGTAGCTTTGTGGAGAACAT
GGAAGTGCCAGCGCCTGAATGCAATCTCCGGCTACTTGCCGGTACAGCCGCTAGACACTCTGA
GGATCCTGAATCTCCAGGAGAGTCCCAGGGCACGCCAACGTCGCCAGCAGCAGCAGCAGGAG
GAGGATCAAGAAGAGAACCCGAGAGCCGGCCTGGACCCTCCGGCGGAGGAGGAGGAGTAGCT
GACCTGTTTCCTGAACTGCGCCGGGTGCTGACTAGGTCTTCGAGTGGTCGGGAGAGGGGGATT
AAGCGGGAGAGGCATGATGAGACTAATCACAGAACTGAACTGACTGTGGGTCTGATGAGTCGC
AAGCGCCCAGAAACAGTGTGGTGGCATGAGGTGCAGTCGACTGGCACAGATGAGGTGTCGGT
GATGCATGAGAGGTTTTCTCTAGAACAAGTCAAGACTTGTTGGTTAGAGCCTGAGGATGATTGG
GAGGTAGCCATCAGGAATTATGCCAAGCTGGCTCTGAGGCCAGACAAGAAGTACAAGATTACTA
AGCTGATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTGAAGTGGAGATCTGTCT
CCAGGAAAGGGTGGCTTTCAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGGGCATGGA
TGGGGTTACCTTTATGAACATGAGGTTCAGGGGAGATGGGTATAATGGCACGGTCTTTATGGCC
AATACCAAGCTGACAGTCCATGGCTGCTCCTTCTTTGGGTTTAATAACACCTGCATCGAGGCCT
GGGGTCAGGTCGGTGTGAGGGGCTGCAGTTTTTCAGCCAACTGGATGGGGGTCGTGGGCAGG
ACCAAGAGTATGCTGTCCGTGAAGAAATGCTTGTTTGAGAGGTGCCACCTGGGGGTGATGAGC
GAGGGCGAAGCCAGAATCCGCCACTGCGCCTCTACCGAGACGGGCTGCTTTGTGCTGTGCAAG
GGCAATGCTAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCAGATG
CTGACCTGCGCCGGCGGGAACAGCCATATGCTGGCCACCGTACATGTGGCTTCCCATGCTCGC
AAGCCCTGGCCCGAGTTCGAGCACAATGTCATGACCAGGTGCAATATGCATCTGGGGTCCCGC

CGAGGCATGTTCATGCCCTACCAGTGCAACCTGAATTATGTGAAGGTGCTGCTGGAGCCCGAT
GCCATGTCCAGAGTGAGCCTGACGGGGGTGTTTGACATGAATGTGGAGGTGTGGAAGATTCTG
AGATATGATGAATCCAAGACCAGGTGCCGAGCCTGCGAGTGCGGAGGGAAGCATGCCAGGTTC
CAGCCCGTGTGTGTGGATGTGACGGAGGACCTGCGACCCGATCATTTGGTGTTGCCCTGCACC
GGGACGGAGTTCGGTTCCAGCGGGGAAGAATCTGACTAGAGTGAGTAGTGTTCTGGGGCGGG
GGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATG
AGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCC
TGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGC
GAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCA
GCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCAC
TCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCT
GCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTC
AGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAAT
AAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGC
GGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGT
AGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTC
CATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGG
GCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAG
GTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCC
TGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACC
ACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGA
AAGAATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGA
TGGGCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGG
TCCTGGGTGAGGTCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGG
ACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGA
GCTCGGAGGGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGG
GAGATGAGCTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCC
GTAGATGACCCCGATGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCC
GGAGGAGGGGGGCCACCTCGTTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCG
CCAGGAGGCGCTCTCCCCCCAGGGATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGC
TTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAG
AGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGAC
GGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAG
GGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGG

CGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCC
TGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCT
TTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGC
GTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCG
CAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTC
CCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGA
CAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGC
GGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACG
AAGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGT
ATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACG
TGACCGGGGGTCCCGGCCGGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGTC
TTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCAT
GACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGC
GGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGC
TTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTG
GTTTTTTTCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCAC
GCACTTCCATTCGGGGAAGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCG
ATTATGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCA
GCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCG
GGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGA
AGTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACGGCCAGCGCGCGCTCGTAGGGAC
TGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCG
TAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGAT
GCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCAGGTTG
GTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGA
GATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGA
AGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCG
CAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGCTCGC
GGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCTG
CACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCA
CGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTC
CCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCA
GAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAA
GAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGG

CCCGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCA
CGATGTAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGT
AGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGG
GTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGT
ACTGACGGAACTGCTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGG
TCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCG
GTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCAT
CCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGAT
GGGGAAGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAA
ATGCCGACGGCGCGCCGAACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACG
CTGCACGGGATGCACGTGCTGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGG
GAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTC
TTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGC
GAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACG
CTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGG
CGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTCGATGGCTT
GCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGC
GACGGGGGCGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAG
GGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCGGGTAG
GTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCTGGA
TCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCGACAG
AATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTTGT
CCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGG
CGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTC
ATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCAT
GACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCT
GGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGC
GGAGCGGCATCTCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCA
CGGCGAAGTTGAAAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGA
TGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCC
TCTTCTTCCTCCTCCACTAACATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAG
GGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCC
GCGCCGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAG
ACGCCGCCGCGCATCTCCAGGTGGCCGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGA

CGATGCATCTTATCAATTGCCCCGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCA
CGGGATCTGAAAACCGCTGAACGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCA
CGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCGGGCGATGCTGCTGGTGATGAAGT
TGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCT
TGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTT
GTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGC
GCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGC
GAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTG
GTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTG
GCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGT
CGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAG
AGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGT
AGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAAC
TCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTG
GCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGC
TCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAAT
CTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGC
ACCAACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACT
AGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCA
GGGTTGCGTTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTG
GCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCT
CTTTTGTTTTGTTTGTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTC
CACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAG
CCACGACCGCCGCGGCCGCCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTG
GAAGAGGGCGAGGGGCTGGCGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTG
CAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGC
GGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGTTCCACGCGGGGCGGGAGCTGCGGCGC
GGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGAT
CAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTG
AAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCACCCTGATCGCGCGCGAG
GAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCCAC
CAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTT
CAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACA
TTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATC

AACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGTACGTGC
CCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCCT
GAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGGC
GGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGG
GACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGG
CCTTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGG
CGAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTC
CTGATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGAT
TGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAG
CAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCC
CACGCACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCG
ACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAAC
GTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGC
GGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGC
CCGCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATG
GTGACCGAGGTGCCCCAGAGCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGT
CGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGG
CGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGC
CTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCAACCGCAACTCGTACCTG
GGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTAC
CAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCAC
CCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGA
GGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGTTCCTGATGCAGGAGGGGG
CCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGC
AACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGGCCGCCATGAACTCTGACT
ATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGGGGTTCTACACGGGCGAGT
ACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGACAGCAGCGTGTTCTCCC
CCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCCCGTCC
TCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCC
CGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCG
CGCTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAAC
TTCCCCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAG
GAGCACAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTA
AACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGAC

AGCAGCGTGTTGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGG
CGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGT
TCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTC
CCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGA
GGCTCCTTACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTC
GGAGCTGGCACCCTTGTACGATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACAT
CGCCTCGCTGAACTACCAGAACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGA
CTTCACCCCCACGGAGGCCAGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCG
GCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAA
GTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTA
TGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTC
TCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGG
CGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCT
GGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATC
CCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGC
TGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGG
GGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAAAGCAAGGAGGATGCAGCA
GCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGGGGCGATAATTTTGCAAGC
GCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAGCCGGT
GGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTACCG
CAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCT
CACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCA
AGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCT
CCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGC
CTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCC
GCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCG
CTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTG
CCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTA
AATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGAT
GTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCG
CTCCCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGAC
CAGGTGGTGGCCGACGCGCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGC
CGTCATCGACAGCGTGGTGGCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGG
CGCATCGCCCGGCGGCACCGGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCA

GGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGC
CAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATG
TCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGC
CCGTGCGCACCCGCCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGC
GGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGAT
CTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGA
CAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCG
GCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTG
GTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGG
GGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCA
GCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCG
AGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAA
GCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGG
AAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCC
ATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGGAGCC
CATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGG
ATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTG
CTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGC
TTCTACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCG
CACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCAC
CTCTGACCCTGCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCtGCTTTGC
AGATCAATGGCCCTCACATGCCGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCG
CGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCA
TCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATC
GGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTT
GGAAACATCTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGAC
AGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCA
CCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGA
GCGGGCTTAAGAATTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCA
CAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGG
CTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAA
CAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAG
CTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCT
GACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGC

GGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACTTG
CCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGC
CCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCA
TCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTT
AACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGT
GAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTAC
ATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGC
GCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCAC
GATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGA
CAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACAT
GGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTC
CGGCACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAA
AGCCGATGGTGAAACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATT
AACATCACAAAAGATGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAA
AACCTATCAGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAA
AAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAA
GCCTACTAATAAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATAT
GACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTG
TTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGCACA
GATGACAGCAGCTCTTCTATTAATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACATTG
GTTTCAGAGACAACTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGC
CGGTCAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTAC
CAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGG
ACAGCTATGATCCTGATGTGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTAT
TGTTTCCCTCTGGATGCTGTTGGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGA
TCAAACCACATGGACCAAAGATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCA
TTCGCCATGGAAATCAACATCCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCC
TGTACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACAC
CTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGG
GGCGCGCTGGTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCGGG
GCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCC
CCAGAAATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAAC
TTCCGCAAGGACGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGG
GCCTCCATCTCCTTCACCAGCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGG

CCTCCACGCTCGAGGCCATGCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCT
CGGCGGCCAACATGCTCTACCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCT
CGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCC
TCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGC
ACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGC
CCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAG
GGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCAC
TACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCT
TCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGG
CCGTCACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCGGCTACCTCGCGCCCACCATGC
GCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACC
AGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACT
TCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACG
CGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAA
GTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCG
CACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCG
GGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCT
GGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGT
CAACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGC
TCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACC
AGTTCGAGTACGAGGGCCTGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACC
CTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTG
CATGTTCCTGCACGCCTTCGTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAA
CTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCC
GCAACCAGGAGGCGCTCTACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCG
CGCGCATCGAGAAGGCCACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTAT
GTTAAATGTCTTTAATAAACAGCACTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAAT
CGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTAC
TTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTC
GGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAAT
CGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGGAAC
ACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTCGCGTCGGTGATGCTCTCCACGTCG
AGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGC
ACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCG

GCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCC
TTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCG
GCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCG
GTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGC
CACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTG
CCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTG
GGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGG
TCTTGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGA
TGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCA
CGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGG
GCAGGCTCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCT
CGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGTAGCTGA
AGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCT
GCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTG
GAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGA
GGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCG
CCGCGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGC
GCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGG
AGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGA
ATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATG
CAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGA
GGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGA
GAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGGG
AGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACC
GCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGC
CGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTC
TACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGA
TCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGC
GCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGC
GACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGC
CCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGA
CCCATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACCAGG
TGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGC
AAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAG

TTTGGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCT
GCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTT
CAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTC
CTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCG
GGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGA
CGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTC
CTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCT
GGCCGACCTCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTAT
GAGCCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCC
ACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCT
GTGGAGCCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGA
GGACGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACC
GCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGC
AAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTG
TGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCT
ACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCG
ATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCG
GGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCG
AGGAAACAAGAAGCTGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGA
ACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGGGACAGCACTCAGGCAGAGGAG
GACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGAGGAGGTGGAAGAAGC
AGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCAAGCAGCACGGATACCATCTCCG
CTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAAC
CCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAAAAA
CGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCT
CTTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCC
TACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAA
TCCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGG
AGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGC
AGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAG
CGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCT
CACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGT
GCCCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTG
GAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGA

ATTGGCTCAGCGCCGGGCCCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACC
AGATACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTG
GCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGC
CCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTC
GTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAAC
GACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCG
GGGAGATCTTCCTTCACGCCTCGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCC
CGCTCGGGTGGCATCGGCACTCTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAAC
CCCTTCTCCGGCTCCCCCGGCCACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGC
GAGTCGGTGGACGGCTACGATTGAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGA
CACCTGGACCACTGCCGCCGCTTCCGCTGCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTT
GAGCTGCCCGAGGAGCACCCTCAGGGCCCGGCCCACGGAGTGCGGATCGTCGTCGAAGGGG
GCCTCGACTCCCACCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGGTCGAGCGCGAGCAAG
GACAGACCCTTCTGACTCTGTACTGCATCTGCAACCACCCCGGCCTGCATGAAAGTCTTTGTTG
TCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGTTCCT
GAATCCATCAACCAGTCTTTGTTCTTCACCGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGC
CCCACAAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCG
ACAACGACGGAGTCCTGCTGAGCGGCCCTGCCAACCTTACTTTTTCCACCCGCAGAAGCAAGC
TCCAGCTCTTCCAACCCTTCCTCCCCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACAC
CTTCCACCTGATCCCGAATACCACAGCGTCGCTCCCCGCTACTAACAACCAAACTAACCTCCAC
CAACGCCACCGTCGCGACCTTTCTGAATCTAATACTACCACCCACACCGGAGGTGAGCTCCGA
GGTCAACCAACCTCTGGGATTTACTACGGCCCCTGGGAGGTGGTTGGGTTAATAGCGCTAGGC
CTAGTTGCGGGTGGGCTTTTGGTTCTCTGCTACCTATACCTCCCTTGCTGTTCGTACTTAGTGGT
GCTGTGTTGCTGGTTTAAGAAATGGGGAAGATCACCCTAGTGAGCTGCGGTGCGCTGGTGGCG
GTGTTGCTTTCGATTGTGGGACTGGGCGGTGCGGCTGTAGTGAAGGAGAAGGCCGATCCCTGC
TTGCATTTCAATCCCAACAAATGCCAGCTGAGTTTTCAGCCCGATGGCAATCGGTGCGCGGTAC
TGATCAAGTGCGGATGGGAATGCGAGAACGTGAGAATCGAGTACAATAACAAGACTCGGAACAA
TACTCTCGCGTCCGTGTGGCAGCCCGGGGACCCCGAGTGGTACACCGTCTCTGTCCCCGGTG
CTGACGGCTCCCCGCGCACCGTGAATAATACTTTCATTTTTGCGCACATGTGCGACACGGTCAT
GTGGATGAGCAAGCAGTACGATATGTGGCCCCCCACGAAGGAGAACATCGTGGTCTTCTCCAT
CGCTTACAGCCTGTGCACGGCGCTAATCACCGCTATCGTGTGCCTGAGCATTCACATGCTCATC
GCTATTCGCCCCAGAAATAATGCCGAAAAAGAAAAACAGCCATAACGTTTTTTTTCACACCTTTTT
CAGACCATGGCCTCTGTTAAATTTTTGCTTTTATTTGCCAGTCTCATTGCCGTCATTCATGGAATG
AGTAATGAGAAAATTACTATTTACACTGGCACTAATCACACATTGAAAGGTCCAGAAAAAGCCAC

AGAAGTTTCATGGTATTGTTATTTTAATGAATCAGATGTATCTACTGAACTCTGTGGAAACAATAA
CAAAAAAAATGAGAGCATTACTCTCATCAAGTTTCAATGTGGATCTGACTTAACCCTAATTAACAT
CACTAGAGACTATGTAGGTATGTATTATGGAACTACAGCAGGCATTTCGGACATGGAATTTTATC
AAGTTTCTGTGTCTGAACCCACCACGCCTAGAATGACCACAACCACAAAAACTACACCTGTTACC
ACTATGCAGCTCACTACCAATAACATTTTTGCCATGCGTCAAATGGTCAACAATAGCACTCAACC
CACCCCACCCAGTGAGGAAATTCCCAAATCCATGATTGGCATTATTGTTGCTGTAGTGGTGTGC
ATGTTGATCATCGCCTTGTGCATGGTGTACTATGCCTTCTGCTACAGAAAGCACAGACTGAACG
ACAAGCTGGAACACTTACTAAGTGTTGAATTTTAATTTTTTAGAACCATGAAGATCCTAGGCCTTT
TAATTTTTTCTATCATTACCTCTGCTCTATGCAATTCTGACAATGAGGACGTTACTGTCGTTGTCG
GATCAAATTATACACTGAAAGGTCCAGCGAAGGGTATGCTTTCGTGGTATTGCTATTTTGGATCT
GACACTACAGAAACTGAATTATGCAATCTTAAGAATGGCAAAATTCAAAATTCTAAAATTAACAAT
TATATATGCAATGGTACTGATCTGATACTCCTCAATATCACGAAATCATATGCTGGCAGTTACAC
CTGCCCTGGAGATGATGCTGACAGTATGATTTTTTACAAAGTAACTGTTGTTGATCCCACTACTC
CACCTCCACCCACCACAACTACTCACACCACACACACAGATCAAACCGCAGCAGAGGAGGCAG
CAAAGTTAGCCTTGCAGGTCCAAGACAGTTCATTTGTTGGCATTACCCCTACACCTGATCAGCG
GTGTCCGGGGCTGCTAGTCAGCGGCATTGTCGGTGTGCTTTCGGGATTAGCAGTCATAATCATC
TGCATGTTCATTTTTGCTTGCTGCTATAGAAGGCTTTACCGACAAAAATCAGACCCACTGCTGAA
CCTCTATGTTTAATTTTTTCCAGAGTCATGAAGGCAGTTAGCGCTCTAGTTTTTTGTTCTTTGATT
GGCATTGTTTTTTGCAATCCTATTCCTAAAGTTAGCTTTATTAAAGATGTGAATGTTACTGAGGGG
GGCAATGTGACACTGGTAGGTGTAGAGGGTGCTGAAAACACCACCTGGACAAAATACCACCTCA
ATGGGTGGAAAGATATTTGCAATTGGAGTGTATTAGTTTATACATGTGAGGGAGTTAATCTTACC
ATTGTCAATGCCACCTCAGCTCAAAATGGTAGAATTCAAGGACAAAGTGTCAGTGTATCTAATGG
GTATTTTACCCAACATACTTTTATCTATGACGTTAAAGTCATACCACTGCCTACGCCTAGCCCAC
CTAGCACTACCACACAGACAACCCACACTACACAGACAACCACATACAGTACATTAAATCAGCCT
ACCACCACTACAGCAGCAGAGGTTGCCAGCTCGTCTGGGGTCCGAGTGGCATTTTTGATGTGG
GCCCCATCTAGCAGTCCCACTGCTAGTACCAATGAGCAGACTACTGAATTTTTGTCCACTGTCG
AGAGCCACACCACAGCTACCTCCAGTGCCTTCTCTAGCACCGCCAATCTCTCCTCGCTTTCCTC
TACACCAATCAGTCCCGCTACTACTCCTAGCCCCGCTCCTCTTCCCACTCCCCTGAAGCAAACA
GACGGCGGCATGCAATGGCAGATCACCCTGCTCATTGTGATCGGGTTGGTCATCCTGGCCGTG
TTGCTCTACTACATCTTCTGCCGCCGCATTCCCAACGCGCACCGCAAGCCGGTCTACAAGCCCA
TCATTGTCGGGCAGCCGGAGCCGCTTCAGGTGGAAGGGGGTCTAAGGAATCTTCTCTTCTCTTT
TACAGTATGGTGATTGAACTATGATTCCTAGACAATTCTTGATCACTATTCTTATCTGCCTCCTCC
AAGTCTGTGCCACCCTCGCTCTGGTGGCCAACGCCAGTCCAGACTGTATTGGGCCCTTCGCCT
CCTACGTGCTCTTTGCCTTCACCACCTGCATCTGCTGCTGTAGCATAGTCTGCCTGCTTATCACC

TTCTTCCAGTTCATTGACTGGATCTTTGTGCGCATCGCCTACCTGCGCCACCACCCCCAGTACC
GCGACCAGCGAGTGGCGCGGCTGCTCAGGCTCCTCTGATAAGCATGCGGGCTCTGCTACTTCT
CGCGCTTCTGCTGTTAGTGCTCCCCCGTCCCGTCGACCCCCGGTCCCCCACCCAGTCCCCCGA
GGAGGTCCGCAAATGCAAATTCCAAGAACCCTGGAAATTCCTCAAATGCTACCGCCAAAAATCA
GACATGCATCCCAGCTGGATCATGATCATTGGGATCGTGAACATTCTGGCCTGCACCCTCATCT
CCTTTGTGATTTACCCCTGCTTTGACTTTGGTTGGAACTCGCCAGAGGCGCTCTATCTCCCGCC
TGAACCTGACACACCACCACAGCAACCTCAGGCACACGCACTACCACCACTACAGCCTAGGCC
ACAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCTATTAGT
TACTTCAATCTAACCGGCGGAGATGACTGACCCACTGGCCAACAACAACGTCAACGACCTTCTC
CTGGACATGGACGGCCGCGCCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCAGCA
GGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCT
TCTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTA
CGAGCTCCTGCAGCAGCGCCAGAAGTTCACCTGCCTGGTCGGAGTCAACCCCATCGTCATCAC
CCAGCAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCA
CACTCTGATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCACCCCCTTAT
CCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAAT
AAAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTG
ATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTAC
TGCAGGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGT
CCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTC
GACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTC
GTCTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCC
GTCACCACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCG
GGAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCA
TTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCAC
CATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGA
CTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACAT
AAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCT
GGGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTT
GGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTG
GCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTT
GTGGACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACAC
TTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGA
AACCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACG

GTGTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATA
GATGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACA
AAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTAT
GCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATA
CACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACA
TCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAA
CAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATT
CGAGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAG
CCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCA
GAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCAC
CTCACAGCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCA
GAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCC
CCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGG
ACTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAG
CGCATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTGTTCAAC
AGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCG
TCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGA
TCTCCTTGGGCATGTGGCGGTTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCA
GCCCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAG
ACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAG
CTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTC
CTCGGGGGTCAAAACCATATCCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAG
AACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCAC
CGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCG
GCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTG
CTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCG
GCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACC
GTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACA
TCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACG
GCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTACTT
CAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCA
GGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCA
CATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTACAC
TCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCcTGAGGT

AAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCAC ACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATATCA AAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTC TCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAAC CGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCG GTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGA AAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCG TTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCT GGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGA CCCTCGTAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGT GAATGATTCGACAAGATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAGCGCC CGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGAT GAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGC CCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACA CAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATAT ATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACAC GCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCC CAAAACTGCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCG ACCGTTAAAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCC CGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATT GATGATGG
C68 de ATCC VR-594 (SEQ ID NO:10); secuenciado de forma independiente; C68 de longitud completa
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTG GGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGT TTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAAC GAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTT TCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAAT CTGAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGA TTACGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGT GTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCC ACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAA GATGAGGCACCTGAGAGACCTGCCCGATGAGAAAATCATCATCGCTTCCGGGAACGAGATTCT GGAACTGGTGGTAAATGCCATGATGGGCGACGACCCTCCGGAGCCCCCCACCCCATTTGAGAC

ACCTTCGCTGCACGATTTGTATGATCTGGAGGTGGATGTGCCCGAGGACGATCCCAATGAGGA
GGCGGTAAATGATTTTTTTAGCGATGCCGCGCTGCTAGCTGCCGAGGAGGCTTCGAGCTCTAG
CTCAGACAGCGACTCTTCACTGCATACCCCTAGACCCGGCAGAGGTGAGAAAAAGATCCCCGA
GCTTAAAGGGGAAGAGATGGACTTGCGCTGCTATGAGGAATGCTTGCCCCCGAGCGATGATGA
GGACGAGCAGGCGATCCAGAACGCAGCGAGCCAGGGAGTGCAAGCCGCCAGCGAGAGCTTTG
CGCTGGACTGCCCGCCTCTGCCCGGACACGGCTGTAAGTCTTGTGAATTTCATCGCATGAATAC
TGGAGATAAAGCTGTGTTGTGTGCACTTTGCTATATGAGAGCTTACAACCATTGTGTTTACAGTA
AGTGTGATTAAGTTGAACTTTAGAGGGAGGCAGAGAGCAGGGTGACTGGGCGATGACTGGTTT
ATTTATGTATATATGTTCTTTATATAGGTCCCGTCTCTGACGCAGATGATGAGACCCCCACTACA
AAGTCCACTTCGTCACCCCCAGAAATTGGCACATCTCCACCTGAGAATATTGTTAGACCAGTTCC
TGTTAGAGCCACTGGGAGGAGAGCAGCTGTGGAATGTTTGGATGACTTGCTACAGGGTGGGGT
TGAACCTTTGGACTTGTGTACCCGGAAACGCCCCAGGCACTAAGTGCCACACATGTGTGTTTAC
TTGAGGTGATGTCAGTATTTATAGGGTGTGGAGTGCAATAAAAAATGTGTTGACTTTAAGTGCGT
GGTTTATGACTCAGGGGTGGGGACTGTGAGTATATAAGCAGGTGCAGACCTGTGTGGTTAGCT
CAGAGCGGCATGGAGATTTGGACGGTCTTGGAAGACTTTCACAAGACTAGACAGCTGCTAGAG
AACGCCTCGAACGGAGTCTCTTACCTGTGGAGATTCTGCTTCGGTGGCGACCTAGCTAGGCTA
GTCTACAGGGCCAAACAGGATTATAGTGAACAATTTGAGGTTATTTTGAGAGAGTGTTCTGGTCT
TTTTGACGCTCTTAACTTGGGCCATCAGTCTCACTTTAACCAGAGGATTTCGAGAGCCCTTGATT
TTACTACTCCTGGCAGAACCACTGCAGCAGTAGCCTTTTTTGCTTTTATTCTTGACAAATGGAGT
CAAGAAACCCATTTCAGCAGGGATTACCAGCTGGATTTCTTAGCAGTAGCTTTGTGGAGAACAT
GGAAGTGCCAGCGCCTGAATGCAATCTCCGGCTACTTGCCGGTACAGCCGCTAGACACTCTGA
GGATCCTGAATCTCCAGGAGAGTCCCAGGGCACGCCAACGTCGCCAGCAGCAGCAGCAGGAG
GAGGATCAAGAAGAGAACCCGAGAGCCGGCCTGGACCCTCCGGCGGAGGAGGAGGAGTAGCT
GACCTGTTTCCTGAACTGCGCCGGGTGCTGACTAGGTCTTCGAGTGGTCGGGAGAGGGGGATT
AAGCGGGAGAGGCATGATGAGACTAATCACAGAACTGAACTGACTGTGGGTCTGATGAGTCGC
AAGCGCCCAGAAACAGTGTGGTGGCATGAGGTGCAGTCGACTGGCACAGATGAGGTGTCGGT
GATGCATGAGAGGTTTTCTCTAGAACAAGTCAAGACTTGTTGGTTAGAGCCTGAGGATGATTGG
GAGGTAGCCATCAGGAATTATGCCAAGCTGGCTCTGAGGCCAGACAAGAAGTACAAGATTACTA
AGCTGATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTGAAGTGGAGATCTGTCT
CCAGGAAAGGGTGGCTTTCAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGGGCATGGA
TGGGGTTACCTTTATGAACATGAGGTTCAGGGGAGATGGGTATAATGGCACGGTCTTTATGGCC
AATACCAAGCTGACAGTCCATGGCTGCTCCTTCTTTGGGTTTAATAACACCTGCATCGAGGCCT
GGGGTCAGGTCGGTGTGAGGGGCTGCAGTTTTTCAGCCAACTGGATGGGGGTCGTGGGCAGG
ACCAAGAGTATGCTGTCCGTGAAGAAATGCTTGTTTGAGAGGTGCCACCTGGGGGTGATGAGC

GAGGGCGAAGCCAGAATCCGCCACTGCGCCTCTACCGAGACGGGCTGCTTTGTGCTGTGCAAG
GGCAATGCTAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCAGATG
CTGACCTGCGCCGGCGGGAACAGCCATATGCTGGCCACCGTACATGTGGCTTCCCATGCTCGC
AAGCCCTGGCCCGAGTTCGAGCACAATGTCATGACCAGGTGCAATATGCATCTGGGGTCCCGC
CGAGGCATGTTCATGCCCTACCAGTGCAACCTGAATTATGTGAAGGTGCTGCTGGAGCCCGAT
GCCATGTCCAGAGTGAGCCTGACGGGGGTGTTTGACATGAATGTGGAGGTGTGGAAGATTCTG
AGATATGATGAATCCAAGACCAGGTGCCGAGCCTGCGAGTGCGGAGGGAAGCATGCCAGGTTC
CAGCCCGTGTGTGTGGATGTGACGGAGGACCTGCGACCCGATCATTTGGTGTTGCCCTGCACC
GGGACGGAGTTCGGTTCCAGCGGGGAAGAATCTGACTAGAGTGAGTAGTGTTCTGGGGCGGG
GGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATG
AGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCC
TGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGC
GAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCA
GCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCAC
TCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCT
GCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTC
AGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAAT
AAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGC
GGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGT
AGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTC
CATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGG
GCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAG
GTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCC
TGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACC
ACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGA
AAGAATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGA
TGGGCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGG
TCCTGGGTGAGGTCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGG
ACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGA
GCTCGGAGGGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGG
GAGATGAGCTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCC
GTAGATGACCCCGATGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCC
GGAGGAGGGGGGCCACCTCGTTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCG
CCAGGAGGCGCTCTCCCCCCAGGGATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGC

TTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAG
AGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGAC
GGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAG
GGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGG
CGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCC
TGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCT
TTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGC
GTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCG
CAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTC
CCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGA
CAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGC
GGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACG
AAGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGT
ATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACG
TGACCGGGGGTCCCGGCCGGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGTC
TTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCAT
GACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGC
GGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGC
TTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTG
GTTTTTTTCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCAC
GCACTTCCATTCGGGGAAGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCG
ATTATGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCA
GCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCG
GGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGA
AGTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACGGCCAGCGCGCGCTCGTAGGGAC
TGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCG
TAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGAT
GCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCAGGTTG
GTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGA
GATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGA
AGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCG
CAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGCTCGC
GGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCTG
CACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCA

CGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTC
CCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCA
GAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAA
GAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGG
CCCGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCA
CGATGTAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGT
AGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGG
GTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGT
ACTGACGGAACTGCTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGG
TCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCG
GTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCAT
CCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGAT
GGGGAAGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAA
ATGCCGACGGCGCGCCGAACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACG
CTGCACGGGATGCACGTGCTGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGG
GAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTC
TTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGC
GAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACG
CTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGG
CGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTCGATGGCTT
GCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGC
GACGGGGGCGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAG
GGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCGGGTAG
GTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCTGGA
TCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCGACAG
AATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTTGT
CCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGG
CGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTC
ATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCAT
GACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCT
GGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGC
GGAGCGGCATCTCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCA
CGGCGAAGTTGAAAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGA
TGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCC

TCTTCTTCCTCCTCCACTAACATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAG
GGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCC
GCGCCGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAG
ACGCCGCCGCGCATCTCCAGGTGGCCGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGA
CGATGCATCTTATCAATTGCCCCGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCA
CGGGATCTGAAAACCGCTGAACGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCA
CGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCGGGCGATGCTGCTGGTGATGAAGT
TGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCT
TGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTT
GTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGC
GCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGC
GAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTG
GTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTG
GCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGT
CGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAG
AGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGT
AGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAAC
TCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTG
GCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGC
TCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAAT
CTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGC
ACCAACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACT
AGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCA
GGGTTGCGTTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTG
GCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCT
CTTTTGTTTTGTTTGTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTC
CACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAG
CCACGACCGCCGCGGCCGCCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTG
GAAGAGGGCGAGGGGCTGGCGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTG
CAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGC
GGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGTTCCACGCGGGGCGGGAGCTGCGGCGC
GGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGAT
CAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTG
AAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCACCCTGATCGCGCGCGAG

GAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCCAC
CAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTT
CAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACA
TTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATC
AACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGTACGTGC
CCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCCT
GAGCGACGATCTGGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGGC
GGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGG
GACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGG
CCTTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGG
CGAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTC
CTGATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGAT
TGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAG
CAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCC
CACGCACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCG
ACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAAC
GTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGC
GGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGC
CCGCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATG
GTGACCGAGGTGCCCCAGAGCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGT
CGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGG
CGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGC
CTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCAACCGCAACTCGTACCTG
GGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTAC
CAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCAC
CCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGA
GGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGTTCCTGATGCAGGAGGGGG
CCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGC
AACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGGCCGCCATGAACTCTGACT
ATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGGGGTTCTACACGGGCGAGT
ACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGACAGCAGCGTGTTCTCCC
CCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCCCGTCC
TCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCC
CGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCG

CGCTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAAC
TTCCCCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAG
GAGCACAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTA
AACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGAC
AGCAGCGTGTTGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGG
CGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGT
TCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTC
CCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGA
GGCTCCTTACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTC
GGAGCTGGCACCCTTGTACGATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACAT
CGCCTCGCTGAACTACCAGAACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGA
CTTCACCCCCACGGAGGCCAGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCG
GCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAA
GTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTA
TGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTC
TCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGG
CGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCT
GGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATC
CCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGC
TGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGG
GGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAAAGCAAGGAGGATGCAGCA
GCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGGGGCGATAATTTTGCAAGC
GCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAGCCGGT
GGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTACCG
CAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCT
CACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCA
AGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCT
CCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGC
CTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCC
GCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCG
CTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTG
CCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTA
AATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGAT
GTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCG

CTCCCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGAC
CAGGTGGTGGCCGACGCGCGCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGC
CGTCATCGACAGCGTGGTGGCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGG
CGCATCGCCCGGCGGCACCGGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCA
GGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGC
CAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATG
TCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGC
CCGTGCGCACCCGCCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGC
GGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGAT
CTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGA
CAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCG
GCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTG
GTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGG
GGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCA
GCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCG
AGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAA
GCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGG
AAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCC
ATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGGAGCC
CATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGG
ATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTG
CTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGC
TTCTACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCG
CACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCAC
CTCTGACCCTGCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCtGCTTTGC
AGATCAATGGCCCTCACATGCCGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCG
CGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCA
TCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATC
GGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTT
GGAAACATCTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGAC
AGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCA
CCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGA
GCGGGCTTAAGAATTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCA
CAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGG

CTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAA
CAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAG
CTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCT
GACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGC
GGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACTTG
CCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGC
CCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCA
TCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTT
AACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGT
GAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTAC
ATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGC
GCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCAC
GATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGA
CAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACAT
GGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTC
CGGCACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAA
AGCCGATGGTGAAACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATT
AACATCACAAAAGATGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAA
AACCTATCAGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAA
AAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAA
GCCTACTAATAAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATAT
GACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTG
TTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGCACA
GATGACAGCAGCTCTTCTATTAATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACATTG
GTTTCAGAGACAACTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGC
CGGTCAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTAC
CAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGG
ACAGCTATGATCCTGATGTGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTAT
TGTTTCCCTCTGGATGCTGTTGGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGA
TCAAACCACATGGACCAAAGATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCA
TTCGCCATGGAAATCAACATCCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCC
TGTACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACAC
CTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGG
GGCGCGCTGGTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCGGG

GCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCC
CCAGAAATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAAC
TTCCGCAAGGACGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGG
GCCTCCATCTCCTTCACCAGCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGG
CCTCCACGCTCGAGGCCATGCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCT
CGGCGGCCAACATGCTCTACCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCT
CGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCC
TCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGC
ACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGC
CCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAG
GGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCAC
TACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCT
TCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGG
CCGTCACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCGGCTACCTCGCGCCCACCATGC
GCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACC
AGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACT
TCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACG
CGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAA
GTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCG
CACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCG
GGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCT
GGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGT
CAACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGC
TCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACC
AGTTCGAGTACGAGGGCCTGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACC
CTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTG
CATGTTCCTGCACGCCTTCGTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAA
CTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCC
GCAACCAGGAGGCGCTCTACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCG
CGCGCATCGAGAAGGCCACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTAT
GTTAAATGTCTTTAATAAACAGCACTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAAT
CGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTAC
TTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTC
GGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAAT

CGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGGAAC
ACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTCGCGTCGGTGATGCTCTCCACGTCG
AGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGC
ACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCG
GCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCC
TTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCG
GCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCG
GTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGC
CACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTG
CCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTG
GGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGG
TCTTGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGA
TGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCA
CGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGG
GCAGGCTCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCT
CGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGTAGCTGA
AGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCT
GCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTG
GAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGA
GGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCG
CCGCGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGC
GCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGG
AGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGA
ATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATG
CAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGA
GGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGA
GAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGGG
AGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACC
GCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGC
CGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTC
TACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGA
TCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGC
GCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGC
GACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGC

CCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGA
CCCATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACCAGG
TGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGC
AAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAG
TTTGGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCT
GCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTT
CAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTC
CTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCG
GGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGA
CGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTC
CTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCT
GGCCGACCTCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTAT
GAGCCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCC
ACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCT
GTGGAGCCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGA
GGACGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACC
GCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGC
AAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTG
TGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCT
ACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCG
ATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCG
GGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCG
AGGAAACAAGAAGCTGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGA
ACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGGGACAGCACTCAGGCAGAGGAG
GACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGAGGAGGTGGAAGAAGC
AGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCAAGCAGCACGGATACCATCTCCG
CTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAAC
CCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAAAAA
CGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCT
CTTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCC
TACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAA
TCCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGG
AGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGC
AGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAG

CGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCT
CACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGT
GCCCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTG
GAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGA
ATTGGCTCAGCGCCGGGCCCGCGATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACC
AGATACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTG
GCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGC
CCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTC
GTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAAC
GACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCG
GGGAGATCTTCCTTCACGCCTCGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCC
CGCTCGGGTGGCATCGGCACTCTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAAC
CCCTTCTCCGGCTCCCCCGGCCACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGC
GAGTCGGTGGACGGCTACGATTGAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGA
CACCTGGACCACTGCCGCCGCTTCCGCTGCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTT
GAGCTGCCCGAGGAGCACCCTCAGGGCCCGGCCCACGGAGTGCGGATCGTCGTCGAAGGGG
GCCTCGACTCCCACCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGGTCGAGCGCGAGCAAG
GACAGACCCTTCTGACTCTGTACTGCATCTGCAACCACCCCGGCCTGCATGAAAGTCTTTGTTG
TCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGTTCCT
GAATCCATCAACCAGTCTTTGTTCTTCACCGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGC
CCCACAAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCG
ACAACGACGGAGTCCTGCTGAGCGGCCCTGCCAACCTTACTTTTTCCACCCGCAGAAGCAAGC
TCCAGCTCTTCCAACCCTTCCTCCCCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACAC
CTTCCACCTGATCCCGAATACCACAGCGTCGCTCCCCGCTACTAACAACCAAACTAACCTCCAC
CAACGCCACCGTCGCGACCTTTCTGAATCTAATACTACCACCCACACCGGAGGTGAGCTCCGA
GGTCAACCAACCTCTGGGATTTACTACGGCCCCTGGGAGGTGGTTGGGTTAATAGCGCTAGGC
CTAGTTGCGGGTGGGCTTTTGGTTCTCTGCTACCTATACCTCCCTTGCTGTTCGTACTTAGTGGT
GCTGTGTTGCTGGTTTAAGAAATGGGGAAGATCACCCTAGTGAGCTGCGGTGCGCTGGTGGCG
GTGTTGCTTTCGATTGTGGGACTGGGCGGTGCGGCTGTAGTGAAGGAGAAGGCCGATCCCTGC
TTGCATTTCAATCCCAACAAATGCCAGCTGAGTTTTCAGCCCGATGGCAATCGGTGCGCGGTAC
TGATCAAGTGCGGATGGGAATGCGAGAACGTGAGAATCGAGTACAATAACAAGACTCGGAACAA
TACTCTCGCGTCCGTGTGGCAGCCCGGGGACCCCGAGTGGTACACCGTCTCTGTCCCCGGTG
CTGACGGCTCCCCGCGCACCGTGAATAATACTTTCATTTTTGCGCACATGTGCGACACGGTCAT
GTGGATGAGCAAGCAGTACGATATGTGGCCCCCCACGAAGGAGAACATCGTGGTCTTCTCCAT

CGCTTACAGCCTGTGCACGGCGCTAATCACCGCTATCGTGTGCCTGAGCATTCACATGCTCATC
GCTATTCGCCCCAGAAATAATGCCGAAAAAGAAAAACAGCCATAACGTTTTTTTTCACACCTTTTT
CAGACCATGGCCTCTGTTAAATTTTTGCTTTTATTTGCCAGTCTCATTGCCGTCATTCATGGAATG
AGTAATGAGAAAATTACTATTTACACTGGCACTAATCACACATTGAAAGGTCCAGAAAAAGCCAC
AGAAGTTTCATGGTATTGTTATTTTAATGAATCAGATGTATCTACTGAACTCTGTGGAAACAATAA
CAAAAAAAATGAGAGCATTACTCTCATCAAGTTTCAATGTGGATCTGACTTAACCCTAATTAACAT
CACTAGAGACTATGTAGGTATGTATTATGGAACTACAGCAGGCATTTCGGACATGGAATTTTATC
AAGTTTCTGTGTCTGAACCCACCACGCCTAGAATGACCACAACCACAAAAACTACACCTGTTACC
ACTATGCAGCTCACTACCAATAACATTTTTGCCATGCGTCAAATGGTCAACAATAGCACTCAACC
CACCCCACCCAGTGAGGAAATTCCCAAATCCATGATTGGCATTATTGTTGCTGTAGTGGTGTGC
ATGTTGATCATCGCCTTGTGCATGGTGTACTATGCCTTCTGCTACAGAAAGCACAGACTGAACG
ACAAGCTGGAACACTTACTAAGTGTTGAATTTTAATTTTTTAGAACCATGAAGATCCTAGGCCTTT
TAATTTTTTCTATCATTACCTCTGCTCTATGCAATTCTGACAATGAGGACGTTACTGTCGTTGTCG
GATCAAATTATACACTGAAAGGTCCAGCGAAGGGTATGCTTTCGTGGTATTGCTATTTTGGATCT
GACACTACAGAAACTGAATTATGCAATCTTAAGAATGGCAAAATTCAAAATTCTAAAATTAACAAT
TATATATGCAATGGTACTGATCTGATACTCCTCAATATCACGAAATCATATGCTGGCAGTTACAC
CTGCCCTGGAGATGATGCTGACAGTATGATTTTTTACAAAGTAACTGTTGTTGATCCCACTACTC
CACCTCCACCCACCACAACTACTCACACCACACACACAGATCAAACCGCAGCAGAGGAGGCAG
CAAAGTTAGCCTTGCAGGTCCAAGACAGTTCATTTGTTGGCATTACCCCTACACCTGATCAGCG
GTGTCCGGGGCTGCTAGTCAGCGGCATTGTCGGTGTGCTTTCGGGATTAGCAGTCATAATCATC
TGCATGTTCATTTTTGCTTGCTGCTATAGAAGGCTTTACCGACAAAAATCAGACCCACTGCTGAA
CCTCTATGTTTAATTTTTTCCAGAGTCATGAAGGCAGTTAGCGCTCTAGTTTTTTGTTCTTTGATT
GGCATTGTTTTTTGCAATCCTATTCCTAAAGTTAGCTTTATTAAAGATGTGAATGTTACTGAGGGG
GGCAATGTGACACTGGTAGGTGTAGAGGGTGCTGAAAACACCACCTGGACAAAATACCACCTCA
ATGGGTGGAAAGATATTTGCAATTGGAGTGTATTAGTTTATACATGTGAGGGAGTTAATCTTACC
ATTGTCAATGCCACCTCAGCTCAAAATGGTAGAATTCAAGGACAAAGTGTCAGTGTATCTAATGG
GTATTTTACCCAACATACTTTTATCTATGACGTTAAAGTCATACCACTGCCTACGCCTAGCCCAC
CTAGCACTACCACACAGACAACCCACACTACACAGACAACCACATACAGTACATTAAATCAGCCT
ACCACCACTACAGCAGCAGAGGTTGCCAGCTCGTCTGGGGTCCGAGTGGCATTTTTGATGTtGG
CCCCATCTAGCAGTCCCACTGCTAGTACCAATGAGCAGACTACTGAATTTTTGTCCACTGTCGA
GAGCCACACCACAGCTACCTCCAGTGCCTTCTCTAGCACCGCCAATCTCTCCTCGCTTTCCTCT
ACACCAATCAGTCCCGCTACTACTCCTAGCCCCGCTCCTCTTCCCACTCCCCTGAAGCAAACAG
ACGGCGGCATGCAATGGCAGATCACCCTGCTCATTGTGATCGGGTTGGTCATCCTGGCCGTGT
TGCTCTACTACATCTTCTGCCGCCGCATTCCCAACGCGCACCGCAAGCCGGTCTACAAGCCCAT

CATTGTCGGGCAGCCGGAGCCGCTTCAGGTGGAAGGGGGTCTAAGGAATCTTCTCTTCTCTTTT
ACAGTATGGTGATTGAACTATGATTCCTAGACAATTCTTGATCACTATTCTTATCTGCCTCCTCCA
AGTCTGTGCCACCCTCGCTCTGGTGGCCAACGCCAGTCCAGACTGTATTGGGCCCTTCGCCTC
CTACGTGCTCTTTGCCTTCACCACCTGCATCTGCTGCTGTAGCATAGTCTGCCTGCTTATCACCT
TCTTCCAGTTCATTGACTGGATCTTTGTGCGCATCGCCTACCTGCGCCACCACCCCCAGTACCG
CGACCAGCGAGTGGCGCGGCTGCTCAGGCTCCTCTGATAAGCATGCGGGCTCTGCTACTTCTC
GCGCTTCTGCTGTTAGTGCTCCCCCGTCCCGTCGACCCCCGGTCCCCCACCCAGTCCCCCGAG
GAGGTCCGCAAATGCAAATTCCAAGAACCCTGGAAATTCCTCAAATGCTACCGCCAAAAATCAG
ACATGCATCCCAGCTGGATCATGATCATTGGGATCGTGAACATTCTGGCCTGCACCCTCATCTC
CTTTGTGATTTACCCCTGCTTTGACTTTGGTTGGAACTCGCCAGAGGCGCTCTATCTCCCGCCT
GAACCTGACACACCACCACAGCAACCTCAGGCACACGCACTACCACCACTACAGCCTAGGCCA
CAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCTATTAGTT
ACTTCAATCTAACCGGCGGAGATGACTGACCCACTGGCCAACAACAACGTCAACGACCTTCTCC
TGGACATGGACGGCCGCGCCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCAGCAG
GAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCTT
CTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTAC
GAGCTCCTGCAGCAGCGCCAGAAGTTCACCTGCCTGGTCGGAGTCAACCCCATCGTCATCACC
CAGCAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCAC
ACTCTGATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCACCCCCTTATC
CAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATA
AAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTGAT
ACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTG
CAGGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGTCC
CTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGA
CCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGT
CTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGT
CACCACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGGTGGACCTCGATTCCTCGG
GAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCAT
TTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACC
ATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACT
CCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAA
AGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTG
GGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTG
GAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGG

CCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGT
GGACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACACTT
TGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGAAA
CCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGT
GTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGA
TGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAA
GTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGC
TTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACA
CCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATC
GCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAACA
AACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATTCG
AGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAGCC
TTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGA
GCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCT
CACAGCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGA
AGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCC
GCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGAC
TCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCG
CATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGT
CCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCG
TACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGATCT
CCTTGGGCATGTGGCGGTTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCAGC
CCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAGAC
CCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCT
GAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCT
CGGGGGTCAAAACCATATCCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAGAA
CAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACC
GGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGG
CCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTGCT
TTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGC
GGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGT
GCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATC
GACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACGGC
GGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTACTTCA

AAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCAGG TCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCACAT CCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTACACTC CTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCcTGAGGTAA ATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACAC CCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATATCAAA ATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTCTC CGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAACCG AAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTG ATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAAAA ATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTC CAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGC GAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGACCC TCGTAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAA TGATTCGACAAGATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCGCCCGA GGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAA GCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCC CGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAAC AGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAG CCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCC AGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCAAAA CTGCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCG TTAAAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCA TCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATG ATGG
ChAdV68.4WTnt.GFP (SEQ ID NO:11); AC_000011.1 con secuencias E1 (nt 577 a 3403) y E3 (nt 27,816- 31,332) eliminadas; nucleótidos de ATCC VR-594 correspondientes sustituidos en cuatro posiciones; GFP indicadora bajo control del promotor/potenciador de CMV insertado en el lugar de E1 eliminada
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTG GGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGT TTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAAC GAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTT TCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAAT

CTGAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGA
TTACGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGT
GTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCC
ACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAA
GTAGGGATAACAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCC
AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT
TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC
ATGgTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCC
AAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
CTATATAAGCAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGA
CCTCCATAGAAGACAGCGATCGCGccaccATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT
GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG
AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAG
CTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC
TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG
GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTC
GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCA
CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT
GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTtTACAAGTAGtgaGTTT
AAACTCCCATTTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTT
GGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGC
TTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTT
CAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAA
TAACTATAACGGTCCTAAGGTAGCGAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGA
GGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATGAGCGGAAGCGGCTCC
TTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGT
CAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGCGAACTCTTCAACCCT

GACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCAGCTGCTGCATCTGC
CGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACT
CGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCCAGC
TCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAG
ACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACG
GTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGCGGTAGGCCCTGGACC
ACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGTAGAGGTGGGCTTGGA
TGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTCCATTGCAGGGCCTCG
TGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGTGTTGCACA
ATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTGT
TGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTG
GCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCACCAGCACGGTGTAT
CCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACG
CCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGATGGGCCCGTGGGCG
GCGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCA
TCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCG
ATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGG
ATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGC
CGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGA
TGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCC
ACCTCGTTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCT
CCCCCCAGGGATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGC
CATGGGCATTTTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTG
CTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAG
GGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCG
CGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTG
CGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAG
GTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTT
ACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGG
CGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCGCAGACGGTCTCGCA
CTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTCCCGCCGTTCTTTTT
GATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCC
GTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGCGGTCCTCCTCGTAG
AGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAAGGAGGCCACGTG

GGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATGTC
CCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCC
GGCCGGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGTCTTCCGGATCGCTGTC
CAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAG
GTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAG
AGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGCTTGGTGGCGAAGGAGC
CGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTTTCCTTGTCGG
CGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGA
AGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCGATTATGCAGGGTGATGA
GGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGCCC
TTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGA
TGGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCC
AGGGCAGCTTGCCATTCGCGCACGGCCAGCGCGCtCTCGTAGGGACTGAGGGGCGTGCCCCA
GGGCATGGGATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCT
CCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAG
TCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTC
GGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGA
AGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCT
TGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTC
CTGGATGATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTT
CGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCA
TGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCT
GGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTCCCTGACCATGACCTTGAGG
AACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCAGAGCTGGAAGTCCGTGCGC
TTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGCGG
GGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGATGACCTGG
GCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAAT
CGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGGTC
GCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAA
GTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTGCTGCCC
GACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCC
ATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCCGGAGAGTTTC
ATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACA
TCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCC

TGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAA
CACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGC
TGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCC
TGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTC
ATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGA
GGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGT
GGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGAT
GGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCT
GGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCT
TCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCG
GAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCG
GAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCTGGATCTGACGCCTCTGGGTGA
AGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCGACAGAATCAATCTCGGTATCGTT
GACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTTGTCCTGGTAGGCGATCTCGGT
CATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGC
GAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCATGCCCGCCTCGTTCCAGAC
GCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCATGACCACCTGGGCGAGGTTGA
GCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTG
GTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACG
TCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGG
GAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGGC
GCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCCTCCTCCACTAAC
ATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCG
GCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCT
CGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAG
GTGGCCGGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCC
CCGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAA
CGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCAT
GTTGGTTGGGAGCGGGGCGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGA
CGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTC
GGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCG
CTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCT
GGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGCGAGGATGGCTTGCTGGATCTG
GGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGT

GTAGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGT
GGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGG
TACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGC
GGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGTAGCCGTAGATGTACCTGGACA
TCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAACTCGCGGACGCGGTTCCAGAT
GTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCCGTGAGGCGCGCGCAGT
CGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTCCGTGGCCTGGAGG
CTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCTCGAATCAGGCTGGAGCCGC
AGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGA
GGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCC
GACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCC
CCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAG
ACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGC
CAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCT
CCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGC
CGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGG
CGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCG
CGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGA
TGCGCGCGGCCCGGTTCCACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGT
GCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCAC
GTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCA
AAAATCCTTCAACAACCACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGAT
GCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGC
AGCTGTTCCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATA
TCACCGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGC
AGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTG
GGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAG
ATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTG
TACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGGCGGCGCGAGCTGAGCGACCA
GGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGGGACCGAGGGGGAGAGCTACT
TTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCCTTGGAGGCGGCGGCAGG
ACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGAT
GGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATGCGGGC
GGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAAC

GCATCATGGCGCTGACGACCCGCAACCCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGG
CTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGCACGAGAAGGTCCT
GGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCCGGCCTGGTGT
ACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGGAC
CGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAA
CCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGG
GCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGA
GCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCG
TGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGG
GACCGCGCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGC
CCCCTTCACGGACAGCGGCAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTAC
CGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTGAG
CCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAA
CCGGTCGCAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTT
ACGTGCAGCAGAGCGTGGGCCTGTTCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTC
GACATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAA
CTGATGGACTACTTGCATCGGGCGGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGA
ATCCCCACTGGCTCCCGCCGCCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAAT
GACGGGTTCCTGTGGGACGATGTGGACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGA
GCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCCCGTCCTCGGCGCTGTCCGGCCGCG
AGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCCCGAGCTTGCCCTTCTCGCTG
AACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGAAGAGGA
GTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAA
AGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCG
GGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGGTGGCACGACAGG
CAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGTTGGACTTGGG
TGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACC
GAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTTGT
TGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCGTGAT
GCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCCG
CGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTAC
GATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAG
AACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCC
AGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATG

CACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATG
GTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAG
CTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCG
ACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGCGGCAGAACGGGGTGCTGG
AGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACC
GAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCC
GGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCA
GCCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCT
CCTGGATGTCGACGCCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGT
AGCTACCGCCTCTACCGAGGTCAGGGGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGG
CCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACA
GGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTACCGCAGCTGGTACCTAGCCTACAA
CTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACCACCTCGGACGTCACCT
GCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCAAGACCCGGTCACCTTCCGCT
CCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGA
GCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACG
TCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCG
TCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGG
GAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGCCCCTACGTCTACAAGGCCC
TGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCC
CAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAAC
GCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCAAG
GGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCGC
GCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTG
GCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACC
GGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACG
CAGGGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGG
AGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGA
ACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCT
CGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGC
GCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGA
AGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTG
ATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGC
GGGCGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCG

CTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCA
GGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAG
AGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCGAGCCTCAAGCCCGTGACCTTG
CAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAAGCGCGAGGGCGAGGATCTGT
ACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGA
AGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGG
CCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGGAGCCCATGGAAACGCAGACCGAGC
CCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTC
CTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTG
CATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCGCGGTCATACCAGC
AGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCGCACCGCCGCTGCAACCACCCC
TGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGCCGCGCGCG
CGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCtGCTTTGCAGATCAATGGCCCTCACATGC
CGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGG
GAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGA
GGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGC
TTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACC
aATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTT
CGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACC
AGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGG
TCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGAT
AAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGG
GTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCC
GCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGG
GGCGAGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGC
CCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCC
ACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGC
CCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGGCA
CCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGA
GTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGTA
TGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTT
GCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGAC
GCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGT
CTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATGTGACCACCGACCGCAG

CCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACACCTACTCGTACAAAGT
GCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCACCTACTTTGACAT
CCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACCGCCTACAACAGTCT
GGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAAGCCGATGGTGAAACTGCCAC
AGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTC
AACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCA
AGTGGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTT
AAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGG
TCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTG
ACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGA
TTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGCACAGATGACAGCAGCTCTTCTATTA
ATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACATTGGTTTCAGAGACAACTTTATCGG
GCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTGAA
TGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCTG
GGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGC
GCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTT
GGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAG
ATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACAT
CCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTAC
AAGTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCC
GGGTGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGAT
CCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATG
CTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCA
AGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACA
TGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCA
GCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCA
TGCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTA
CCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTT
CCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCG
ACCCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACAC
CTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCT
GACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTACAACGTGGCCCAGT
GCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGG
GCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCAT

GAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTCACCCTGGCCTACC
AGCACAACAACTCGGGCTTCGTCGGCTACCTCGCGCCCACCATGCGCCAGGGCCAGCCCTACC
CCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACCAGCGTCACCCAGAAAAAGT
TCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGC
TCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACGCGCTAGACATGAATTTCGA
AGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAAGTCTTCGACGTCGTCCGA
GTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCCCCTTCTCGGCCGG
TAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGCAGGAG
CTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGC
TTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGA
GACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCTCT
TCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCC
TGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCACCCAG
ACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTC
GTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCC
CAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCT
ACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCA
CCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAAC
AGCACTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGT
CTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACT
CGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTC
AGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTT
CTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCT
TCACGCTCGCCAGCACCGTCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCA
TCCCGAAGGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGT
TGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACA
TGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGA
AGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAG
CGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGC
CCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCAT
GTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCCCTCGGCCTCGGTGCA
CCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTGGGCGATCTGGGAATGCG
CGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCTTGTTGCTAGTGAAGG
TCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTCGC

CCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCACGCGGTAGCGGTCCATCA
GCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTTCTT
CACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTCAAAGCT
CCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGGTAGCTGAAGCCCACGGCCGCCAGCT
CCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGGACCACATGCTTGGT
CTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTGGAGATGGCGAGGGGGAGC
GCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGT
ATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACTTGGCGGATGGC
TGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTCCG
CGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACC
TCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCG
CCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCAT
CGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTT
CACAAGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTG
GGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGGGGAGGACGCGCTCATCAAGCAT
CTGGCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAG
CGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGC
GCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGC
CCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGC
CAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGC
CTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAA
CGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCG
ACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTC
TGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGC
CCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAG
CAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTC
ATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGC
GGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCA
GGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGA
GAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACA
TCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGGCATGGGCGTGTGGCAGC
AGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGGTC
TGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTCATTTTCCCCGAGC
GCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCAAAACT

TTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGG
ACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGC
GCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGGCCTG
CTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAG
CTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGG
TTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAA
GTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCC
CAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCAT
CCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGAC
CGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTG
GAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGA
GGAGATGGAGGAAGACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGG
AGGAAGACGAGGAGGAGGCAGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTC
CTCGGCGGGGGAGAAAGCAAGCAGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTC
GACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGA
AGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAG
GCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTT
CCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAG
CAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGT
GGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCC
CACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCG
TTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACT
CTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCTCACTCTTAAAGAGTAGCCCGCG
CCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGCCCTTCGCCCTAGCCGCCTCC
ACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGC
CTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGC
GATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGTCAGC
GCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCA
GGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGAC
TAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTCAGGGTA
TAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCG
CTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGCCT
CGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACT
CTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCC

ACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATT
GAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGGACCACTGCCGCCGCT
TCCGCTGCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCCGAGGAGCACCCTC
AGGGCCCGGCCCACGGAGTGCGGATCGTCGTCGAAGGGGGCCTCGACTCCCACCTGCTTCGG
ATCTTCAGCCAGCGTCCGATCCTGGTCGAGCGCGAGCAAGGACAGACCCTTCTGACTCTGTAC
TGCATCTGCAACCACCCCGGCCTGCATGAAAGTCTTTGTTGTCTGCTGTGTACTGAGTATAATAA
AAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGTTCCTGAATCCATCAACCAGTCTTTGTTC
TTCACCGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGCCCCACAAGAAGTACCTCACCTGG
CTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCGACAACGACGGAGTCCTGCTGAGC
GGCCCTGCCAACCTTACTTTTTCCACCCGCAGAAGCAAGCTCCAGCTCTTCCAACCCTTCCTCC
CCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACCA
CAGCGTCGCTCCCCGCTACTAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACGGCC
ACAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCTATTAGT
TACTTCAATCTAACCGGCGGAGATGACTGACCCACTGGCCAACAACAACGTCAACGACCTTCTC
CTGGACATGGACGGCCGCGCCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCAGCA
GGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCT
TCTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTA
CGAGCTCCTGCAGCAGCGCCAGAAGTTCACCTGCCTGGTCGGAGTCAACCCCATCGTCATCAC
CCAGCAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCA
CACTCTGATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCACCCCCTTAT
CCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAAT
AAAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTG
ATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTAC
TGCAGGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGT
CCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTC
GACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTC
GTCTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCC
GTCACCACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCG
GGAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCA
TTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCAC
CATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGA
CTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACAT
AAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCT
GGGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTT

GGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTG
GCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTT
GTGGACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACAC
TTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGA
AACCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACG
GTGTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATA
GATGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACA
AAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTAT
GCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATA
CACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACA
TCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAA
CAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATT
CGAGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAG
CCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCA
GAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCAC
CTCACAGCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCA
GAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCC
CCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGG
ACTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAG
CGCATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTGTTCAAC
AGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCG
TCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGA
TCTCCTTGGGCATGTGGCGGTTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCA
GCCCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAG
ACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAG
CTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTC
CTCGGGGGTCAAAACCATATCCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAG
AACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCAC
CGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCG
GCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTG
CTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCG
GCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACC
GTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACA
TCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACG

GCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTACTT CAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCA GGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCA CATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTACAC TCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCGTGAGG TAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCAC ACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATATCA AAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTC TCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAAC CGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCG GTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGA AAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCG TTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCT GGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGA CCCTCGTAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGT GAATGATTCGACAAGATGAATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCGCC CGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGAT GAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGC CCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACA CAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATAT ATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACAC GCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCC CAAAACTGCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCG ACCGTTAAAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCC CGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATT GATGATGG
ChAdV68.4WTnt.MAG25mer (SEQ ID NO:12); AC_000011.1 con secuencias E1 (nt 577 a 3403) y E3 (nt 27.816- 31.332) eliminadas; nucleótidos de ATCC VR-594 correspondientes sustituidos en cuatro posiciones; casete de neoantígeno modelo bajo control del promotor/potenciador de CMV insertado en el lugar de E1 eliminada
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTG GGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGT TTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAAC GAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTT

TCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAAT
CTGAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGA
TTACGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGT
GTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCC
ACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAA
GTAGGGATAACAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCC
AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT
TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC
AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC
ATGgTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCC
AAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT
CTATATAAGCAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGA
CCTCCATAGAAGACAGCGATCGCGccaccATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTG
CACACCCTATGATATTAACCAGATGCTGAATGTCCTGGGAGACCACCAGGTCTCTGGCCTGGAG
CAGCTGGAGAGCATCATCAACTTCGAGAAGCTGACCGAGTGGACAAGCTCCAATGTGATGCCTA
TCCTGTCCCCACTGACCAAGGGCATCCTGGGCTTCGTGTTTACCCTGACAGTGCCTTCTGAGCG
GGGCCTGTCTTGCATCAGCGAGGCAGACGCAACCACACCAGAGTCCGCCAATCTGGGCGAGG
AGATCCTGTCTCAGCTGTACCTGTGGCCCCGGGTGACATATCACTCCCCTTCTTACGCCTATCA
CCAGTTCGAGCGGAGAGCCAAGTACAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGCTGTT
CGGCTACCCCGTGTACGTGTTCGGCGATTGCGTGCAGGGCGACTGGGATGCCATCCGGTTTAG
ATACTGCGCACCACCTGGATATGCACTGCTGAGGTGTAACGACACCAATTATTCCGCCCTGCTG
GCAGTGGGCGCCCTGGAGGGCCCTCGCAATCAGGATTGGCTGGGCGTGCCAAGGCAGCTGGT
GACACGCATGCAGGCCATCCAGAACGCAGGCCTGTGCACCCTGGTGGCAATGCTGGAGGAGA
CAATCTTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCGGCCCCAAGACAAACATCAT
CGTGGATTCCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGTTTGTGGACTGGGAG
AACGTGAGCCCAGAGCTGAATTCCACCGATCAGCCATTCTGGCAGGCAGGAATCCTGGCAAGG
AACCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCCAGAGCCTG
GTCATCAGCGCCTCCATCATCGTGTTTAACCTGCTGGAGCTGGAGGGCGACTATCGGGACGAT
GGCAACGTGTGGGTGCACACCCCACTGAGCCCCAGAACACTGAACGCCTGGGTGAAGGCCGT
GGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGGCCTCCATGACCAATATGGAGCTGAT
GTCTAGCATCGTGCACCAGCAGGTGAGGACATACGGACCCGTGTTCATGTGCCTGGGAGGCCT

GCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGCTGTTCAGAGCCG
CCCAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCGGAGCAGCCTTTCGCCAGG
TGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGTGGAACAATGAGACAA
CACAGCCTCAGATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTGGCTGCACTACTATAGCGT
GAGGGATACCCTGTGGCCCCGCGTGACATACCACATGAATAAGTACGCCTATCACATGCTGGA
GAGGCGCGCCAAGTATAAGAGAGGCCCTGGCCCAGGCGCAAAGTTTGTGGCAGCATGGACCC
TGAAGGCCGCCGCCGGCCCCGGCCCCGGCCAGTATATCAAGGCTAACAGTAAGTTCATTGGAA
TCACAGAGCTGGGACCCGGACCTGGATAATGAGTTTAAACTCCCATTTAAATGTGAGGGTTAAT
GCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGA
AAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATA
AACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTCAGGGGGAGATGTGGGAGGTT
TTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGGTAGCGAGT
GAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTT
TCTGTGTGTTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTAT
CTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGA
CGGCCGGCCCGTGCAGCCCGCGAACTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTC
GTTGGACGCAGCTGCCGCCGCAGCTGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCA
TGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCC
TGAACGAGGAGAAGCTGTTGCTGCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGC
GAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAA
ATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAAT
CTTTATTTGATTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGT
GGATCTTTTCCAGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGT
CCCGGGGGTGGAGGTAGCTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACC
CAGTCATAGCAGGGGCGCAGGGCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCC
ACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGG
GAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTG
GGGTTCATGTTGTGCAGGACCACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCA
ACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATGC
ACTCATCCATGATGATGGCGATGGGCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGG
TCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCATCATAGGCCATTTTAATGAATTTGGGGC
GGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAG
ATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGGATCATGTCCACCTGCGGGGCGATAAAG
AACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGA

CTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGACCGGCTGCAGGTGGTAGTTGAGGG
AGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGTTCATCATCTCGCGCACGTGC
ATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGGGATAGGAGCTCCTGGAGC
GAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATTTTGGAGAGGGTTTGTTGC
AAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCT
CCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCA
GCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTCACGGTGAA
GGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAA
ACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAGGTAGCAATTGACCATGAGTTCGTAGTTGA
GCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAAGTCTGCCCGCAGGCGGGA
CAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGGGGCGTAGG
CGTCCGCGCCGCAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCGGGCTG
GTCGGGGTCAAAAACCAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGA
GCTCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCC
GGTCCTCGAGCGGTGTGCCGCGGTCCTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAA
GCCCGGGTCCAGGCCAGCACGAAGGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAG
CGGGTCCACCTTTTCCAGGGTATGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATT
GGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGGCCGGGGGGGGTATAAAAGGGTGCGGG
TCCCTGCTCGTCCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTA
TTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGA
TTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAG
ACGATCTTTTTGTTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTG
GCGATGGAGCGCATGGTCTGGTTTTTTTCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGC
TGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGAAGACGGTGGTCAGCTCGTCGGGCAC
GATTCTGACCTGCCAGCCCCGATTATGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCC
GCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGGGGCAGG
GGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGAGGTC
GGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACGG
CCAGCGCGCtCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAG
GCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGG
GTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGA
GGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAA
ATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAG
TCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGAGCTCGGCGG

TGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATACTTGAGCTGTC
CCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCCAGTACTCTTCGAG
GGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTGTA
GGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTG
TGCGTGAGGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCG
TCGCAGCCCCCCTGCTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAA
GCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGG
AAAGGTTGGGGCACCTCGGCCCGGTTGTTGATGACCTGGGCGGCGAGCACGATCTCGTCGAA
GCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAATCGCGGACGGCCCTTGACGTGGGG
CAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGGTCGCTGAGCCCGTGCTGCTCGAGCGC
CCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAGGGCG
GTTTGCAGACGGTCCCGGTACTGACGGAACTGCTGCCCGACGGCCATTTTTTCGGGGGTGACG
CAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAG
GGCGAGCTCGACGAGCCGGTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCT
GCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGG
TGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGT
TGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAACACTCGTGCTTGTGTTTATACAAGCG
GCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAGCTGTACCTGAGTTCCTTT
GACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACTACGTCGTG
GTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGC
AGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCT
GTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTT
GCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGG
TGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTCT
TCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGAC
GCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGG
CGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGA
CGGTTGACGTCCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTG
AAAGAGAGTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCTTGCA
CGTCGCCCGAGTTGTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTCTTGAAG
GTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCCCATGAGCT
GCGAGAAGGCGTTCATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGA
TCGCgGGCGCGCATGACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTGAAGACCGCGTA
GTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGGTGACGAAGAAATA

CATGATCCAGCGGCGGAGCGGCATCTCGCTGACGTCGCCCAGCGCCTCCAAACGTTCCATGGC
CTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGGGAGTTGCGCGCCGAGACGGTCAACTCCTC
CTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCCCGGGA
GTTCCTCCACTTCCTCTTCTTCCTCCTCCACTAACATCTCTTCTACTTCCTCCTCAGGCGGCAGT
GGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCT
CGATGGTCTCGCCGCGCCGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCCTCGCGGGG
CCGCAGCGTGAAGACGCCGCCGCGCATCTCCAGGTGGCCGGGGGGGGTCCCCGTTGGGCAGG
GAGAGGGCGCTGACGATGCATCTTATCAATTGCCCCGTAGGGACTCCGCGCAAGGACCTGAGC
GTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCGAGCCAGTCGCAGTCGCAA
GGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCGGGCGATGCT
GCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGT
CTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCCTGACAC
CTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCG
CGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGCGA
CGACGCGCTCGGCGAGGATGGCTTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGT
CGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAG
TTGACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGT
GTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCG
GCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAG
CATGGTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGG
AGGCGCGCGGGAACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATG
GTGGGCACGGTCTGGCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGA
AAGCGGTCAGCGGCTCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGT
ACCCCGGTTCGAATCTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCT
CGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGA
GGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTG
GAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGG
CTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTA
CGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGCCAGATGCATCCCGTACTGCGGCAGATGC
GCCCCCACCACCCTCCACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCC
CAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGCCGTGAGCGGGGCTGGACAGAGTTATGA
TCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGGCGCGCCTGGGGGCGTCGTCGCCGGAG
CGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAGCAGAACCT
GTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGTTCCACGCGGGG

CGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTCGAGGCGG
ACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGC
GTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCAC
CCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCA
TCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCCTGGTGGTGCAGCATAGTC
GGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCACCGAGCCCGAGGGCCGCTGGCTC
CTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCGGGCTGCCGCTGTCCGA
GAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGGAAGATCTAC
AAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGACC
CTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGC
GGTGAGCGCCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGG
GCCCTGACCGGGGCCGGGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTG
GCAGCCCAGCCGCCGGGCCTTGGAGGCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGAT
GAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGATGGCGCGACCGTATTTTTGCTAGATGC
AACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGC
ATTAACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACGACCCGCAAC
CCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGT
GCCCTCGCGCTCCAACCCCACGCACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGA
ACAAGGCCATCCGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCC
CGCTACAACAGCACCAACGTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGC
CGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACG
CCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATCA
GCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTGTACCAGTCCGGGCCGGAC
TACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGCTTTCAAGAACT
TGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTGTCGAGCCTGCT
GACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCA
ACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACG
TGGACGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCG
GGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGCCCCAG
TACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGTT
CCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGGAGC
CCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGGGCGG
CCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCGCCGG
GGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGG

ACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGC
AGCGACCGACGCCCGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCG
AGGCCGCCAGTCCTTTCCCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGG
GCAGGATCACGCGCCCGCGCTTGCTGGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGAC
CCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAAGCCTGGTGGACAAGATGAGCCGCT
GGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTCGCAGGGGGCCACGAGCCG
GGGCAGCGCCGCCCGTAAACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGACGAT
GAGGACTCCGCCGACGACAGCAGCGTGTTGGACTTGGGTGGGAGTGGTAACCCGTTCGCTCA
CCTGCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGC
CATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGC
GTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGG
CGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGG
CGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTACGATACCACCCGGTTGTACCTGGTG
GACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGACCACAGCAACTTCCTGACC
ACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCACCCAGACCATCAACTTTGAC
GAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGTGAAC
GAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGG
GTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTG
AGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGA
CAATTACTTGGCGGTGGGGCGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCG
ACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTAC
ACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAG
AGCCGCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCAGATC
ATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAA
AGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGG
GGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAA
GATAGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAA
GATAAACACCGCCTACCGCAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGT
GCGCTCCTGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTC
GCTGCCCGACATGATGCAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCC
GGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCTA
CTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAACCA
GATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCTCAC
AGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTG

ACGCCAGACGCCGCACCTGCCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTC
CTCTCGAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCC
TGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGC
GTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACCA
CCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCGCGCAACTACACCCCCGCCGCCGCG
CCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTGGCcGACGCGCGCCGGTACGCCCG
CGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGGAGCACCCCCGCCATGCGCGCG
GCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCA
GACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGG
CAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCC
GCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCTCGCACTTGAAGATGTTCACTTCGC
GATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGCGCAAATTCAAGGAAGAGATGCTCC
AGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAAGGAGGAAAGAAAGCCCCGCAAAA
TCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACGGATTGGTGGAGTTTG
TGCGCGAGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCT
GAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCT
CCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAG
TTTGCTTACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGAC
CACGGCAACCCCACGCCGAGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGC
GCCGCGCCGGGGGTTCAAGCGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGC
CCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCC
GAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACAT
CAAGATTCCCACGGAGCCCATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCAC
CATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTA
CGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGG
CTACCGCGGCACGCGCTTCTACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTC
GCCGCCGCCGTCGCCGCACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTAC
CGCCGCGGCCGCGCACCTCTGACCCTGCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTA
AACTTTCGCCtGCTTTGCAGATCAATGGCCCTCACATGCCGCCTTCGCGTTCCCATTACGGGCTA
CCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCACCACCACC
GGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCCCCATC
ATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGCG
CCACTGAGACACACTTGGAAACATCTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTG
ATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCGACACGGCACGC

GGCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATT
GGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGG
CGTGGAACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAG
AAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGGGTGGTGGACCTGGCCAACCAGGCCGT
GCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGCTCCGTGGAGATGCCG
CAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCCCGATG
CGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTG
GGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCC
CGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCC
GCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGGCACCGCCCGCCCTCATGCGAACTGGCAG
AGCACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAA
ACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGTATGTATTATGTCGCCGCCGCCGCTGTC
CACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGGCCACCCCATCGATGCT
GCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGTCCGGGTC
TGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCA
CGGTGGCGCCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTG
CCCGTGGACCGCGAGGACAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGA
CAACCGCGTGCTGGACATGGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCC
CTAGCTTCAAACCCTACTCCGGCACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACA
CTTGTCAGTGGACATATAAAGCCGATGGTGAAACTGCCACAGAAAAAACCTATACATATGGAAAT
GCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTCAACTTGGAACTGACACCGATGATC
AGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGA
CATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCT
TGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAAC
AGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCT
GGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATAT
TGTATACAAAGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTCAGCAAGCCATGCCC
AACAGACCTAACTACATTGGTTTCAGAGACAACTTTATCGGGCTCATGTACTACAACAGCACTGG
CAATATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAG
AAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAGT
ATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGCGCATTATTGAAAATCATGGTGTGG
AGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAGAACAGATACTTATCAGGGA
ATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAGATGACAGTGTCAATGATGCTAATGA
GATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACATCCAAGCCAACCTGTGGAGGAACTTC

CTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCC
TGCCCACCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGG
ACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATGGACAACGTGAACCCCTTCA
ACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTG
CCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGGT
CCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGATCCTGCAGAGCTCCCTCGGCA
ACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGCATCAACCTCTACGCCACCTTCTT
CCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAACGACACCAACGACC
AGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACGCCACCA
ACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTTCCGCGGCTGGTCCTTCACGCGTC
TCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCT
CCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTT
CGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAA
GCGCACCGTCGACGGCGAGGGCTACAACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCC
TGGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGGGCTTCTACGTGCCCGAGGGCTACA
AGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGG
TCAACTACAAGGACTACCAGGCCGTCACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCG
GCTACCTCGCGCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCA
TCGGCAAGAGCGCCGTCACCAGCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGC
GCATCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGC
TCTATGCCAACTCCGCCCACGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCAC
CCTTCTCTATGTTGTCTTCGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTC
ATCGAGGCCGTCTACCTGCGCACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCT
TCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTG
GGCTGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCAC
AAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGG
CCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACG
AGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGCAGCGCCCTGGCC
ACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGC
CGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTCGTGCACTGGCCCGACCGCCCCAT
GGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCCA
GGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCTACCGCTTCCTCAACTCCCACTCCGC
CTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCACCGCCTTCGACCGCATGAATCAAGA
CATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTTTCATGTTACACATGCATCT

GAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGCGGGCAGGGA
CACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATCAGCAGTTTGGGCAGCGG
GGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTCGG
GCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACG
GGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTCGCGTC
GGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAAGGGGGTCATCTTGCAGGT
CTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCA
GCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGGCCTTCATGAAAGCCTCCAATTG
CCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCCGCAGGACTTGCTAGAGAA
CTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTCGTTGTTGGCCAGCTGCA
CCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTCGGGGTTCTCCTTCAGCGCG
CGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCC
CGTGCAGGCACCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCG
GTGCACTCCCAGTTCTTGTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCG
GCCCATCATGGTGGTCAGGGTCTTGTTGCTAGTGAAGGTCAGCGGAATGCCGCGGTGCTCCTC
GTTGATGTACAGGTGGCAGATGCGGCGGTACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTT
GGCTTTCAGGTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAGTCATGATTTCCATACCCTTC
TCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCC
GCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATC
CGCACCGGGGGGGTAGCTGAAGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTC
GCTGTCCTGGCTGACGTCCTGCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGG
CAGCGGCGGCGGAGATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTA
TCTCTTCCTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCG
GAGGCGACGGGCTCTCGCCGCCGCGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCG
GGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAG
GGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCGCCGAC
GAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTCCGA
CGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGAC
GCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACA
GCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACC
TCCACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATCGTC
AAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTA
CGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCGAGC
CCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTACCACA

TCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCGACGCCC
TTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAAGAT
CTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAG
AGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTC
AAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTCTGAACCTGCCCCCCAAAGTCATG
AGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGCCCATCTCCGAGGACGAGGGCAT
GCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGT
CCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTCATGATGGCCGTGGTCCTGGTG
ACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGAGACCCTGCGCAAGGTCGA
GGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCAACGT
GGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGT
GCTGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCT
ACCTCTGCCACACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAAC
CTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAG
CGCACCACCGCCTCGGACCTGGCCGACCTCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCG
CAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCAAAACTTTCGCTCTTTCATCCTCGAAC
GCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCC
GCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCC
TACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTG
CAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGAT
CATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGGTC
TGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACC
ATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCT
GCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAAT
TCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCC
GGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGCCGCCCGTGGAGG
ATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGG
GACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGC
AGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCA
AGCAGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGA
CGAGACCGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACA
AGTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCT
CCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCATTA
CTACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAGACCA

GCAGAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCGCGGCG
AACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCCACCCTCTATGCCATCTTC
CAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACC
CGCAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCT
CTCTTCAACAAGTACTGCGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCCAGTCGCAGAAAA
AGGCGGGAATTACGTCACCTGTGCCCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAA
GAGATTCCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGCCTGGCCGCCGGTGCCGC
CCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGCGATGATCTCACGGGTGAA
TGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGTCAGCGCTCACCGCCACGCCCCG
CAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGCCCAC
GACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCT
GGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCC
GGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGAC
GGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGCCTCGTCAGGCCGTCCTGACT
TTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACTCTCCAGTTCGTGGAGGAG
TTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCCACTACCCGGACGAGTTCA
TCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATTGAATGTCCCATGGTGGCG
CAGCTGACCTAGCTCGGCTTCGACACCTGGACCACTGCCGCCGCTTCCGCTGCTTCGCTCGGG
ATCTCGCCGAGTTTGCCTACTTTGAGCTGCCCGAGGAGCACCCTCAGGGCCCGGCCCACGGAG
TGCGGATCGTCGTCGAAGGGGGCCTCGACTCCCACCTGCTTCGGATCTTCAGCCAGCGTCCGA
TCCTGGTCGAGCGCGAGCAAGGACAGACCCTTCTGACTCTGTACTGCATCTGCAACCACCCCG
GCCTGCATGAAAGTCTTTGTTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTA
CTCCGGACTTCCGTGTGTTCCTGAATCCATCAACCAGTCTTTGTTCTTCACCGGGAACGAGACC
GAGCTCCAGCTCCAGTGTAAGCCCCACAAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCG
ATCGCCGTTGTCAACCACTGCGACAACGACGGAGTCCTGCTGAGCGGCCCTGCCAACCTTACT
TTTTCCACCCGCAGAAGCAAGCTCCAGCTCTTCCAACCCTTCCTCCCCGGGACCTATCAGTGCG
TCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACCACAGCGTCGCTCCCCGCTAC
TAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACGGCCACAATACATGCCCATATTAG
ACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCTATTAGTTACTTCAATCTAACCGGCGG
AGATGACTGACCCACTGGCCAACAACAACGTCAACGACCTTCTCCTGGACATGGACGGCCGCG
CCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCAGCAGGAGAGAGCCGTCAAGGAG
CTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCTTCTGCCTGGTGAAACAGGCC
AAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTACGAGCTCCTGCAGCAGCGC
CAGAAGTTCACCTGCCTGGTCGGAGTCAACCCCATCGTCATCACCCAGCAGTCTGGCGATACC

AAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCACACTCTGATCAAGACCCTC
TGCGGCCTCCGCGACCTCCTCCCCATGAACTAATCACCCCCTTATCCAGTGAAATAAAGATCAT
ATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATAAAGATACAATCATATTGA
TGATTTGAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCAT
GTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTGCAGGCCCCGGCGGGCT
GCAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTATC
TTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCCGTCTACCCCTACG
ATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGTCTCTTCAGATGGATTCCA
AGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCAAGAACGGGG
AAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCGGGAAAACTCATCTCCAACA
CGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCA
CCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAATATACTGAGAAC
AAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGG
CAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAAGCTTACCTTAGACAGA
GGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTTTAAAATT
TGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTGGAAGCAGTAGTACAGAA
ACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTTGACAGTAC
AGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACACCTGATCCA
TCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAACTAACACTTTGCTTGACTAAATGTGG
TAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCATTACTG
GCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGTGTTCTTTTAACAGAACAT
TCTACACTAAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATATACCA
ATGCTGTAGGATTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATA
ATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGCTTCTCACTATAACCCTC
AATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACACCTGGACTAATGGAAG
CTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATCGCCCAAGAATGAACAC
TGTATCCCACCCTGCATGCCAACCCTTCCCACCCCACTCTGTGGAACAAACTCTGAAACACAAA
ATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGATTCGAGCAGTTATTTTTCCTC
CACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCCGCACAGCCTTGAACATCTGAATGCC
ATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCTCGGG
TCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGA
GGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAAT
CATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGC
CGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCC

CACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCA
GGTCGCTGCAGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCT
CCAGCCGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTA
AATCAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCG
GTTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCG
GAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAAT
GGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGG
CACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCAT
ATCCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCA
CAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCA
GAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGG
CGGGACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTGCTTTCGGACATTTTCGTAC
TTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGCGGCGGTCTCGGCGCTT
GGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGCC
TCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGTGGAATGGG
CCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAG
GAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGA
TGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCT
CGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAG
CGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCATGTTACACTCCTGCACCATCCCCAGATAA
TTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGTTCGTGAGGTAAATCCAAGCCAGCCATGAT
AAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAAGATAT
TCTGCTCCTGGTTCACCTGCAGCAGATTGACAAGCGGAATATCAAAATCTCTGCCGCGATCCCT
GAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTCTCCGAAATTTTTAGCCATAG
GACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAACCGAAGTCCTCCCCAGTGAG
CATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATAACT
GGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACG
TTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGC
TGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATC
GTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGACCCTCGTAAAAATTGTCGCTA
TGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGATTCGACAAGATGAAT
ACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAAGCGCCCGAGGAAGCAATAAGGCACTA
CAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGC
GTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTCCAGGTACACA

TACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGA AAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGACG TAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAACC GGTGACACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCAAAACTGCCGTCATTTCCGGGT TCCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAAAACGTCACCCGCC CCGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACC TCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG
ChAdV68.5WTnt.GFP (SEQ ID NO:13); AC_000011.1 con secuencias E1 (nt 577 a 3403) y E3 (nt 27.125- 31.825) eliminadas; nucleótidos de ATCC VR-594 correspondientes sustituidos en cinco posiciones; GFP informante bajo control del promotor/potenciador de CMV insertado en el lugar de E1 eliminada
CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTG GGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGT TTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAAC GAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTT TCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGGAAGTGAAAAT CTGAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGA TTACGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGGTGTCAAAGTCCGGT GTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCC ACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACACTTTGAAA GTAGGGATAACAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGTCATTAGTTCA TAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCC AACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTT TCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT CATATGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCC AGTACATGACCTTACGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACC ATGgTGATGCGGTTTTGGCAGTACACCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCC AAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGT CTATATAAGCAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGA CCTCCATAGAAGACAGCGATCGCGccaccATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGT GGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCG AGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAG CTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC

TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAG
GAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAG
GGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATC
CTGGGGCACAAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAG
AAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTC
GCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCA
CTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCT
GCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTtTACAAGTAGtgaGTTT
AAACTCCCATTTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTT
GGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGC
TTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTT
CAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTAAAA
TAACTATAACGGTCCTAAGGTAGCGAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGA
GGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATGAGCGGAAGCGGCTCC
TTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCCTGGGCGGGAGTGCGT
CAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGCGAACTCTTCAACCCT
GACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCAGCTGCTGCATCTGC
CGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACT
CGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCCAGC
TCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAG
ACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACG
GTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTCGCGCGCGGTAGGCCCTGGACC
ACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAGGACCCGGTAGAGGTGGGCTTGGA
TGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCTCCATTGCAGGGCCTCG
TGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGTGTTGCACA
ATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTGT
TGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTG
GCGATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGCAGGACCACCAGCACGGTGTAT
CCGGTGCACTTGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACG
CCTTTGTGCCCGCCCAGGTTTTCCATGCACTCATCCATGATGATGGCGATGGGCCCGTGGGCG
GCGGCCTGGGCAAAGACGTTTCGGGGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCA
TCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCG
ATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGG
ATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGC

CGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGA
TGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCC
ACCTCGTTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCT
CCCCCCAGGGATAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGC
CATGGGCATTTTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTG
CTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAG
GGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCG
CGTCAGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTG
CGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAG
GTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTT
ACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGG
CGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCGCAGACGGTCTCGCA
CTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTCCCGCCGTTCTTTTT
GATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCC
GTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGCGGTCCTCCTCGTAG
AGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAAGGAGGCCACGTG
GGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATGTC
CCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCC
GGCCGGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACTGTCTTCCGGATCGCTGTC
CAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAG
GTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAG
AGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGTTGTCGAGCTTGGTGGCGAAGGAGC
CGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGCGCATGGTCTGGTTTTTTTCCTTGTCGG
CGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCCACGCACTTCCATTCGGGGA
AGACGGTGGTCAGCTCGTCGGGCACGATTCTGACCTGCCAGCCCCGATTATGCAGGGTGATGA
GGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGCCC
TTGCGCGAGCAGAAGGGGGGCAGGGGGTCCAGCATGACCTCGTCGGGGGGGTCGGCATCGA
TGGTGAAGATGCCGGGCAGGAGGTCGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCC
AGGGCAGCTTGCCATTCGCGCACGGCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCA
GGGCATGGGATGGGTAAGCGCGGAGGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCT
CCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAG
TCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTC
GGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGA
AGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCT

TGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTC
CTGGATGATGTCATACTTGAGCTGTCCCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTT
CGCGGTCCTTCCAGTACTCTTCGAGGGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCA
TGTAGAACTGGTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCT
GGGCGGCCTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTCCCTGACCATGACCTTGAGG
AACTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCAGAGCTGGAAGTCCGTGCGC
TTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGCGG
GGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGATGACCTGG
GCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAAT
CGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGGTC
GCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAA
GTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTGCTGCCC
GACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCC
ATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCCGGAGAGTTTC
ATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTGTAGGTTTCCACA
TCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACTGGATCTCC
TGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAA
CACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGC
TGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGGGAAGTGGAGTCGTGGCGCC
TGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTC
ATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGGCGCGAGCGGGTCGGAGAGCGA
GGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTGAGACGCTGCGGAGTCAGGTCAGT
GGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGGCGCGCGGGAGGTCCAGAT
GGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTCGATGGCTTGCAGGGTCCCGTGCCCCT
GGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGGCGGTGCCTCT
TCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCG
GAGGCAGGGGCGGCAGGGGCACGTCGGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCG
GAGAAGACTGGCGTGAGCGACGACGCGACGGTTGACGTCCTGGATCTGACGCCTCTGGGTGA
AGGCCACGGGACCCGTGAGTTTGAACCTGAAAGAGAGTTCGACAGAATCAATCTCGGTATCGTT
GACGGCGGCCTGCCGCAGGATCTCTTGCACGTCGCCCGAGTTGTCCTGGTAGGCGATCTCGGT
CATGAACTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGC
GAGGTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCATGCCCGCCTCGTTCCAGAC
GCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCATGACCACCTGGGCGAGGTTGA
GCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTG

GTGGCGATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACG
TCGCCCAGCGCCTCCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGG
GAGTTGCGCGCCGAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGGC
GCGCACCTCGCGCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCCTCCTCCACTAAC
ATCTCTTCTACTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCG
GCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCT
CGGTGACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAG
GTGGCCGGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCC
CCGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAA
CGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCAT
GTTGGTTGGGAGCGGGGCGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGA
CGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAGACGGTC
GGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGCATGAGCCG
CTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGCGCGTGAGCCCGAAGCCGCGCT
GGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGCGAGGATGGCTTGCTGGATCTG
GGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGT
GTAGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGT
GGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGTCGTTGCAGGTGCGCACCAGG
TACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGC
GGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGTAGCCGTAGATGTACCTGGACA
TCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGAACTCGCGGACGCGGTTCCAGAT
GTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGTCTGGCCCGTGAGGCGCGCGCAGT
CGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAGCGGCTCGACTCCGTGGCCTGGAGG
CTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTCGAATCTCGAATCAGGCTGGAGCCGC
AGCTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGA
GGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCC
GACCGCGATGGCTCGCTGCCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCC
CCGGTTCGAGGCCGGCCGGATTCCGCGGCTAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAG
ACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGC
CAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCT
CCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGC
CGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGG
CGCGCCTGGGGGCGTCGTCGCCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCG
CGAGGCCTACGTGCCCAAGCAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGA

TGCGCGCGGCCCGGTTCCACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGT
GCTGAGGGACGAGGATTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCAC
GTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCA
AAAATCCTTCAACAACCACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGAT
GCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGC
AGCTGTTCCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATA
TCACCGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGC
AGGAGCGCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTG
GGCAAGTACTACGCTAGGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAG
ATCGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTG
TACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGGCGGCGCGAGCTGAGCGACCA
GGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGGGACCGAGGGGGAGAGCTACT
TTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCCTTGGAGGCGGCGGCAGG
ACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGAT
GGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATGCGGGC
GGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAAC
GCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGG
CTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACGCACGAGAAGGTCCT
GGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGACGAGGCCGGCCTGGTGT
ACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAACGTGCAGACCAACCTGGAC
CGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCGCGAGTCCAA
CCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCCGGG
GCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGA
GCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCG
TGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCCTGTGGGGCGTGCAGGCCCCGGTCGGG
GACCGCGCGACGGTGTCGAGCCTGCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGC
CCCCTTCACGGACAGCGGCAGCATCAACCGCAACTCGTACCTGGGCTACCTGATTAACCTGTAC
CGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTACCAGGAGATCACCCACGTGAG
CCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAA
CCGGTCGCAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTT
ACGTGCAGCAGAGCGTGGGCCTGTTCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTC
GACATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAA
CTGATGGACTACTTGCATCGGGCGGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGA
ATCCCCACTGGCTCCCGCCGCCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAAT

GACGGGTTCCTGTGGGACGATGTGGACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGA
GCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCCCGTCCTCGGCGCTGTCCGGCCGCG
AGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCCCGAGCTTGCCCTTCTCGCTG
AACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGAAGAGGA
GTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAA
AGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCG
GGCGTCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGGTGGCACGACAGG
CAGCGGGGACAGATGTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGTTGGACTTGGG
TGGGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACC
GAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTTGT
TGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCGTGAT
GCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGCCCCCG
CGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCTTGTAC
GATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAG
AACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCC
AGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATG
CACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATG
GTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAG
CTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCTCGGTGACCATGACCATCG
ACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGCGGCAGAACGGGGTGCTGG
AGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGGACCCCGTGACC
GAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCCC
GGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTCGCAAGAGGCA
GCCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCT
CCTGGATGTCGACGCCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGT
AGCTACCGCCTCTACCGAGGTCAGGGGCGATAATTTTGCAAGCGCCGCAGCAGTGGCAGCGG
CCGAGGCGGCTGAAACCGAAAGTAAGATAGTCATTCAGCCGGTGGAGAAGGATAGCAAGAACA
GGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTACCGCAGCTGGTACCTAGCCTACAA
CTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACCACCTCGGACGTCACCT
GCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCAAGACCCGGTCACCTTCCGCT
CCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGA
GCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACG
TCTTCAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCG
TCAGTGAAAACGTTCCTGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGG

GAGTCCAGCGCGTGACCGTTACTGACGCCAGACGCCGCACCTGCCCCTACGTCTACAAGGCCC
TGGGCATAGTCGCGCCGCGCGTCCTCTCGAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCC
CAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAAC
GCTCCACGCAACACCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCAAG
GGCCGCGTGCGGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCGC
GCAACTACACCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTG
GCcGACGCGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACC
GGAGCACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACG
CAGGGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGG
AGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGGA
ACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCCCCT
CGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCCAAGC
GCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGA
AGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTG
ATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCGGCGGCGCGTGCAGTGGCGC
GGGCGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCG
CTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCA
GGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAG
AGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCGAGCCTCAAGCCCGTGACCTTG
CAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAAGCGCGAGGGCGAGGATCTGT
ACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGA
AGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGG
CCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACGGAGCCCATGGAAACGCAGACCGAGC
CCATGATCAAGCCCAGCACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTC
CTAGTCGAAGACCCCGGCGCAAGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTG
CATCCTTCCATCATCCCCACGCCGGGCTACCGCGGCACGCGCTTCTACCGCGGTCATACCAGC
AGCCGCCGCCGCAAGACCACCACTCGCCGCCGCCGTCGCCGCACCGCCGCTGCAACCACCCC
TGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGCCGCGCGCG
CGCTACCACCCGAGCATCGCCATTTAAACTTTCGCCtGCTTTGCAGATCAATGGCCCTCACATGC
CGCCTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGG
GAACGGGATGCGTCGCCACCACCACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGA
GGCTTCCTGCCCGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGC
TTCCGTGGCGGTGCAGGCCTCTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACC
aATGGACTCTGACGCTCCTGGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTT

CGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACC
AGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGG
TCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGAT
AAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGG
GTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCC
GCCCGCCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGG
GGCGAGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGC
CCCCGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCC
ACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCGC
CCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGGCA
CCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAGTGCAGA
GTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGTA
TGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTT
GCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGAC
GCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGT
CTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATGTGACCACCGACCGCAG
CCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACACCTACTCGTACAAAGT
GCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCCAGCACCTACTTTGACAT
CCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCGGCACCGCCTACAACAGTCT
GGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAAGCCGATGGTGAAACTGCCAC
AGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTC
AACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCA
AGTGGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTT
AAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGG
TCAGGCAAATGTGAAAACAGGAACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTG
ACAACAGAAGTGCGGCTGCTGCTGGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGA
TTTGGAAACTCCAGATACCCATATTGTATACAAAGCAGGCACAGATGACAGCAGCTCTTCTATTA
ATTTGGGTCAGCAAGCCATGCCCAACAGACCTAACTACATTGGTTTCAGAGACAACTTTATCGG
GCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTGAA
TGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCTG
GGTGACAGAACCCGGTATTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGC
GCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTT
GGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAG
ATGACAGTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACAT

CCAAGCCAACCTGTGGAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTAC
AAGTACACGCCGGCCAATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCC
GGGTGGTGGCGCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGAT
CCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATG
CTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCA
AGAGCCTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACA
TGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCA
GCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCA
TGCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTA
CCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCCTT
CCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCG
ACCCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAACCACAC
CTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGCCCGGCAACGACCGGCTCCT
GACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTACAACGTGGCCCAGT
GCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATCGGCTACCAGG
GCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCAT
GAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTCACCCTGGCCTACC
AGCACAACAACTCGGGCTTCGTCGGCTACCTCGCGCCCACCATGCGCCAGGGCCAGCCCTACC
CCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACCAGCGTCACCCAGAAAAAGT
TCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGC
TCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCACGCGCTAGACATGAATTTCGA
AGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTTCGAAGTCTTCGACGTCGTCCGA
GTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCCCCTTCTCGGCCGG
TAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGCAGGAG
CTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGC
TTCCCGGGATTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGA
GACCGGGGGCGAGCACTGGCTGGCCTTCGCCTGGAACCCGCGCTCGAACACCTGCTACCTCT
TCGACCCCTTCGGGTTCTCGGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCC
TGCTGCGCCGCAGCGCCCTGGCCACCGAGGACCGCTGCGTCACCCTGGAAAAGTCCACCCAG
ACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTC
GTGCACTGGCCCGACCGCCCCATGGACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCC
CAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCT
ACCGCTTCCTCAACTCCCACTCCGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCA
CCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAAC

AGCACTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGT
CTCGGCATGGCCCGCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACT
CGGGGATCAGCAGTTTGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTC
AGTTGCAGGGCGCCCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTT
CTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCT
TCACGCTCGCCAGCACCGTCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCA
TCCCGAAGGGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGT
TGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACA
TGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGA
AGACCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAG
CGCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGC
CCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCAT
GTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCCCTCGGCCTCGGTGCA
CCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTGGGCGATCTGGGAATGCG
CGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCTTGTTGCTAGTGAAGG
TCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTCGC
CCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCACGCGGTAGCGGTCCATCA
GCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTTCTT
CACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTCGTCCAGGGTCTCAAAGCT
CCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGGTAGCTGAAGCCCACGGCCGCCAGCT
CCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCTGCAGGACCACATGCTTGGT
CTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGTTGGAGATGGCGAGGGGGAGC
GCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGTCCGAGGCCACGCGGCGGTAGGT
ATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACTTGGCGGATGGC
TGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTCCG
CGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACC
TCGCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCG
CCGCCCAGCCCCGCCACCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCAT
CGAGATTGACCTGGGCTATGTGACGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTT
CACAAGAAGAGATACACCAAGAACAGCCAGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTG
GGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCAT
CTGGCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAG
CGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGC
GCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGC

CCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGC
CAACCGCACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGC
CTCCTTGGAAGAGGTTCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAA
CGCTCTGCAAGGAGAAGGAGGAGAGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCG
ACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTC
TGAACCTGCCCCCCAAAGTCATGAGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGC
CCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAG
CAGCTGGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTC
ATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGC
GGAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCA
GGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGA
GAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACA
TCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGGCATGGGCGTGTGGCAGC
AGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACCTCAAGGGTC
TGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTCATTTTCCCCGAGC
GCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCAAAACT
TTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGG
ACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGC
GCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTCAGCGGCGAGGGCCTG
CTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAG
CTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGG
TTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGCTGTGGACCTCGGCCTACTTGCGCAA
GTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGAGGACCAATCCCATCCGCC
CAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAATTGCAAGCCAT
CCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAGAC
CGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTG
GAGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGA
GGAGATGGAGGAAGACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGG
AGGAAGACGAGGAGGAGGCAGAGGAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTC
CTCGGCGGGGGAGAAAGCAAGCAGCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTC
GACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGA
AGGAGCGGCAGGGATACAAGTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAG
GCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTT
CCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAG

CAGCAGCAGAAAAAGACCAGCAGAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGT
GGACTGAGGATCGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCC
CACCCTCTATGCCATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCG
TTCTCTGCGCTCGCTCACCCGCAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACT
CTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCGCGCTCACTCTTAAAGAGTAGCCCGCG
CCCGCCCAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGCCCTTCGCCCTAGCCGCCTCC
ACCCATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGC
CTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGC
GATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGTCAGC
GCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCA
GGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGAC
TAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTCAGGGTA
TAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCG
CTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTCCTTCACGCCT
CGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGTGGCATCGGCACT
CTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTCCCCCGGCC
ACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATT
GAAACTAATCACCCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAA
AATAATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAACAAAAAAATAAAG
AATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCC
CCTCTTCCCAGCTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACTTCCTCCACACGCTGAAGG
GGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGT
CCGGGTGGATGATGACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCC
CTTCATCAACCCCCCCTTCGTCTCTTCAGATGGATTCCAAGAGAAGCCCCTGGGGGTGTTGTCC
CTGCGACTGGCCGACCCCGTCACCACCAAGAACGGGGAAATCACCCTCAAGCTGGGAGAGGG
GGTGGACCTCGATTCCTCGGGAAAACTCATCTCCAACACGGCCACCAAGGCCGCCGCCCCTCT
CAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATT
ATCCTTACAAGTTTCTCCACCATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGG
TTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACAT
TTGATACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCA
ATTGAAAGCAACATAAGCTGGGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACAT
TGGAAATGGGTTAGAGTTTGGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATC
CAAGTTAAACTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAG
AAGACGATAAACTCACTTTGTGGACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAA

AATGATGCAAAACTAACACTTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGT
CTTAGTTGTAGGAAGTGGAAACCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTT
CTACGTTTTGATGCAAACGGTGTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTAT
AGGCAGGGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAA
AAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATG
GAGATGTTTCAAAACCTATGCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACA
TATTCAATGTCATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAA
CTCTTATACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCTT
CCCACCCCACTCTGTGGAACAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGAT
TCAACAGTTTTACAGGATTCGAGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAATACACCA
CCCTCTCCCCCCCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTC
CACGTTCCACACAGTTTCAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGG
GCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAGGATTGTCCTCGGTGGTCGGGATCAC
GGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCG
GTGGTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCA
GGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTG
GTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAG
AACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGG
ATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACA
CGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGGTTCACCACCTCCCGGTACCACATCA
CCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCG
CCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTA
CCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCAT
CTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCCAGGGCACGGGGAACTCTTGCA
GGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGG
TATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCAC
AGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGAC
CGTGTCATGATGCAGTTGCTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCG
CTGCACACCGATCGCCGGCGGCGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAAC
AGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGC
CTGATGGCTCTGATCACATCGACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTT
GTTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCA
AACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTT
GGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCA

GCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATT
CCTCAATCATCATGTTACACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGA
TTCGAACTAGTTCcTGAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTC
CACCGGCATTCTTAAGCACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAG
ATTGACAAGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGT
AAGTACTCTTTCATATCCTCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCA
AGCCACAGTACAGATAAACCGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAA
GCATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTT
TAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAA
GTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAA
CATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGCAGGCC
ACGGGGTCTCCGGCGCGACCCTCGTAAAAATTGTCGCTATGATTGAAAACCATCACAGAGAGAC
GTTCCCGGTGGCCGGCGTGAATGATTCGACAAGATGAATACACCCCCGGAACATTGGCGTCCG
CGAGTGAAAAAAAGCGCCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAA
AGCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCC
TGCACAGGCAGCAAAGCCCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTT
ACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACC
CGCTCTCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCG
CCAAATAATCACACACGCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGC
ACTTCCTCAAACGCCCAAAACTGCCGTCATTTCCGGGTTCCCACGCTACGTCATCAAAACACGA
CTTTCAAATTCCGTCGACCGTTAAAAACGTCACCCGCCCCGCCCCTAACGGTCGCCCGTCTCTC
AGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAA
GTTTGAGGTATATTATTGATGATGG

XV.B. Evaluación del vector de administración del casete de neoantígeno de ChAd

XV.B.1. Métodos y materiales de evaluación del vector de ChAd

Transfección de células HEK293A mediante el uso de lipofectamina

[00812] El ADN para las construcciones ChAdV68 (ChAdV68.4WTnt.GFP, ChAdV68.5WTnt.GFP, ChAdV68.4WTnt.MAG25mer y ChAdV68.5WTnt.MAG25mer) se preparó y se transfectó en células HEK293 mediante el uso del siguiente protocolo.

[00813] Se digirieron 10 ug de ADN plásmido con PacI para liberar el genoma viral. El ADN fue purificado posteriormente mediante el uso de columnas de limpieza Micro GeneJet DNA (Thermo Fisher) de acuerdo con las instrucciones del fabricante para los fragmentos de ADN largos y se eluyó en 20 ul de agua previamente calentada; las columnas se dejaron a 37 grados durante 0,5-1 horas antes de la

etapa de elución.

[00814] Las células HEK293 se introdujeron en placas de 6 pocillos a una densidad celular de 10^6 células/pocillos 14-18 horas antes de la transfección. Las células se superpusieron con 1 ml de medio fresco (DMEM-hiFBS al 10 % con pen/estrep y glutamato) por pocillo. 1-2 ug de ADN purificado se utilizaron por pocillo en una transfección con dos veces el volumen en ul (2-4 ul) de Lipofectamina2000, de acuerdo con el protocolo del fabricante. 0,5 ml de medio OPTI-MEM que contiene la mezcla de transfección se agregaron al 1 ml de medio de crecimiento normal en cada pocillo y se dejó en las células durante la noche.

[00815] Los cultivos celulares transfectados se incubaron a 37 ºC durante al menos 5-7 días. Si las placas virales no fueron visibles en el día 7 luego de la transfección, las células se dividieron 1:4 o 1:6 y se incubaron a 37 ºC para monitorear el desarrollo de la placa. Alternativamente, las células transfectadas fueron recolectadas y sometidas a 3 ciclos de congelamiento y descongelamiento y los lisados celulares se utilizaron para infectar las células HEK293A y las células se incubaron hasta que se observaron placas de virus.

Transfección de vectores ChAdV68 en células HEK293A mediante el uso de fosfato de calcio y generación de la solución madre viral terciaria

[00816] El ADN para las construcciones ChAdV68 (ChAdV68.4WTnt.GFP, ChAdV68.5WTnt.GFP, ChAdV68.4WTnt.MAG25mer, ChAdV68.5WTnt.MAG25mer) se preparó y se transfectó en células HEK293 mediante el uso del siguiente protocolo.

[00817] Se cultivaron células HEK293 un día luego de la transfección a 10^6 células/pocillo de una placa de 6 pocillos en 5 % de BS/DMEM/ 1XP/S, 1XGlutamax. Se necesitaron dos pocillos por transfección. Dos o cuatro horas antes de la transfección el medio se cambió a medio fresco. El plásmido ChAdV68.4WTnt.GFP se linealizó con PacI. El ADN linealizado fue extraído luego con cloroformo de fenol y se precipitó con el uso de un décimo de volumen de acetato de sodio 3 M pH 5.3 y dos volúmenes de etanol al 100 %. El ADN precipitado se sedimentó mediante centrifugación a 12.000 xg durante 5 min antes de lavar 1 X con etanol al 70 %. El sedimento se secó al aire y se resuspendió en 50 µL de agua estéril. La concentración de ADN se determinó mediante el uso de un NanoDrop™ (ThermoFisher) y el volumen se ajustó hasta 5 µg de AND/50µL.

[00818] Se agregaron 169 µL de agua estéril a un tubo de microcentrífuga. 5 µL de 2M $CaCl_2$ se agregaron luego al agua y se mezclaron suavemente mediante pipeteo. 50 µL de ADN se agregaron en gotas a la solución de agua con $CaCl_2$. Veintiséis µL de $CaCl_2$ 2 M se agregaron luego y se mezclaron suavemente mediante pipeteo dos veces con una micropipeta. Esta solución final debe consistir en 5 µg de DNA en 250 µL de $CaCl_2$ 0,25 M. Un segundo tubo se preparó luego con 250 µL de 2XHBS (solución amortiguada Hepes). Mediante el uso de una pipeta estéril de 2 mL acoplada a un Pipet-Aid se burbujeó

aire lentamente a través de la solución 2XHBS. Al mismo tiempo la solución de ADN en la solución $CaCl_2$ 0,25 M se agregó en un gotas. El burbujeo continuó durante aproximadamente 5 segundos luego de la adición de la gota de ADN final. La solución se incubó posteriormente a temperatura ambiente durante hasta 20 minutos antes de agregarla a células 293A. 250 µL de la solución de fosfato de ADN/Calcio se agregó en gotas a una monocapa de células 293A que se cultivaron un día antes a 10^6 células por pocillo de una placa de 6 pocillos. Las células se regresaron a la incubadora y se incubaron durante la noche. El medio se cambió 24 h más tarde. Luego de las 72 h las células se dividieron 1:6 en una placa de 6 pocillos. Las monocapas se monitorearon diariamente mediante microscopía de luz para determinar la evidencia de efecto citopático (CPE). 7-10 días luego de la transfección se observaron placas virales y la monocapa se cultivó mediante pipeteo del medio en los pocillos para elevar las células. Las células recolectadas y el medio se transfirieron a un tubo de centrífuga 50 mL seguido de tres rondas de congelamiento descongelamiento (a -80 ºC y 37 ºC). El lisado posterior, denominado solución madre de virus primaria, se clarificó mediante centrifugación a velocidad total en una centrífuga de mesada (4300Xg) y una proporción del lisado 10-50 %) se utilizó para infectar células 293A en un matraz T25. Las células infectadas se incubaron durante 48 h antes de recolectar las células y el medio en CPE completo. Las células se recolectaron nuevamente, se congelaron descongelaron y se clarificaron antes de utilizar esta solución madre viral secundaria para infectar un matraz T150 sembrado con 1,5x 10^7 células por matraz. Una vez que se logró el CPE completo a 72 h el medio y las células se recolectaron y se trataron como con soluciones madre virales iniciales para generar una solución madre terciaria.

Producción en células 293F

[00819] La producción del virus ChAdV68 se llevó a cabo en células 293F cultivadas en medio 293 FreeStyleT^M (ThermoFisher) en una incubadora a CO_2 al 8 %. En el día de la infección, las células se diluyeron a 10^6 células por mL, con una viabilidad del 98 % y se utilizaron 400 mL por corrida de producción en matraces Shake de 1L (Corning). Se utilizaron 4 mL de la solución madre viral terciaria con un MOI objetivo de > 3,3 por infección. Las células se incubaron durante 48-72 h hasta que la viabilidad fue de 70 % como se midió mediante azul de tripano. Las células infectadas se recolectaron luego mediante centrifugación, centrífuga de mesada a velocidad total y se lavaron en 1XPBS, se recentrifugaron y luego se resuspendieron en 20 mL de Tris 10 mM pH 7,4. El sedimento celular se lisó mediante congelamiento descongelamiento 3X y se clarificó mediante centrifugación a 4300 Xg durante 5 minutos.

Purificación mediante centrifugación con CsCl

[00820] Se purificó el ADN viral mediante centrifugación con CsCl. Se realizaron dos corridas de gradiente discontinua. La primera para purificar el virus a partir de los componentes celulares y la segunda para refinar adicionalmente la separación a partir de los componentes celulares y separar

partículas defectuosas de las infecciosas.

[00821] Se agregaron 10 mL de CsCl 1,2 (CsCl 26,8 g disueltos en 92 mL de Tris 10 mM pH 8,0) a tubos de polialómero. Luego se agregaron cuidadosamente 8 mL de CsCl 1,4 (CsCl 53 g disueltos en 87 mL de Tris 10 mM pH 8,0) mediante el uso de una pipeta hacia el fondo del tubo. El virus clarificado fue cuidadosamente colocado en una capa en la superficie de la capa 1,2. De ser necesario se agregó más Tris 10 mM al balance de tubos. Los tubos se colocaron luego en un rotor SW-32Ti y se centrifugó durante 2 h 30 min a 10 ºC. El tubo se retiró luego a un gabinete de flujo laminar y la banda del virus se retiró mediante una aguja de calibre 18 y una jeringa de 10 mL. Se evitó cuidadosamente retirar el ADN y proteína de la célula hospedadora contaminante. La banda se diluyó luego al menos 2X con Tris 10 mM pH 8,0 y se colocó en capas como antes en un gradiente discontinuo como se describe anteriormente. La corrida se llevó a cabo como se describe anteriormente salvo que en este caso la corrida se llevó a cabo durante la noche. El día siguiente la banda se retiró con cuidado para evitar retirar cualquier banda de partículas defectuosas. El virus se dializó luego mediante el uso de un casete Slide-a-LyzerT^M (Pierce) contra el amortiguador ARM (Tris 20 mM pH 8,0, NaCl 25 mM, glicerol al 2,5 %). Esto se llevó a cabo 3X, 1 h por intercambio de amortiguador. El virus se separó en alícuotas para el almacenamiento a -80 ºC.

Ensayos virales

[00822] La concentración de VP se llevó a cabo mediante el uso de un ensayo de OD 260 en función del coeficiente de extinción de 1,1 x 10^{12} partículas virales (VP) es equivalente a un valor de absorción de 1 a OD260 nm. Las dos diluciones (1:5 y 1:10) de adenovirus se llevaron a cabo en un amortiguador de lisis viral (SDS al 0,1 %, Tris 10 mM pH 7,4, EDTA 1 mM). Se midió la OD en duplicado en ambas diluciones y la concentración de VP/mL se midió mediante la multiplicación del valor de OD260 X el factor de dilución X 1,1 x 10^{12}VP.

[00823] Se calculó una titulación de la unidad infecciosa (IU, por su siglas en inglés) mediante un ensayo de dilución limitante de la solución madre viral. El virus se diluyó inicialmente 100X en DMEM/NS al 5 %/1X PS y luego se diluyó posteriormente mediante el uso de diluciones de 10 veces hasta 1 x 10^{-7}. 100 µL de estas diluciones se agregaron a células 293A que se cultivaron al menos una hora antes de a 3e5 células/pocillos de una placa de 24 pocillos. Esto se lleva a cabo en duplicado. Las placas se incubaron durante 48 h en una incubadora de CO2 (5 %) a 37 ºC. Las células se lavaron luego con 1XPBS y luego se fijaron con metanol frío al 100 % (-20 ºC). Las placas se incubaron luego a -20 ºC durante un mínimo de 20 minutos. Los pocillos se lavaron con 1XPBS luego se bloquearon en 1XPBS/BSA al 0,1 % durante 1 h a temperatura ambiente. Un anticuerpo anti-Ad de conejo (Abcam, Cambridge, MA) se agregó a una dilución de 1:8000 en amortiguador de bloqueo (0,25 ml por pocillo) y se incubó durante 1 h a temperatura ambiente. Los pocillos se lavaron 4X con PBS 0,5 mL por pocillo.

Se agregó un anticuerpo anti-conejo de cabra conjugado con HRP (Bethyl Labs, Montgomery Texas) diluido 1000X por pocillo y se incubó durante 1 h antes de una ronda final de lavado. Se llevaron a cabo 5 lavados con PBS y los placas se desarrollaron mediante el uso de un sustrato de DAB (tetraclorhidrato de diaminobencidina) en solución salina amortiguada con Tirs (DAB 0,67 mg/mL en Tris 50 mM pH 7,5, NaCl 150 mM) con H_2O_2 al 0,01 %. Se desarrollaron pocillos durante 5 minutos antes del conteo. Se contaron las células en un objetivo de 10X mediante el uso de una dilución que proporcionó entre 4-40 células teñidas por campo de visión. El campo de visión que se utilizó fue una rejilla de 0,32 mm^2 que es equivalente a 625 por campo de visión en una placa de 24 pocillos. La cantidad de virus infeccioso/mL puede determinarse mediante la cantidad de células teñidas por rejillas multiplicada por la cantidad de rejillas por campo de visión multiplicado por un factor de dilución factor 10. De forma similar, cuando se trabaja con células que expresan GFP puede usarse fluorescencia en lugar de tinción de la cápside para determinar la cantidad de viriones que expresan GFP por mL.

Inmunizaciones

[00824] Se inyectaron ratones hembra C57BL/6J y ratones hembra Balb/c con $1x10^8$ partículas virales (VP) de ChAdV68.5WTnt.MAG25mer en una inyección intramuscular bilateral de 100 uL de volumen (50 uL por pata).

Disociación de esplenocitos

[00825] Se agruparon los bazos y ganglios linfáticos para cada ratón en 3 mL de RPMI completo (RPMI, FBS al 10 %, penicilina/estreptomicina). Se realizó una disociación mecánica mediante el uso del Disociador gentleMACS (Miltenyi Biotec), siguiendo el protocolo del fabricante. Se filtraron las células disociadas a través de un filtro de 40 micrómetros y se lisaron los glóbulos rojos con solución amortiguadora de lisis de ACK (NH_4Cl 150 mM, $KHCO_3$ 10 mM, Na_2EDTA 0,1 mM). Las células se filtraron nuevamente a través de un filtro de 30 micrómetros y luego se volvieron a suspender en RPMI completo. Se contaron las células en el citómetro de flujo Attune NxT (Thermo Fisher) mediante el uso de tinción con yoduro de propidio para excluir células muertas y apoptóticas. Luego, las células se ajustaron a la concentración adecuada de células hepáticas para un análisis posterior.

Análisis immunospot ligado a enzimas (ELISPOT) ex vivo

[00826] Se realizó un análisis ELISPOT de acuerdo con lineamientos de armonización ELISPOT {DOI: 10.1038/nprot.2015.068} con el kit ELISpotPLUS de IFNg de ratón (MABTECH). Se incubaron $5x10^4$ esplenocitos con 10 uM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNg de 96 pocillos. Se desarrollaron manchas mediante el uso de fosfatasa alcalina. Se marca una duración de 10 minutos para la reacción, tras cuyo plazo se pone fin a la reacción al colocar la placa bajo agua del grifo. Se contaron las manchas mediante el uso de un AID vSpot Reader Spectrum. Para el análisis ELISPOT, los pocillos con una saturación > 50 % se registraron como

«cantidad excesiva para contar". Las muestras con una desviación de los pocillos replicados > 10 % se excluyeron del análisis. Luego, se corrigieron los conteos de manchas con respecto a la confluencia de pocillos mediante el uso de la fórmula: conteo de manchas + 2 x (conteo de manchas x % de confluencia /[100 % - % de confluencia]). Se corrigió el fondo negativo mediante la sustracción de los conteos de manchas en los pocillos de estimulación peptídica de caracter negativo de los pocillos estimulados con antígenos. Finalmente, los pocillos etiquetados como «cantidad excesiva para contar» se ajustaron al valor corregido observado más alto, con redondeo hasta el múltiplo de cien más cercano.

XV.B.2. Producción de partículas de administración viral ChAdV68 luego de la transfección de ADN

[00827] En un ejemplo, se transfectó ADN de ChAdV68.4WTnt.GFP (Figura 21) y ChAdV68.5WTnt.GFP (Figura 22) en células HEK293A y se observó la replicación del virus (placas virales) 7-10 días luego de la transfección. Se visualizaron placas virales de ChAdV68 mediante el uso de microscopía de luz (Figura 21A y 22A) y fluorescente (Figura 21 B-C y Figura 22 B-C). La GFP denota una producción de partículas de administración viral de ChAdV68 productiva.

XV.B.3. Expansión de las partículas de administración viral de ChAdV68

[00828] En un ejemplo, se expandieron los virus ChAdV68.4WTnt.GFP, ChAdV68.5WTnt.GFP y ChAdV68.5WTnt.MAG25mer en células HEK293F y se produjo una solución madre de virus purificada 18 días luego de la transfección (Figura 23). Las partículas virales se cuantifican en las soluciones madre del virus ChAdV68 purificadas y se compararon con el adenovirus tipo 5 (Ad5) y las soluciones madre virales de ChAdVY25 (un ChAdV estrechamente relacionado; Dicks, 2012, PloS ONE 7, e40385) producida mediante el uso del mismo protocolo. Las titulaciones virales de ChAdV68 fueron comparables a Ad5 y ChAdVY25 (Tabla 7).

Tabla 7. Producción de vectores adenovirales en células de suspensión 293F

Construcción	**VP promedio/célula+/- SD**
Vectores de Ad5 (vectores múltiples)	2,96e4+/- 2,26e4
GFP de Ad5	3,89e4
GFP de chAdY25	1,75e3+/-6,03e1
ChAdV68.4WTnt.GFP	1,2e4 +/-6,5e3
ChAdV68.5WTnt.GFP	1,8e3
ChAdV68.5WTnt.MAG25mer	1,39e3+/-1,1e3

*SD es solamente descrita cuando se llevaron a cabo las múltiples corridas de producción

XV.B.4. Evaluación de la inmunogenicidad en modelos tumorales

[00829] Se evaluaron antígenos tumorales de ratón que expresan el vector C68 en estudios de inmunogenicidad de ratón para demostrar que el vector de C68 produce respuestas de linfocitos T. Las respuestas de linfocitos T al epítopo de MHC clase I SIINFEKL (SEQ ID NO: 57) se midieron en ratones hembra C57BL/6J y el epítopo de MHC clase I AH1-A5 (Slansky et ál., 2000, Immunity13:529-538) se midió en ratones Balb/c. Como se muestra en la Figura 29, se midieron las respuestas de linfocitos T fuertes en comparación con el control luego de la inmunización de ratones con ChAdV68.5WTnt.MAG25mer. Se observaron las respuestas inmunitarias celulares promedio de 8957 o 4019 células formadoras de manchas (SFC) por 10^6 esplenocitos en ensayos ELISpot cuando se inmunizaron ratones C57BL/6J o Balb/c con ChAdV68.5WTnt.MAG25mer, respectivamente, 10 días luego de la inmunización.

XVI. Vector de administración del casete de neoantígenos de alfavirus

XVI.A. Materiales y métodos de evaluación del vector de administración de alfavirus

Transcripción in vitro para generar ARN

[00830] ***Para evaluación in vitro:*** ADN plásmido se linealizó mediante la digestión de la restricción con PmeI, columna purificada de acuerdo con el protocolo del fabricante (kit de limpieza de ADN GeneJet, Thermo) y se utilizó como plantilla. Se llevó la transcripción in vitro mediante el uso del sistema de producción de ARN a gran escala RiboMAX (Promega) con el análogo de casquete m^7G (Promega) de acuerdo con el protocolo del fabricante. El ARNm se purificó mediante el uso del kit RNeasy (Qiagen) de acuerdo con el protocolo del fabricante.

[00831] ***Para estudios in vivo:*** Se generó el ARN y se purificó mediante TriLInk Biotechnologies y se tapó con Cap1 enzimática.

Transfección de ARN

[00832] Las células HEK293A se cultivaron a 6e4 células/pocillo para 96 pocillos y 2e5 células/pocillos para 24 pocillos, ~16 horas antes de la transfección. Las células se transfectaron con ARNm mediante el uso de lipofectamina MessengerMAX (Invitrogen) y según el protocolo del fabricante. Para 96 pocillos, se utilizaron 0,15 uL de lipofectamina y 10 ng de ARNm por pocillo, y para 24 pocillos, se utilizaron 0,75 uL de lipofectamina y 150 ng de ARNm por pocillo. Se utilizó un ARNm que expresa GFP TriLink Biotechnologies) como un control de transfección.

Ensayo de luciferasa

[00833] Se llevó a cabo un ensayo indicador de luciferasa en placas de 96 pocillos de paredes blancas con cada condición en triplicado mediante el ensayo de luciferasa ONE-Glo (Promega) de acuerdo con

el protocolo del fabricante. Se midió la luminiscencia mediante el uso del SpectraMax.

qRT-PCR

[00834] Las células transfectadas se enjuagaron y sustituyeron con medio fresco 2 horas después de la transfección para eliminar cualquier ARNm sin transfectar. Las células se recolectaron luego en diversos momentos en RLT más amortiguador de lisis (Qiagen), se homogenizaron mediante un QiaShredder (Qiagen) y el ARN se extrajo mediante el uso del kit RNeasy (Qiagen), todo de acuerdo con el protocolo del fabricante. Se cuantificó el ARN total mediante el uso de un Nanodrop (Thermo Scientific). Se llevó a cabo una qRT-PCR mediante el uso del kit RT-PCR de una etapa para sondas Quantitect (Qiagen) en el qTower^3 (Analytik Jena) de acuerdo con el protocolo del fabricante, mediante el uso de 20 ng de ARN total por reacción. Cada muestra se ejecutó en triplicado para cada sonda. Se utilizaron actina o GusB como genes de referencia. Se generaron cebadores/sondas a medida mediante IDT (Tabla 8).

Tabla 8. Sondas/cebadores de qPCR

Objetivo			SEQ ID NO:
Luci	Cebador1	GTGGTGTGCAGCGAGAATAG	142
	Cebador2	CGCTCGTTGTAGATGTCGTTAG	143
	Sonda	/56-FAM/TTGCAGTTC/ZEN/TTCATGCCCGTGTTG/3IABkFQ/	144
GusB	Cebador1	GTTTTTGATCCAGACCCAGATG	145
	Cebador2	GCCCATTATTCAGAGCGAGTA	146
	Sonda	/56-FAM/TGCAGGGTT/ZEN/TCACCAGGATCCAC/3IABkFQ/	147
ActB	Cebador1	CCTTGCACATGCCGGAG	148
	Cebador2	ACAGAGCCTCGCCTTTG	149
	Sonda	/56-FAM/TCATCCATG/ZEN/GTGAGCTGGCGG/3IABkFQ/	150
MAG-25mer	Cebador1	CTGAAAGCTCGGTTTGCTAATG	151
Conjunto1	Cebador2	CCATGCTGGAAGAGACAATCT	152
	Sonda	/56-FAM/CGTTTCTGA/ZEN/TGGCGCTGACCGATA/3IABkFQ/	153
MAG-25mer	Cebador1	TATGCCTATCCTGTCTCCTCTG	154

Conjunto2	Cebador2	GCTAATGCAGCTAAGTCCTCTC	155
	Sonda	/56-FAM/TGTTTACCC/ZEN/TGACCGTGCCTTCTG/3IABkFQ/	156

Modelo tumoral B16-OVA

[00835] Los ratones C57BL/6J recibieron la inyección en el flanco abdominal izquierdo inferior de 10^5 células B16-OVA/animal. Se dejó que los tumores se desarrollaran durante 3 días antes de la inmunización.

Modelo tumoral CT26

[00836] Los ratones Balb/c recibieron la inyección en el flanco abdominal izquierdo inferior de 10^6 células CT26/animal. Se dejó que los tumores se desarrollaran durante 7 días antes de la inmunización.

Inmunizaciones

[00837] Para la vacuna de srRNA, los ratones recibieron la inyección de 10 ug de ARN en una inyección intramuscular bilateral de 100 uL de volumen (50 uL por pata). Para la vacuna de Ad5, los ratones recibieron la inyección de $5x10^{10}$ partículas virales (VP) en una inyección intramuscular bilateral de 100 uL de volumen (50 uL por pata). Los animales recibieron la inyección con anti-CTLA-4 (clon 9D9, BioXcell), anti-PD-1 (clon RMP1-14, BioXcell) o anti-IgG (clon MPC-11, BioXcell), en dosis de 250 ug, 2 veces por semana, a través de inyección intraperitoneal.

Imagenología luminiscente in vivo

[00838] En cada punto temporal, se inyectaron ratones con un sustrato de luciferina 150 mg/kg vía inyección intraperitoneal y la bioluminiscencia se midió mediante el sistema de imagenología in vivo IVIS (PerkinElmer) 10-15 minutos luego de la inyección.

Disociación de esplenocitos

[00839] Se agruparon los bazos y ganglios linfáticos para cada ratón en 3 mL de RPMI completo (RPMI, FBS al 10 %, penicilina/estreptomicina). Se realizó una disociación mecánica mediante el uso del Disociador gentleMACS (Miltenyi Biotec), siguiendo el protocolo del fabricante. Se filtraron las células disociadas a través de un filtro de 40 micrómetros y se lisaron los glóbulos rojos con solución amortiguadora de lisis de ACK (NH_4Cl 150 mM, $KHCO_3$ 10 mM, Na_2EDTA 0,1 mM). Las células se filtraron nuevamente a través de un filtro de 30 micrómetros y luego se volvieron a suspender en RPMI completo. Se contaron las células en el citómetro de flujo Attune NxT (Thermo Fisher) mediante el uso de tinción con yoduro de propidio para excluir células muertas y apoptóticas. Luego, las células se ajustaron a la concentración adecuada de células hepáticas para un análisis posterior.

Análisis immunospot ligado a enzimas (ELISPOT) ex vivo

[00840] Se realizó un análisis ELISPOT de acuerdo con lineamientos de armonización ELISPOT

{DOI: 10.1038/nprot.2015.068} con el kit ELISpotPLUS de IFNg de ratón (MABTECH). Se incubaron $5x10^4$ esplenocitos con 10 uM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNg de 96 pocillos. Se desarrollaron manchas mediante el uso de fosfatasa alcalina. Se marca una duración de 10 minutos para la reacción, tras cuyo plazo se pone fin a la reacción al colocar la placa bajo agua del grifo. Se contaron las manchas mediante el uso de un AID vSpot Reader Spectrum. Para el análisis ELISPOT, los pocillos con una saturación > 50 % se registraron como «cantidad excesiva para contar". Las muestras con una desviación de los pocillos replicados > 10 % se excluyeron del análisis. Luego, se corrigieron los conteos de manchas con respecto a la confluencia de pocillos mediante el uso de la fórmula: conteo de manchas + 2 x (conteo de manchas x % de confluencia /[100 % - % de confluencia]). Se corrigió el fondo negativo mediante la sustracción de los conteos de manchas en los pocillos de estimulación peptídica de caracter negativo de los pocillos estimulados con antígenos. Finalmente, los pocillos etiquetados como «cantidad excesiva para contar» se ajustaron al valor corregido observado más alto, con redondeo hasta el múltiplo de cien más cercano.

XVI.B. Vector del alfavirus

XVI.B.1. Evaluación *in vitro* del vector del alfavirus

[00841] En una implementación de la presente invención, se generó una estructura principal de ARN de alfavirus para el sistema de expresión de los neoantígenos a partir de una encefalitis equina venezolana (VEE, por sus siglas en inglés) (Kinney, 1986, Virology 152: 400-413) en función del vector de ARN autorreplicante (srRNA). En un ejemplo, las secuencias que codifican las proteínas estructurales de VEE ubicadas 3' del promotor subgenómico 26S se eliminaron (secuencias VEE 7544 a 11.175 eliminadas; numeración basada en Kinney *et ál* 1986; SEQ ID NO:6) y se sustituyeron por secuencias de antígeno (SEQ ID NO:14 y SEQ ID NO:4) o un indicador de luciferasa (por ejemplo, VEE-Luciferasa, SEQ ID NO:15) (Figura 24). Se transcribió el ARN a partir del vector de ADN de srRNA in vitro, transfectado en células HEK293A y se midió la expresión del indicador de luciferasa. Además, se transfectó un ARNm (de no replicación) que codifica luciferasa con fines comparativos. Se observó un aumento de ~30.000 en la señal indicadora de srRNA para srRNA de VEE-luciferasa cuando se comparó la medición de 23 horas con la medición de 2 horas (Tabla 9). En contraste, el indicador de ARNm exhibió un aumento de 10 veces en la señal en el mismo período de tiempo (Tabla 9).

Tabla 9. La expresión de luciferasa del vector de autorreplicación de VEE aumenta con el tiempo. Células HEK293A transfectadas con 10 ng de srRNA de VEE-Luciferasa o 10 ng de ARNm de luciferasa de no replicación (TriLink L-6307) por pocillo en 96 pocillos. Se midió la luminiscencia en diversos momentos luego de la transfección. La expresión de luciferasa se describe como unidades relativas de luminiscencia (RLU, por sus siglas en inglés). Cada punto de datos es el promedio +/- SD de 3 pocillos

transfectados.

Construcción	Momento (hr)	RLU promedio	Desv. estándar (pocillos triplicados)
ARNm	2	878,6666667	120,7904522
ARNm	5	1847,333333	978,515372
ARNm	9	4847	868,3271273
ARNm	23	8639,333333	751,6816702
SRRNA	2	27	15
SRRNA	5	4884,333333	2955,158935
SRRNA	9	182065,5	16030,81784
SRRNA	23	783658,3333	68985,05538

[00842] En otro ejemplo, la replicación del srRNA fue confirmada directamente mediante la medición de los niveles de ARN luego de la transfección del srRNA que codifica luciferasa (VEE-Luciferasa) o un srRNA que codifica un casete de múltiple epítopo (VEE-MAG25mer) mediante el uso de una reacción en cadena de la polimerasa con transcripción inversa cuantitativa (qRT-PCR). Se observó un aumento de ~150 veces en el ARN para el srRNA de VEE-luciferasa (Tabla 10), mientras que se observó un aumento de 30-50 veces en el ARN para el srRNA de VEE-MAG25mer (Tabla 11). Estos datos confirman que los vectores de srRNA de VEE se replican cuando se transfectan a las células.

Tabla 10. Medición directa de la replicación de ARN en células transfectadas de srRNA de VEE-Luciferasa. Células HEK293A transfectadas con srRNA de VEE-Luciferasa (150 ng por pocillo, 24 pocillos) y niveles de ARN cuantificados mediante qRT-PCR en diversos momentos luego de la transfección. Cada medición se normalizó en función del gen de referencia de Actina y se presentó un cambio de múltiplo relativo al momento de 2 horas.

Momento (hr)	Ct de luciferasa	Ct de actina	dCt	dCt de ref	ddCt	Cambio múltiplo relativo
2	20,51	18,14	2,38	2,38	0,00	1,00
4	20,09	18,39	1,70	2,38	-0,67	1,59
6	15,50	18,19	-2,69	2,38	-5,07	33,51
8	13,51	18,36	-4,85	2,38	-7,22	149,43

Tabla 11. Medición directa de la replicación de ARN en células transfectadas de srRNA de VEE-MAG25mer. Células HEK293 transfectadas con srRNA de VEE-MAG25mer (150 ng por pocillo, 24 pocillos) y niveles de ARN cuantificados mediante qRT-PCR en diversos momentos luego de la transfección. Cada medición se normalizó en función del gen de referencia de GusB y se presentó un cambio de múltiplo relativo al momento de 2 horas. Las líneas diferentes líneas en la gráfica representan 2 conjuntos de cebadores/sondas de qPCR, los cuales detectan la región del casete del epítopo del srRNA.

Cebador/ sonda	**Momento (hr)**	**Ct**	**Ct de GusB**	**dCt**	**dCt de ref**	**ddCt**	**Relativo Cambio múltiplo**
Conjunto1	2	18,96	22,41	-3,45	-3,45	0,00	1,00
Conjunto1	4	17,46	22,27	-4,81	-3,45	-1,37	2,58
Conjunto1	6	14,87	22,04	-7,17	-3,45	-3,72	13,21
Conjunto1	8	14,16	22,19	-8,02	-3,45	-4,58	23,86
Conjunto1	24	13,16	22,01	-8,86	-3,45	-5,41	42,52
Conjunto1	36	13,53	22,63	-9,10	-3,45	-5,66	50,45
Conjunto2	2	17,75	22,41	-4,66	-4,66	0,00	1,00
Conjunto2	4	16,66	22,27	-5,61	-4,66	-0,94	1,92
Conjunto2	6	14,22	22,04	-7,82	-4,66	-3,15	8,90
Conjunto2	8	13,18	22,19	-9,01	-4,66	-4,35	20,35
Conjunto2	24	12,22	22,01	-9,80	-4,66	-5,13	35,10
Conjunto2	36	13,08	22,63	-9,55	-4,66	-4,89	29,58

XVI.B.2 Evaluación *in vivo* del vector del alfavirus

[00843] En otro ejemplo, se evaluó la expresión del indicador de VEE-Luciferasa in vivo. Se administraron inyecciones a ratones de 10 ug de srRNA de VEE-Luciferasa encapsulado en una nanopartícula lipídica (MC3) y se obtuvieron imágenes a las 24 y 48 horas, y 7 y 14 días después de la inyección para determinar la presencia de señales bioluminiscentes. Se detectó una señal de luciferasa 24 horas después de la inyección, la cual aumentó con el paso del tiempo y pareció llegar a su punto máximo 7 días después de la inyección de srRNA (Figura 25).

XVI.B.3 Evaluación del modelo tumoral con el vector del alfavirus

[00844] En una implementación, para determinar si el vector de srRNA de VEE dirige respuestas inmunitarias específicas para el antígeno *in vivo*, se generó un vector de srRNA de VEE (VEE-UbAAY,

SEQ ID NO:14) que expresa 2 epítopos tumorales de ratones del MHC clase I, SIINFEKL (SEQ ID NO: 57) y AH1-A5 (Slansky et ál., 2000, Immunity 13:529-538). El epítopo SFL (SIINFEKL (SEQ ID NO: 57)) es expresado por la línea celular de melanoma B16-OVA, y el epítopo AH1-A5 (SPSYAYHQF (SEQ ID NO: 58); Slansky et ál., 2000, Immunity) induce linfocitos T que se dirigen a un epítopo relacionado (AH1/ SPSYVYHQF (SEQ ID NO: 193); Huang et ál., 1996, Proc Natl Acad Sci EUA 93:9730-9735) expresado por la línea celular de carcinoma de colon CT26. En un ejemplo, para los estudios *in vivo*, se generó srRNA de VEE-UbAAY mediante la transcripción *in vitro* con polimerasa T7 (TriLink Biotechnologies), el cual se encapsuló en una nanopartícula lipídica (MC3).

[00845] Se observó una fuerte respuesta de linfocitos T específicos del antígeno dirigida al SFL en comparación con el control dos semanas después de la inmunización de ratones con tumores B16-OVA con srRNA de VEE-UbAAY formulado en MC3. En un ejemplo, se midió una mediana de 3835 células formadoras de manchas (SFC) por cada 10^6 esplenocitos después de la estimulación con el péptido SFL en ensayos ELISpot (Figura 26A, Tabla 12) y un 1,8 % (mediana) de los linfocitos T CD8 eran específicos para el antígeno SFL tal como se midió mediante tinción con pentámeros (Figura 26B, Tabla 12). En otro ejemplo, la administración conjunta de un anticuerpo monoclonal (mAb) anti-CTLA-4 con la vacuna de srRNA de VEE produjo un aumento moderado en las respuestas de linfocitos T a nivel general con una mediana de 4794,5 SFC por cada 10^6 esplenocitos medidas en el ensayo ELISpot (Figura 26A, Tabla 12).

Tabla 12. Resultados de los ensayos ELISPOT y de tinción con pentámeros del MHCI 14 días después de la inmunización con srRNA de VEE en ratones C57BL/6J con tumores B16-OVA.

Grupo	Ratón	SFC/1e6 esplenocitos	Resultado positivo para pentámeros (% de CD8)	Grupo	Ratón	SFC/1e6 esplenocitos	Resultado positivo para pentámeros (% de CD8)
Control	1	47	0,22	Vax	1	6774	4,92
	2	80	0,32		2	2323	1,34
	3	0	0,27		3	2997	1,52
	4	0	0,29		4	4492	1,86
	5	0	0,27		5	4970	3,7
	6	0	0,25		6		4,13
	7	0	0,23		7	3835	1,66

	8	87	0,25		8	3119	1,64
aCTLA4	1	0	0,24	**Vax + aCTLA4**	1	6232	2,16
	2	0	0,26		2	4242	0,82
	3	0	0,39		3	5347	1,57
	4	0	0,28		4	6568	2,33
	5	0	0,28		5	6269	1,55
	6	0	0,28		6	4056	1,74
	7	0	0,31		7	4163	1,14
	8	6	0,26		8	3667	1,01

** Nótese que se excluyeron del análisis los resultados del ratón #6 en el grupo Vax debido a una variabilidad alta entre pocillos triplicados.*

[00846] En otra implementación, para reflejar una estrategia clínica, se realizó una sensibilización/un refuerzo heterólogos en los modelos tumorales de ratones B16-OVA y CT26, donde los ratones con tumores se inmunizaron primero con un vector adenoviral que expresaba el mismo casete antigénico (Ad5-UbAAY) y luego con una inmunización de refuerzo con la vacuna de srRNA de VEE-UbAAY 14 días después de la sensibilización con Ad5-UbAAY. En un ejemplo, se indujo una respuesta inmunitaria específica del antígeno mediante la vacuna de Ad5-UbAAY, lo cual produjo 7330 (mediana) SFC por cada 10^6 esplenocitos, tal como se midió en el ensayo ELISpot (Figura 27A, Tabla 13), y 2,9 % (mediana) de linfocitos T CD8 dirigidos al antígeno SFL, tal como se midió mediante tinción con pentámeros (Figura 27C, Tabla 13). En otro ejemplo, se mantuvo la respuesta de linfocitos T 2 semanas después del refuerzo con srRNA de VEE-UbAAY en el modelo B16-OVA con 3960 (mediana) SFC específicas del SFL por cada 10^6 esplenocitos, tal como se midió en el ensayo ELISpot (Figura 27B, Tabla 13), y 3,1 % (mediana) de linfocitos T CD8 dirigidos al antígeno SFL, tal como se midió mediante tinción con pentámeros (Figura 27D, Tabla 13).

Tabla 13. Supervisión de la inmunidad de los ratones B16-OVA luego de la sensibilización/el refuerzo heterólogos con sensibilización con vacuna Ad5 y refuerzo con srRNA.

Día 14

Grupo	**Ratón**	**SFC/1e6 esplenocitos**	**Resultado positivo para pentámeros**	**Grupo**	**Ratón**	**SFC/1e6 esplenocitos**	**Resultado positivo para pentámeros**

			(% de CD8)				(% de CD8)
Control	1	0	0,10	Vax	1	8514	1,87
	2	0	0,09		2	7779	1,91
	3	0	0,11		3	6177	3,17
	4	46	0,18		4	7945	3,41
	5	0	0,11		5	8821	4,51
	6	16	0,11		6	6881	2,48
	7	0	0,24		7	5365	2,57
	8	37	0,10		8	6705	3,98
aCTLA4	1	0	0,08	Vax + aCTLA4	1	9416	2,35
	2	29	0,10		2	7918	3,33
	3	0	0,09		3	10153	4,50
	4	29	0,09		4	7212	2,98
	5	0	0,10		5	11203	4,38
	6	49	0,10		6	9784	2,27
	7	0	0,10		8	7267	2,87
	8	31	0,14				

Día 28

Grupo	**Ratón**	**SFC/1e6 esplenocitos**	**Resultado positivo para pentámeros (% de CD8)**	**Grupo**	**Ratón**	**SFC/1e6 esplenocitos**	**Resultado positivo para pentámeros (% de CD8)**
Control	2	0	0,17	Vax	1	5033	2,61
	4	0	0,15		2	3958	3,08
	6	20	0,17		4	3960	3,58
aCTLA4	1	7	0,23	Vax + aCTLA4	4	3460	2,44
	2	0	0,18		5	5670	3,46
	3	0	0,14				

[00847] En otra implementación, se observaron resultados similares después de una sensibilización con Ad5-UbAAY y un refuerzo con srRNA de VEE-UbAAY en el modelo de ratones CT26. En un ejemplo,

se observó una respuesta específica del antígeno AH1 después de la sensibilización con Ad5-UbAAY (día 14) con una media de 5187 SFC por cada 10^6 esplenocitos tal como se midió en el ensayo ELISpot (Figura 28A, Tabla 14) y 3799 SFC por cada 10^6 esplenocitos tal como se midió en el ensayo ELISpot después del refuerzo con srRNA de VEE-UbAAY (día 28) (Figura 28B, Tabla 14).

Tabla 14. Supervisión de la inmunidad después de la sensibilización/el refuerzo heterólogos en el modelo tumoral de ratones CT26.

Día 12			**Día 21**		
Grupo	**Ratón**	**SFC/1e6 esplenocitos**	**Grupo**	**Ratón**	**SFC/1e6 esplenocitos**
Control	1	1799	Control	9	167
	2	1442		10	115
	3	1235		11	347
aPD1	1	737	aPD1	8	511
	2	5230		11	758
	3	332	Vax	9	3133
Vax	1	6287		10	2036
	2	4086		11	6227
Vax + aPD1	1	5363	Vax + aPD1	8	3844
	2	6500		9	2071
				11	4888

XVII. Evaluación del modelo tumoral con combinación de ChAdV/srRNA

[00848] Se evaluaron diversos protocolos de dosificación mediante el uso de ChAdV68 y ARN de autorreplicación (srRNA) en modelos tumorales murinos CT26.

XVII.A Métodos y materiales para la evaluación del modelo tumoral con combinación de ChAdV/srRNA

Inyección tumoral

[00849] Se administraron inyecciones de la línea celular tumoral CT26 a ratones Balb/c. 7 días después de la inyección de células tumorales, los ratones se distribuyeron aleatoriamente en los diferentes grupos de estudio (28-40 ratones por grupo) y se inició el tratamiento. Los ratones Balb/c recibieron la inyección en el flanco abdominal izquierdo inferior de 10^6 células CT26/animal. Se dejó que los tumores se desarrollaran durante 7 días antes de la inmunización. Los grupos de estudio se describen en detalle en la Tabla 15.

Tabla 15 - Grupos de estudio para la evaluación del modelo tumoral con combinación de ChAdV/srRNA

Grupo	N	Tratamiento	Dosis	Volumen	Cronograma	Vía
1	40	Control con chAd68	1e11 vp	2x 50 uL	día 0	IM
		Control con srRNA	10 ug	50 ul	día 14, 28, 42	IM
		Anti-PD1	250 ug	100 ul	2x / semana (día de inicio 0)	IP
2	40	Control con chAd68	1e11 vp	2x 50 uL	día 0	IM
		Control con srRNA	10 ug	50 ul	día 14, 28, 42	IM
		Anti-IgG	250 ug	100 ul	2x / semana (día de inicio 0)	IP
3	28	Vacuna contra chAd68	1e11 vp	2x 50 uL	día 0	IM
		Vacuna de srRNA	10 ug	50 ul	día 14, 28, 42	IM
		Anti-PD1	250 ug	100 ul	2x / semana (día de inicio 0)	IP
4	28	Vacuna contra chAd68	1e11 vp	2x 50 uL	día 0	IM
		Vacuna de srRNA	10 ug	50 ul	día 14, 28, 42	IM
		Anti-IgG	250 ug	100 ul	2x / semana (día de inicio 0)	IP
5	28	Vacuna de srRNA	10 ug	50 ul	día 0, 28, 42	IM
		Vacuna contra chAd68	1e11 vp	2x 50 uL	día 14	IM
		Anti-PD1	250 ug	100 ul	2x / semana (día de inicio 0)	IP

6	28	Vacuna de srRNA	10 ug	50 ul	día 0, 28, 42	IM
		Vacuna contra chAd68	1e11 vp	2x 50 uL	día 14	IM
		Anti-IgG	250 ug	100 ul	2x / semana (día de inicio 0)	IP
7	40	Vacuna de srRNA	10 ug	50 ul	día 0, 14, 28, 42	IM
		Anti-PD1	250 ug	100 ul	2x / semana (día de inicio 0)	IP
8	40	Vacuna de srRNA	10 ug	50 ul	día 0, 14, 28, 42	IM
		Anti-IgG	250 ug	100 ul	2x / semana (día de inicio 0)	IP

Inmunizaciones

[00850] Para la vacuna de srRNA, los ratones recibieron la inyección de 10 ug de srRNA de VEE-MAG25mer en una inyección intramuscular bilateral de 100 uL de volumen (50 uL por pata). Para la vacuna de C68, los ratones recibieron la inyección de $1x10^{11}$ partículas virales (VP) de ChAdV68.5WTnt.MAG25mer en una inyección intramuscular bilateral de 100 uL de volumen (50 uL por pata). Los animales recibieron la inyección de anti-PD-1 (clon RMP1-14, BioXcell) o anti-IgG (clon MPC-11, BioXcell), en dosis de 250 ug, 2 veces por semana, a través de inyección intraperitoneal.

Disociación de esplenocitos

[00851] Se agruparon los bazos y ganglios linfáticos para cada ratón en 3 mL de RPMI completo (RPMI, FBS al 10 %, penicilina/estreptomicina). Se realizó una disociación mecánica mediante el uso del Disociador gentleMACS (Miltenyi Biotec), siguiendo el protocolo del fabricante. Se filtraron las células disociadas a través de un filtro de 40 micrómetros y se lisaron los glóbulos rojos con solución amortiguadora de lisis de ACK (NH_4Cl 150 mM, $KHCO_3$ 10 mM, Na_2EDTA 0,1 mM). Las células se filtraron nuevamente a través de un filtro de 30 micrómetros y luego se volvieron a suspender en RPMI completo. Se contaron las células en el citómetro de flujo Attune NxT (Thermo Fisher) mediante el uso de tinción con yoduro de propidio para excluir células muertas y apoptóticas. Luego, las células se ajustaron a la concentración adecuada de células hepáticas para un análisis posterior.

Análisis immunospot ligado a enzimas (ELISPOT) ex vivo

[00852] Se realizó un análisis ELISPOT de acuerdo con lineamientos de armonización ELISPOT

{DOI: 10.1038/nprot.2015.068} con el kit ELISpotPLUS de IFNg de ratón (MABTECH). Se incubaron $5x10^4$ esplenocitos con 10 uM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNg de 96 pocillos. Se desarrollaron manchas mediante el uso de fosfatasa alcalina. Se marca una duración de 10 minutos para la reacción, tras cuyo plazo se pone fin a la reacción al colocar la placa bajo agua del grifo. Se contaron las manchas mediante el uso de un AID vSpot Reader Spectrum. Para el análisis ELISPOT, los pocillos con una saturación > 50 % se registraron como «cantidad excesiva para contar". Las muestras con una desviación de los pocillos replicados > 10 % se excluyeron del análisis. Luego, se corrigieron los conteos de manchas con respecto a la confluencia de pocillos mediante el uso de la fórmula: conteo de manchas + 2 x (conteo de manchas x % de confluencia /[100 % - % de confluencia]). Se corrigió el fondo negativo mediante la sustracción de los conteos de manchas en los pocillos de estimulación peptídica de carácter negativo de los pocillos estimulados con antígenos. Finalmente, los pocillos etiquetados como «cantidad excesiva para contar» se ajustaron al valor corregido observado más alto, con redondeo hasta el múltiplo de cien más cercano.

XVII.B Evaluación de combinación de ChAdV/srRNA en un modelo tumoral CT26

[00853] Se evaluó la inmunogenicidad y eficacia de la sensibilización/el refuerzo heterólogos con ChAdV68.5WTnt.MAG25mer/ srRNa de VEE-MAG25mer o la sensibilización/el refuerzo homólogos con srRNA de VEE-MAG25mer en el modelo tumoral de ratón CT26. Se administraron inyecciones de la línea celular tumoral CT26 a ratones Balb/c. 7 días después de la inyección de células tumorales, los ratones se distribuyeron aleatoriamente en los diferentes grupos de estudio y se inició el tratamiento. Los grupos de estudio se describen en detalle en la Tabla 15 y más generalmente en la Tabla 16.

Tabla 16 – Grupos de estudio con sensibilización/refuerzo

Grupo	Sensibilización	Refuerzo
1	Control	Control
2	Control + anti-PD-1	Control + anti-PD-1
3	ChAdV68.5WTnt.MAG25mer	srRNA de VEE-MAG25mer
4	ChAdV68.5WTnt.MAG25mer + anti-PD-1	srRNA de VEE-MAG25mer + anti-PD-1
5	srRNA de VEE-MAG25mer	ChAdV68.5WTnt.MAG25mer
6	srRNA de VEE-MAG25mer + anti-PD-1	ChAdV68.5WTnt.MAG25mer + anti-PD-1
7	srRNA de VEE-MAG25mer	srRNA de VEE-MAG25mer
8	srRNA de VEE-MAG25mer +	srRNA de VEE-MAG25mer +

	anti-PD-1	anti-PD-1

[00854] Los bazos se recolectaron 14 días después de la vacunación de sensibilización para la supervisión de la inmunidad. Se realizaron mediciones del peso corporal y de los tumores dos veces por semana y se supervisó la supervivencia. Se observaron respuestas inmunitarias fuertes en todos los grupos de vacunación activos en comparación con el control.

[00855] Se observaron respuestas inmunitarias celulares medianas de 10.630, 12.976, 3319 o 3745 células formadoras de manchas (SFC) por cada 10^6 esplenocitos en ensayos ELISpot en ratones inmunizados con ChAdV68.5WTnt.MAG25mer (ChAdV/grupo 3), ChAdV68.5WTnt.MAG25mer + anti-PD-1 (ChAdV + PD-1/grupo 4), srRNA de VEE-MAG25mer (srRNA/mediana para los grupos 5 y 7 combinados) o srRNA de VEE-MAG25mer + anti-PD-1 (srRNA + PD-1/mediana para los grupos 6 y 8 combinados), respectivamente, 14 días después de la primera inmunización (Figura 30 y Tabla 17). En cambio, el control con vacuna (grupo 1) o el control con vacuna con anti-PD-1 (grupo 2) presentaron una mediana de respuestas inmunitarias celulares de 296 o 285 SFC por cada 10^6 esplenocitos, respectivamente.

Tabla 17 – Respuestas inmunitarias celulares en un modelo tumoral CT26

Tratamiento	**Mediana de SFC/10^6 eplenocitos**
Control	296
PD1	285
ChAdV68.5WTnt.MAG25mer (ChAdV)	10630
ChAdV68.5WTnt.MAG25mer + PD1 (ChAdV + PD-1)	12976
srRNA de VEE-MAG25mer (srRNA)	3319
srRNA de VEE-MAG25mer + PD-1 (srRNA + PD1)	3745

[00856] En consonancia con los datos del ELISpot, el 5,6, 7,8, 1,8 o 1,9 % de los linfocitos T CD8 (mediana) presentó respuestas específicas de antígenos en los análisis de tinción de citocinas intracelulares (ICS) para ratones inmunizados con ChAdV68.5WTnt.MAG25mer (ChAdV/grupo 3), ChAdV68.5WTnt.MAG25mer + anti-PD-1 (ChAdV + PD-1/grupo 4), srRNA de VEE-MAG25mer

(srRNA/mediana para los grupos 5 y 7 combinados) o srRNA de VEE-MAG25mer + anti-PD-1 (srRNA + PD-1/mediana para los grupos 6 y 8 combinados), respectivamente, 14 días después de la primera inmunización (Figura 31 y Tabla 18). Los ratones inmunizados con el control con vacuna o control con vacuna combinado con anti-PD-1 mostraron respuestas de CD8 específicas del antígeno de 0,2 y 0,1 %, respectivamente.

Tabla 18 – Respuestas de linfocitos T CDR8 en un modelo tumoral CT26

Tratamiento	**Mediana de % de CD8 positivo para IFN-gamma**
Control	0,21
PD1	0,1
ChAdV68.5WTnt.MAG25mer (ChAdV)	5,6
ChAdV68.5WTnt.MAG25mer + PD1 (ChAdV + PD-1)	7,8
srRNA de VEE-MAG25mer (srRNA)	1,8
srRNA de VEE-MAG25mer + PD-1 (srRNA + PD1)	1,9

[00857] Se midió el desarrollo del tumor en el modelo de tumor de colón CT26 para todos los grupos, y se presenta el desarrollo tumoral hasta 21 días después del inicio del tratamiento (28 días después de la inyección de células tumorales CT-26). Los ratones se sacrificaron 21 días después del inicio del tratamiento en función de los tamaños grandes de tumores (> 2500 mm^3); por lo tanto, se presentan solamente los primeros 21 días para evitar la desviación analítica. Los volúmenes medios de los tumores a los 21 días fueron de 1129, 848, 2142, 1418, 2198 y 1606 mm^3 para la sensibilización con ChAdV68.5WTnt.MAG25mer/refuerzo con srRNA de VEE-MAG25mer (grupo 3), la sensibilización con ChAdV68.5WTnt.MAG25mer/el refuerzo con srRNA de VEE-MAG25mer + anti-PD-1 (grupo 4), la sensibilización con srRNA de VEE-MAG25mer/el refuerzo con ChAdV68.5WTnt.MAG25mer (grupo 5), la sensibilización con srRNA de VEE-MAG25mer/el refuerzo con ChAdV68.5WTnt.MAG25mer + anti-PD-1 (grupo 6), la sensibilización con srRNA de VEE-MAG25mer/el refuerzo con srRNA de VEE-MAG25mer (grupo 7) y la sensibilización con srRNA de VEE-MAG25mer/el refuerzo con srRNA de VEE-MAG25mer + anti-PD-1 (grupo 8), respectivamente (Figura 32 y Tabla 19). Los volúmenes medios de

los tumores en el control con vacuna o control con vacuna combinado con anti-PD-1 fueron de 2361 o 2067 mm^3, respectivamente. En función de estos datos, el tratamiento con vacuna con ChAdV68.5WTnt.MAG25mer / srRNA de VEE-MAG25mer (grupo 3), ChAdV68.5WTnt.MAG25mer / srRNA de VEE-MAG25mer + anti-PD-1 (grupo 4), srRNA de VEE-MAG25mer/ ChAdV68.5WTnt.MAG25mer + anti-PD-1 (grupo 6) y srRNA de VEE-MAG25mer/srRNA de VEE-MAG25mer + anti-PD-1 (grupo 8) produjo una reducción del desarrollo tumoral a los 21 días que fue significativamente diferente con respecto al control (grupo 1).

Tabla 19 - Tamaño de los tumores en el día 21, medido en el modelo CT26

Tratamiento	**Tamaño del tumor (mm^3)**	**SEM**
Control	2361	235
PD1	2067	137
chAdV/srRNA	1129	181
chAdV/srRNA + PD1	848	182
srRNA/chAdV	2142	233
srRNA/chAdV + PD1	1418	220
srRNA	2198	134
srRNA + PD1	1606	210

[00858] Se supervisó la supervivencia durante 35 días después del inicio del tratamiento en el modelo tumoral CT-26 (42 después de la inyección de células tumorales CT-26). Se observó una supervivencia mejorada después de la vacunación de los ratones con 4 de las combinaciones evaluadas. Después de la vacunación, el 64 %, 46 %, 41 % y 36 % de los ratones sobrevivió con sensibilización con ChAdV68.5WTnt.MAG25mer/refuerzo con VEE-MAG25mer en combinación con anti-PD-1 (grupo 4; $P < 0,0001$ en relación con el grupo de control 1), sensibilización con srRNA de VEE-MAG25mer/refuerzo con srRNA de VEE-MAG25mer srRNA en combinación con anti-PD-1 (grupo 8; $P = 0,0006$ en relación con el grupo de control 1), sensibilización con ChAdV68.5WTnt.MAG25mer/refuerzo con srRNA de VEE-MAG25mer (grupo 3; $P = 0,0003$ en relación con el grupo de control 1) y sensibilización con srRNA de VEE-MAG25mer/refuerzo con ChAdV68.5WTnt.MAG25mer en combinación con anti-PD-1 (grupo 6; $P = 0,0016$ en relación con el grupo de control 1), respectivamente (Figura 33 y Tabla 20). La supervivencia no fue significativamente diferente a la del grupo de control 1 (≤ 14 %) para los grupos de tratamiento

restantes [sensibilización con srRNA de VEE-MAG25mer/refuerzo con ChAdV68.5WTnt.MAG25mer (grupo 5), sensibilización con srRNA de VEE-MAG25mer/refuerzo con srRNA de VEE-MAG25mer (grupo 7) y anti-PD-1 solo (grupo 2)].

Tabla 20 - Supervivencia en el modelo CT26

Momento	Control	PD1	chAdV/ srRNA	chAdV/ srRNA + PD1	srRNA/ chAdV	srRNA/ chAdV + PD1	srRNA	srRNA + PD1
0	100	100	100	100,00	100,00	100	100	100
21	96	100	100	100	100	95	100	100
24	54	64	91	100	68	82	68	71
28	21	32	68	86	45	68	21	64
31	7	14	41	64	14	36	11	46
35	7	14	41	64	14	36	11	46

[00859] En conclusión, el ChAdV68.5WTnt.MAG25mer y srRNA de VEE-MAG25mer provocó respuestas de linfocitos T fuertes a los antígenos tumorales de ratones codificados por las vacunas. La administración de una sensibilización con ChAdV68.5WTnt.MAG25mer y un refuerzo con srRNA de VEE-MAG25mer con o sin la administración conjunta de anti-PD-1, la sensibilización con srRNA de VEE-MAG25mer y el refuerzo con ChAdV68.5WTnt.MAG25mer en combinación con anti-PD-1 o la administración de srRNA de VEE-MAG25mer como una inmunización de sensibilización y refuerzo homólogos en combinación con anti-PD-1 a los ratones con tumores produjo una supervivencia mejorada.

XVIII. Estudios en primates no humanos

[00860] Se evaluaron diversos protocolos de dosificación mediante el uso de ChAdV68 y ARN de autorreplicación (srRNA) en primates no humanos: (NHP).

Materiales y métodos

[00861] Se inyectó por vía intramuscular (IM) una vacuna de sensibilización en cada NHP para iniciar el estudio (vacuna de sensibilización). También se inyectaron una o más vacunas de refuerzo (refuerzo de vacuna) por vía intramuscular en cada NHP. Se administraron inyecciones bilaterales por dosis de acuerdo con los grupos detallados en lastables y resumidos a continuación.

Inmunización

[00862] Se inmunizaron bilateralmente macacos rhesus indios Mamu-A*01 con $1x10^{12}$ partículas virales ($5x10^{11}$ partículas virales por inyección) de ChAdV68.5WTnt.MAG25mer, 30 ug de srRNA de VEE-MAG25MER, 100 ug de srRNA de VEE-MAG25mer o 300 ug de srRNA de VEE-MAG25mer formulados en LNP-1 o LNP-2. Se administraron por vía intramuscular 30 ug, 100 ug o 300 ug de refuerzos de vacuna con srRNA de VEE-MAG25mer en el momento indicado después de la vacunación de sensibilización.

Supervisión de la inmunidad

[00863] Se aislaron PBMC en los momentos indicados después de la vacunación de sensibilización mediante el uso de medio de separación de linfocitos (LSM, MP Biomedicals) y tubos de separación LeucoSep (Greiner Bio-One) y se volvieron a suspender en RPMI que contenía FBS al 10 % y penicilina/estreptomicina. Se contaron las células en el citómetro de flujo Attune NxT (Thermo Fisher) mediante el uso de tinción con yoduro de propidio para excluir células muertas y apoptóticas. Luego, las células se ajustaron a la concentración adecuada de células hepáticas para un análisis posterior. Para cada mono en los estudios, se midieron las respuestas de los linfocitos T mediante el uso de métodos de ELISport o citometría de flujo. Se supervisaron las respuestas de linfocitos T de las PBMC a 6 epítopos diferentes de macacos rhesus Mamu-A*01 clase I codificados en las vacunas mediante la medición de la inducción de citocinas, tales como IFN-gamma, mediante el uso del análisis immunospot ligado a enzimas (ELISpot) ex vivo. Se realizó un análisis ELISpot de acuerdo con lineamientos de armonización ELISPOT {DOI: 10.1038/nprot.2015.068} con el kit ELISpotPLUS de IFNg de mono (MABTECH). Se incubaron 200.000 PBMC con 10 uM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNg de 96 pocillos. Se desarrollaron manchas mediante el uso de fosfatasa alcalina. Se marca una duración de 10 minutos para la reacción, tras cuyo plazo se pone fin a la reacción al colocar la placa bajo agua del grifo. Se contaron las manchas mediante el uso de un AID vSpot Reader Spectrum. Para el análisis ELISPOT, los pocillos con una saturación > 50 % se registraron como «cantidad excesiva para contar". Las muestras con una desviación de los pocillos replicados > 10 % se excluyeron del análisis. Luego, se corrigieron los conteos de manchas con respecto a la confluencia de pocillos mediante el uso de la fórmula: conteo de manchas + 2 x (conteo de manchas x % de confluencia /[100 % - % de confluencia]). Se corrigió el fondo negativo mediante la sustracción de los conteos de manchas en los pocillos de estimulación peptídica de carácter negativo de los pocillos estimulados con antígenos. Finalmente, los pocillos etiquetados como «cantidad excesiva para contar» se ajustaron al valor corregido observado más alto, con redondeo hasta el múltiplo de cien más cercano.

[00864] Se supervisaron las respuestas de linfocitos T CD4 y CD8 específicas de las PBMC a 6 epítopos diferentes de macacos rhesus Mamu-A*01 clase I codificados en las vacunas mediante la medición de la inducción de citocinas intracelulares, tales como IFN-gamma, mediante el uso de

citometría de flujo. Los resultados de ambos métodos indican que las citocinas se indujeron con especificidad de antígeno en respuesta a los epítopos.

Inmunogenicidad macacos rhesus

[00865] Este estudio se diseñó para (a) evaluar la inmunogenicidad y seguridad preliminar de una inmunización de dosis de 30 µg y 100 µg de srRNA de VEE-MAG25mer como sensibilización/refuerzo homólogos o sensibilización/refuerzo heterólogos en combinación con ChAdV68.5WTnt.MAG25mer; (b) comparar las respuestas inmunitarias de srRNA de VEE-MAG25mer en nanopartículas lipídicas mediante el uso de LNP1 en comparación con LNP2; (c) evaluar la cinética de las respuestas de linfocitos T a inmunizaciones con ChAdV68.5WTnt.MAG25mer y srRNA de VEE-MAG25mer.

[00866] El grupo de estudio se llevó a cabo en macacos rhesus indios Mamu-A*01 para demostrar la inmunogenicidad. Los antígenos seleccionados usados en este estudio son reconocidos solamente en los macacos Rhesus, específicamente aquellos con un haplotipo de Mamu- A*01 MHC clase I. Los macacos rhesus indios Mamu A*01 se distribuyeron aleatoriamente en los diferentes grupos de estudio (6 macacos por grupo) y recibieron la administración bilateral de una inyección IM del vector de srRNA de VEE-MAG25mer o ChAdV68.5WTnt.MAG25mer que codifica modelos de antígenos que incluye múltiples epítopos con restricción de Mamu- A*01. Los grupos de estudio fueron como se describen a continuación.

[00867] **Tabla 21:** Estudio de inmunogenicidad de GLP macacos rhesus indios

Grupo	Sensibilización	Refuerzo 1	Refuerzo 2
1	srRNA de VEE-MAG25mer -LNP1 (30 µg)	srRNA de VEE-MAG25mer -LNP1 (30 µg)	srRNA de VEE-MAG25mer -LNP1 (30 µg)
2	srRNA de VEE-MAG25mer -LNP1 (100 µg)	srRNA de VEE-MAG25mer -LNP1 (100 µg)	srRNA de VEE-MAG25mer -LNP1 (100 µg)
3	srRNA de VEE-MAG25mer -LNP2 (100 µg)	srRNA de VEE-MAG25mer -LNP2 (100 µg)	srRNA de VEE-MAG25mer -LNP2 (100 µg)
4	ChAdV68.5WTnt.MAG25mer	srRNA de VEE-MAG25mer -LNP1 (100 µg)	srRNA de VEE-MAG25mer -LNP1 (100 µg)

[00868] Se recolectaron PBMC antes de la inmunización y en las semanas 1, 2, 3, 4, 5, 6, 8, 9 y 10

después de la inmunización inicial, para la supervisión de la inmunidad.

Resultados

[00869] Se midieron las respuestas inmunitarias celulares específicas del antígeno en células mononucleares de sangre periférica (PBMC) a seis epítopos con restricción de Mamu-A*01 antes de la inmunización y 1, 2, 3, 4, 5, 6, 8, 9 y 10 semanas después de la inmunización inicial. Los animales recibieron una inmunización de refuerzo con srRNA de VEE-MAG25mer en las semanas 4 y 8 con dosis de ya sea 30 µg o 100 µg y formulada ya sea con LNP1 o LNP2, como se describe en la Tabla 21. Se graficaron las respuestas inmunitarias combinadas a los seis epítopos para cada momento de supervisión de la inmunidad (FIG. 34A-D y Tabla 25).

[00870] Se observaron respuestas inmunitarias específicas del antígeno combinadas en todas las mediciones con 170, 14, 15, 11, 7, 8, 14, 17, 12 SFC por 10^6 PBMC (seis epítopos combinados) 1, 2, 3, 4, 5, 6, 8, 9 o 10 semanas después de una inmunización de sensibilización inicial con srRNA de VEE-MAG25mer-LNP1 (30 µg), respectivamente (FIG. 34A). Se observaron respuestas inmunitarias específicas del antígeno combinadas en todas las mediciones con 108, -3, 14, 1, 37, 4, 105, 17, 25 SFC por 10^6 PBMC (seis epítopos combinados) 1, 2, 3, 4, 5, 6, 8, 9 o 10 semanas después de una inmunización de sensibilización inicial con srRNA de VEE-MAG25mer-LNP1 (100 µg), respectivamente (FIG. 34B). Se observaron respuestas inmunitarias específicas del antígeno combinadas en todas las mediciones con -17, 38, 14, -2, 87, 21, 104, 129, 89 SFC por 10^6 PBMC (seis epítopos combinados) 1, 2, 3, 4, 5, 6, 8, 9 o 10 semanas después de una inmunización de sensibilización inicial con srRNA de VEE-MAG25mer-LNP2 (100 µg), respectivamente (FIG. 34C). Los valores negativos son un resultado de la normalización a valores previos a la extracción de sangre para cada epítopo/animal.

[00871] Se observaron respuestas inmunitarias específicas del antígeno en todas las mediciones con 1218, 1784, 1866, 973, 1813, 747, 797, 1249 y 547 SFC por cada 10^6 PBMC (seis epítopos combinados) 1, 2, 3, 4, 5, 6, 8, 9 o 10 semanas después de una inmunización de sensibilización inicial con ChAdV68.5WTnt.MAG25mer, respectivamente (FIG. 34D). La respuesta inmunitaria mostró el perfil esperado con respuestas inmunitarias en su punto máximo medidas ~2-3 semanas después de la inmunización de sensibilización, seguidas por una contracción en la respuesta inmunitaria después de 4 semanas. Se midieron respuestas inmunitarias celulares específicas del antígeno combinadas de 1813 SFC por cada 10^6 PBMC (seis epítopos combinados) días después de la inmunización inicial con ChAdV68.5WTnt.MAG25mer (es decir, 1 semana después del primer refuerzo con srRNA de VEE-MAG25mer). La respuesta inmunitaria medida 1 semana después del primer refuerzo con srRNA de VEE-MAG25mer (semana 5) fue comparable a la respuesta inmunitaria máxima medida para la inmunización de sensibilización con ChAdV68.5WTnt.MAG25mer (semana 3) (FIG. 34D). También se midieron respuestas inmunitarias celulares específicas del antígeno combinadas de 1249 SFC por 10^6

PBMC (seis epítopos combinados) 9 semanas después de la inmunización inicial con ChAdV68.5WTnt.MAG25mer, respectivamente (es decir, 1 semana después del segundo refuerzo con srRNA VEE-MAG25mer). Las respuestas inmunitarias medidas 1 semana después del segundo refuerzo con srRNA de VEE-MAG25mer (semana 9) fue ~2 veces superiores a las medidas justo antes de la inmunización de refuerzo (FIG. 34D).

[00872] **Tabla 22**: Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para srRNA de VEE-MAG25mer-LNP1 (30 µg) (Grupo 1)

Semana	Antígeno					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0±0	0±0	0±0	0±0	0±0	0±0
2	39,7±22,7	35,4±25,1	3,2±3,6	33±28,1	30,9±20,3	28,3±17,5
3	2±2,4	0,2±1,8	1,8±2,4	3,7±1,9	1,7±2,8	4,9±2,3
4	1±1,8	0,3±1,2	5,5±3,6	2,3±2,2	5,7±2,7	0,8±0,8
5	0,5±0,9	1,4±3,8	3,1±1,6	2,3±2,7	1,9±2	1,4±1,2
6	1,9±1,8	-0,3±3	1,7±1,2	1,4±1,4	0,8±1,1	1,1±1
8	-0,4±0,8	-0,9±2,9	0,5±1,3	3±1,1	2,2±2,1	3,7±2
9	1±1,7	1,2±4,2	7,2±3,9	0,5±0,7	1,6±3	3±1
10	3,8±1,8	11±5	-1,1±1,1	1,9±0,9	1,3±1,6	0,2±0,5

[00873] **Tabla 23**: Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para srRNA de VEE-MAG25mer-LNP1 (100 µg) (Grupo 2)

Semana	Antígeno					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0±0	0±0	0±0	0±0	0±0	0±0
2	7,9±17,2	23,2±17,4	11,4±4,9	41,7±16,5	15±13,5	8,9±6,2
3	-3,1±4,6	-7,2±6,5	2,3±2,3	-0,3±2,7	2,7±5,1	2,2±1,4
4	1,9±3,8	-6,2±7,6	10,5±4,1	1,2±2,9	5,6±4,9	1,1±0,8
5	-2,6±7	-8±5,9	1,5±1,7	6,4±2,3	0,7±4,3	3,3±1,3

6	6,3±6,3	4,4±8,3	6,6±4,4	5,2±5,2	3,9±5	10,8±6,9
8	-3,6±7,2	-6,8±7,3	-0,8±1,2	3,4±4,2	6,4±7,5	5,7±2,7
9	8,1±2,4	20,6±23,4	18,9±5,7	8,1±8,9	9±11,2	40±17,6
10	3,1±8	-3,9±8,5	3,3±1,8	0,6±2,9	7,4±6,4	6,1±2,5

[00874] **Tabla 24**: Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para srRNA de VEE-MAG25mer-LNP2 (100 µg) (Grupo 3)

Semana	Antígeno					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	0±0	0±0	0±0	0±0	0±0	0±0
2	-5,9±3,8	-0,3±0,5	-0,5±1,5	-5,7±6,1	-1±1,3	-3,2±5,5
3	0,7±5,2	3,4±2,4	4,2±4,6	18,3±15,5	11,9±5,1	-0,4±8,2
4	-3,8±5,5	2,3±1,8	11,3±6,1	-3,1±5,6	8,5±4	-1,5±6,1
5	-3,7±5,7	-0,1±0,7	-0,2±1,6	3,4±8,5	3±3,1	-4,6±5
6	12,3±15	7,8±4,9	24,7±19,8	23,2±22,5	18,7±15,8	0,5±6,2
8	5,9±12,3	-0,1±0,7	-0,5±1,3	8,8±14,4	8,7±8	-1,3±4
9	16,1±13,4	16,5±4	22,9±4,2	13±13,2	16,4±7,8	19,6±9,2
10	29,9±21,8	22±19,5	0,5±2,6	22,2±22,6	35,3±15,8	19,4±17,3

[00875] **Tabla 25**: Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para sensibilización con ChAdV68.5WTnt.MAG25mer

Semana	Antígeno					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
1	178±68,7	206,5±94,8	221,2±120	15,4±16,7	33,3±25,9	563,5±174,4
2	311,2±165,5	278,8±100,9	344,6±110,8	46,3±13,5	181,6±76,8	621,4±220,9
3	277,3±101,1	359,6±90,5	468,2±106,6	41,7±11,1	169,8±57,8	549,4±115,7

4	140±46,5	169,6±46,8	239,4±37	26,5±11,4	75±31,6	322,2±50,7
5	155,6±62,1	406,7±96,4	542,7±143,3	35,1±16,6	134,2±53,7	538,5±91,9
6	78,9±42,5	95,5±29,4	220,9±75,3	-1,4±5,3	43,4±19,6	308,1±42,6
8	88,4±30,4	162,1±30,3	253,4±78,6	21,4±11,2	53,7±22,3	217,8±45,2
9	158,5±69	322,3±87,2	338,2±137,1	5,6±12,4	109,2±17,9	314,8±43,4
10	97,3±32,5	133,2±27	154,9±59,2	10±6	26±16,7	125,5±27,7

Estudio de intervalo de odis de ARN no GLP (dosis altas) en macacos rhesus indios.

[00876] Este estudio se diseñó para (a) evaluar la inmunogenicidad y seguridad preliminar del srRNA de VEE-MAG25mer en una dosis de 300 µg como una sensibilización/un refuerzo homólogos o una sensibilización/un refuerzo heterólogos en combinación con ChAdV68.5WTnt.MAG25mer; (b) comparar las respuestas inmunitarias del srRNA de VEE-MAG25mer en nanopartículas lipídicas mediante el uso de LNP1 con respecto a LNP2 en la dosis de 300 µg; y (c) evaluar la cinética de las respuestas de linfocitos T a las inmunizaciones con srRNA de VEE-MAG25mer y ChAdV68.5WTnt.MAG25mer.

[00877] El grupo de estudio se llevó a cabo en macacos rhesus indios Mamu-A*01 para demostrar la inmunogenicidad. La inmunogenicidad de una vacuna en especies de primates no humanos, tales como Rhesus, es el mejor indicador para predecir la potencia de la vacuna en los seres humanos. Además, los antígenos seleccionados usados en este estudio son reconocidos solamente en los macacos Rhesus, específicamente aquellos con un haplotipo de Mamu- A*01 MHC clase I. Los macacos rhesus indios Mamu-A*01 se distribuyeron aleatoriamente en los diferentes grupos de estudio (6 macacos por grupo) y recibieron la administración de una inyección IM bilateral de ya sea ChAdV68.5WTnt.MAG25mer o srRNA de VEE-MAG25mer que codifica modelos de antígenos que incluye múltiples antígenos con restricción de Mamu- A*01. Los grupos de estudio fueron como se describe a continuación.

[00878] Se recogieron las PBMC antes de la inmunización y 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 o 24 semanas después de la inmunización inicial para monitoreo inmunitario para el grupo 1 (sensibilización/refuerzo heterólogos). Se recogieron las PBMC antes de la inmunización y 4, 5, 7, 8, 10, 11, 12, 13, 14 o 15 semanas después de la inmunización inicial para monitoreo inmunitario para los grupos 2 y 3 (sensibilización/refuerzo homólogos).

[00879] **Tabla 26:** Estudio de inmunogenicidad no GLP en macacos rhesus indios

Grupo	Sensibilización	Refuerzo 1	Refuerzo 2	Refuerzo 3

1	ChAdV68.5WTnt.MAG25mer	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)
2	VEE-MAG25mer srRNA-LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)	
3	VEE-MAG25mer srRNA-LNP1 (300 µg)	VEE-MAG25mer srRNA -LNP1 (300 µg)	VEE-MAG25mer srRNA -LNP1 (300 µg)	

Resultados

[00880] Se inmunizaron macacos rhesus indios Mamu-A*01 con ChAdV68.5-WTnt.MAG25mer. Se midieron las respuestas inmunitarias celulares específicas del antígeno en células mononucleares de sangre periférica (PBMC) a seis epítopos con restricción de Mamu-A*01 antes de la inmunización y 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 o 24 semanas después de la inmunización inicial (FIG. 35 y Tabla 27). Los animales recibieron inmunización de refuerzo con srRNA de VEE-MAG25mer mediante el uso de la formulación de LNP2 en las semanas 4, 12 y 20. Las respuestas inmunitarias específicas del antígeno combinadas de 1750, 4225, 1100, 2529, 3218, 1915, 1708, 1561, 5077, 4543, 4920, 5820, 3395, 2728, 1996, 1465, 4730, 2984, 2828 o 3043 SFC por 10^6 PBMC (seis epítopos combinados) se midieron 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 o 24 semanas después dela inmunización inicial con ChAdV68.5WTnt.MAG25mer (FIG. 35). Las respuestas inmunitarias medidas 1 semana después de la segunda inmunización de refuerzo (semana 13) con srRNA de VEE-MAG25mer fueron ~3 veces superiores a las medidas justo antes de la inmunización de refuerzo (semana 12). Las respuestas inmunitarias medidas 1 semana después de la tercera inmunización de refuerzo (semana 21) con srRNA de VEE-MAG25mer fueron ~3 veces superiores a las medidas justo antes de la inmunización de refuerzo (semana 20), similar a la respuesta observada para el segundo refuerzo.

[00881] Los macacos rhesus indios Mamu-A*01 también se inmunizaron con srRNA de VEE-MAG25mer mediante el uso de dos formulaciones de LNP diferentes (LNP1 y LNP2). Se midieron las respuestas inmunitarias celulares específicas del antígeno en células mononucleares de sangre periférica (PBMC) a seis epítopos con restricción de Mamu-A*01 antes de la inmunización y 4, 5, 6, 7, 8, 10, 11, 12, 13, 14 o 15 semanas después de la inmunización inicial (FIG. 36 y 37, Tablas 28 y 29).

Los animales recibieron inmunización de refuerzo con srRNA de VEE-MAG25mer mediante el uso de la formulación de LNP1 o LNP2 respectiva en las semanas 4 y 12. Se midieron las respuestas inmunitarias específicas del antígeno combinadas de 168, 204, 103, 126, 140, 145, 330, 203, 162 SFC por 106 PBMC (seis epítopos combinados) 4, 5, 7, 8, 10, 11, 13, 14, 15 semanas después de la inmunización con srRNA de VEE-MAG25mer-LNP2 (FIG. 36). Se midieron las respuestas inmunitarias específicas del antígeno combinadas de 189, 185, 349, 437, 492, 570, 233, 886, 369, 381 SFC por 10^6 PBMC (seis epítopos combinados) 4, 5, 7, 8, 10, 11, 12, 13, 14, 15 semanas después de la inmunización con srRNA de VEE-MAG25mer-LNP1 (FIG. 37).

[00882] **Tabla 27:** Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para vacunación de sensibilización con ChAdV68.5WTnt.MAG25mer (Grupo 1)

Semana	**Antígeno**					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	173±41,6	373,5±87,3	461,4±74,2	38,4±26,1	94,5±26	609,2±121,9
5	412,7±138,4	987,8±283,3	1064,4±266,9	85,6±31,2	367,2±135,2	1306,8±332,8
6	116,2±41,2	231,1±46,3	268,3±90,7	86,1±42	174,3±61	223,9±38,1
7	287,4±148,7	588,9±173,9	693,2±224,8	92,1±33,5	172,9±55,6	694,6±194,8
8	325,4±126,6	735,8±212	948,9±274,5	211,3±62,7	179,1±50	817,3±185,2
10	312±129,7	543,2±188,4	618,6±221,7	-5,7±4,1	136,5±51,3	309,9±85,6
11	248,5±81,1	348,7±129,8	581,1±205,5	-3,1±4,4	119±51,2	413,7±144,8
12	261,9±68,2	329,9±83	486,5±118,6	-1,2±5,1	132,8±31,8	350,9±69,3

13	389,3 ±167, 7	1615,8±418, 3	1244,3±403, 6	1,3±8,1	522,5±155	1303,3±385, 6
14	406,3 ±121, 6	1616±491,7	1142,3±247, 2	6,6±11,1	322,7±94,1	1048,6±215, 6
15	446,8 ±138, 7	1700,8±469, 1	1306,3±294, 4	43±24,5	421,2±87,9	1001,5±236, 4
16	686,8 ±268, 8	1979,5±541, 7	1616,8±411, 8	2,4±7,8	381,9±116, 4	1152,8±352, 7
17	375,8 ±109, 3	1378,6±561, 2	773,1±210,3	-1,4±4,3	177,6±93,7	691,7±245
18	255,9 ±99,7	1538,4±498, 1	498,7±152,3	-5,3±3,3	26,2±13,4	413,9±164,8
19	133±6 2,6	955,9±456,8	491,1±121,8	-5,7±4,1	50,3±25,4	371,2±123,7
20	163,7 ±55,8	641,7±313,5	357,9±91,1	2,6±7,5	41,4±24,2	257,8±68,9
21	319,9 ±160, 5	2017,1±419, 9	1204,8±335, 2	-3,7±5,1	268,1±109, 6	924,1±301
22	244,7 ±105, 6	1370,9±563, 5	780,3±390	-3,6±5,1	118,2±68,1	473,3±249,3
23	176,7 ±81,8	1263,7±527, 3	838,6±367,9	-5,7±4,1	73,6±49	480,9±163,9
24	236,5 ±92	1324,7±589, 3	879,7±321	-0,4±5,7	104±53,1	498±135,8

[00883] **Tabla 28:** Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para vacunación de sensibilización con srRNA de VEE-MAG25mer-LNP2 (300 µg)

(Grupo 2)

Semana	**Antígeno**					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	46±27,1	18,4±6,8	58,3±45,8	29,9±20,8	4,9±2,3	10,7±4
5	85,4±54	5,2±5,8	52,4±51,2	34,5±35	11,8±12,2	14,4±7,9
7	18,6±32,5	1,9±1,7	59,4±55,7	9,3±10,7	3,3±3	10,7±6,1
8	36,6±39,4	6,3±3,9	48,7±39,9	13,5±8,8	3,8±3,6	17,2±9,7
10	69,1±59,1	4,4±1,9	39,3±38	14,7±10,8	4,4±5,3	8,5±5,3
11	43±38,8	22,6±21,1	30,2±26,2	3,3±2,2	5,8±3,5	40,3±25,5
13	120,4±78,3	68,2±43,9	54,2±36,8	21,8±7,4	17,7±6,1	47,4±27,3
14	76±44,8	28±19,5	65,9±64,3	-0,3±1,3	2,5±2	31,1±26,5
15	58,9±41,4	19,5±15,1	55,4±51	2,5±2	5,5±3,6	20,1±15,7

[00884] **Tabla 29:** Células formadoras de manchas (SFC) promedio por 10^6 PBMC para cada epítopo ± SEM para vacunación de sensibilización con srRNA de VEE-MAG25mer-LNP1 (300 µg) (Grupo 3)

Semana	**Antígeno**					
	Env CL9	Env TL9	Gag CM9	Gag LW9	Pol SV9	Tat TL8
4	19,5±8,7	13,3±13,1	16,5±15,3	10,5±7,3	35,9±24,8	92,9±91,6
5	87,9±43,9	12,7±11,7	37,2±31,9	21,1±23,8	13,2±13,7	12,6±13,7
7	21,1±13,3	48,8±48,4	51,7±39,5	9,1±10,5	58,6±55,8	159,4±159
8	47,7±21,7	66,4±52,2	59,8±57,4	49,4±28	79,4±63	133,8±132,1

10	49±30,2	42,2±41,1	139,3±139,3	51,6±51,2	78,2±75,8	131,7±131,6
11	42±26,8	20,9±21,4	177,1±162	-6,3±4,3	104,3±104,1	231,5±230,1
12	40,2±19	20,3±11,9	42,2±46,7	3,7±6,7	57±44,7	70±69,2
13	81,2±48,9	38,2±37,6	259,4±222,2	-4±4,1	164,1±159,3	347,3±343,5
14	34,5±31,8	5,3±11,6	138,6±137,3	-4,7±5,2	52,3±52,9	142,6±142,6
15	49±24	6,7±9,8	167,1±163,8	-6,4±4,2	47,8±42,3	116,6±114,5

Estudio de intervalo de dosis de srRNA

[00885] En una implementación de la presente invención, se puede llevar a cabo un estudio del intervalo de dosis de srRNA en macacos rhesus indios A01 para identificar qué dosis de srRNA hacer progresar a los estudios de inmunogenicidad de NHP. En un ejemplo, a macacos rhesus indios Mamu A01 se les puede administrar un vector de srRNA que codifica antígenos del modelo que incluyen múltiples epítopos restringidos por mamu A01 mediante inyección IM. En otro ejemplo, un anticuerpo monoclonal anti-CTLA-4 puede administrarse por vía SC próximo al sitio de la inyección de vacuna IM para dirigirse al ganglio linfático que drena la vacuna en un grupo de animales. Se pueden recoger las PBMC cada 2 semanas después de la vacunación inicial, para monitorear la inmunidad. Los grupos de estudio se describen a continuación (Tabla 30).

Tabla 30: Estudio de intervalo de dosis de ARN no GLP en macacos rhesus indios

Grupo	**Sensibilización**	**Refuerzo 1**	**Refuerzo 2**
1	srRNA-LNP (dosis baja)	srRNA-LNP (dosis baja)	srRNA-LNP (dosis baja)
2	srRNA-LNP (dosis media)	srRNA-LNP (dosis media)	srRNA-LNP (dosis media)
3	srRNA-LNP (dosis alta)	srRNA-LNP (dosis alta)	srRNA-LNP (dosis alta)
4	srRNA-LNP (dosis alta) + anti-CTLA-4	srRNA-LNP (dosis alta) + anti-CTLA-4	srRNA-LNP (dosis alta) + anti-CTLA-4

* Se debe determinar el intervalo de dosis de srRNA con la dosis alta ≤300 µg.

Estudio de inmunogenicidad en macacos rhesus indios

[00886] En una implementación de la presente invención, se pueden llevar a cabo estudios de vacuna en macacos rhesus indios mamu A01 para demostrar la inmunogenicidad. En un ejemplo, a macacos rhesus indios Mamu A01 se les puede administrar una inyección IM de un vector de ChAdV y/o srRNA que codifica antígenos del modelo que incluyen múltiples antígenos restringidos por mamu A01. En otro ejemplo, un anticuerpo monoclonal anti-CTLA-4 se administrará por vía SC próximo al sitio de la inyección de vacuna IM para algunos de los grupos. Se pueden recoger las PBMC cada 2 semanas después de la vacunación inicial, para monitorear la inmunidad. Los grupos de estudio se describen a continuación (Tabla 31).

Tabla 31: Estudio de inmunogenicidad no GLP en macacos rhesus indios

Grupo	**Sensibilización**	**Refuerzo 1**	**Refuerzo 2**
1	ChAdV	srRNA-LNP*	srRNA-LNP
2	srRNA-LNP	ChAdV	srRNA-LNP
3	srRNA-LNP	srRNA-LNP	ChAdV
4	srRNA-LNP + anti-CTLA-4	srRNA-LNP + anti-CTLA-4	srRNA-LNP + anti-CTLA-4
5	ChAdV + anti-CTLA-4	srRNA-LNP + anti-CTLA-4	srRNA-LNP + anti-CTLA-4
6	srRNA-LNP + anti-CTLA-4	ChAdV + anti-CTLA-4	srRNA-LNP + anti-CTLA-4

*La dosis de srRNA se determinará basándose en el estudio de intervalo de dosis de srRNA.

XIX. Identificación de TCR y linfocitos T reactivos para el objetivo peptídico/MHC

[00887] Los linfocitos T pueden aislarse de sangre, nódulos linfáticos o tumores de los pacientes. Los linfocitos T pueden enriquecerse para linfocitos T específicos para antígenos, por ejemplo, al clasificar células que se unen al tetrámero de antígeno-MHC o al clasificar células activadas estimuladas en un cocultivo *in vitro* de linfocitos T y células presentadoras de antígenos pulsadas por antígenos. Se conocen en la técnica diversos reactivos para la identificación de linfocitos T específicos para antígenos que incluyen tetrámeros cargados con antígenos y otros reactivos a base de MHC.

[00888] Los dímeros de TCR alfa-beta (o gamma-delta) pertinentes para los antígenos pueden identificarse mediante la secuenciación de células individuales de TCR de linfocitos T específicos para antígenos. De manera alternativa, puede llevarse a cabo una secuenciación de TCR a granel de linfocitos T específicos para antígenos y los pares alfa-beta con una probabilidad elevada de coincidencia pueden determinarse mediante el uso de un método de emparejamiento de TCR conocido en la técnica.

[00889] De manera alternativa o adicional, los linfocitos T específicos para antígenos pueden obtenerse mediante cebado *in vitro* de linfocitos T sin modificación de donantes saludables. Los linfocitos T obtenidos de PBMC, nódulos linfáticos o sangre del cordón umbilical pueden estimularse de forma repetitiva mediante células presentadoras de antígenos pulsadas por antígenos para cebar la diferenciación de linfocitos T con experimentación de antígenos. Los TCR pueden identificarse luego de forma similar a la descrita anteriormente para los linfocitos T específicos para antígenos de pacientes.

Determinadas secuencias

[00890] A continuación, se muestran secuencias para vectores, casetes y anticuerpos.

Tremelimumab VL (SEQ ID NO:16)
PSSLSASVGDRVTITCRASQSINSYLDWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDFT LTISSLQPEDFATYYCQQYYSTPFTFGPGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNF YPREAKV
Tremelimumab VH (SEQ ID NO:17)
GVVQPGRSLRLSCAASGFTFSSYGMHWVRQAPGKGLEWVAVIWYDGSNKYYADSVKGRFTISR DNSKNTLYLQMNSLRAEDTAVYYCARDPRGATLYYYYYGMDVWGQGTTVTVSSASTKGPSVFP LAPCSRSTSESTAALGCLVKDYFPEPVTVSWNSGALTSGVH
Tremelimumab VH CDR1 (SEQ ID NO:18)
GFTFSSYGMH
Tremelimumab VH CDR2 (SEQ ID NO:19)
VIWYDGSNKYYADSV
Tremelimumab VH CDR3 (SEQ ID NO:20)
DPRGATLYYYYYGMDV
Tremelimumab VL CDR1 (SEQ ID NO:21)
RASQSINSYLD
Tremelimumab VL CDR2 (SEQ ID NO:22)
AASSLQS
Tremelimumab VL CDR3 (SEQ ID NO:23)
QQYYSTPFT

Durvalumab (MEDI4736) VL (SEQ ID NO:24)
EIVLTQSPGTLSLSPGERATLSCRASQRVSSSYLAWYQQKPGQAPRLLIYDASSRATGIPDRFSG SGSGTDFTLTISRLEPEDFAVYYCQQYGSLPWTFGQGTKVEIK
MEDI4736 VH (SEQ ID NO:25)
EVQLVESGGGLVQPGGSLRLSCAASGFTFSRYWMSWVRQAPGKGLEWVANIKQDGSEKYYVD SVKGRFTISRDNAKNSLYLQMNSLRAEDTAVYYCAREGGWFGELAFDYWGQGTLVTVSS
MEDI4736 VH CDR1 (SEQ ID NO:26)
RYWMS
MEDI4736 VH CDR2 (SEQ ID NO:27)
NIKQDGSEKYYVDSVKG
MEDI4736 VH CDR3 (SEQ ID NO:28)
EGGWFGELAFDY
MEDI4736 VL CDR1 (SEQ ID NO:29)
RASQRVSSSYLA
MEDI4736 VL CDR2 (SEQ ID NO:30)
DASSRAT
MEDI4736 VL CDR3 (SEQ ID NO:31)
QQYGSLPWT
Nucleótido de UbA76-25merPDTT (SEQ ID NO:32)
GCCCGGGCATTTAAATGCGATCGCATCGATtacgactctagaatagtctagtccgcaggccaccatgCAGATCT TCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATCGA GAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATC TTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTC AACCCTGCACCTGGTCCTTCGCCTGAGAGGTGcCatgtttcaggcgctgagcgaaggctgcaccccgtatgatatt aaccagatgctgaacgtgctgggcgatcatcaggtctcaggccttgagcagcttgagagtataatcaactttgaaaaactgactgaatgg accagttctaatgttatgCCTATCCTGTCTCCTCTGACAAAGGGCATCCTGGGCTTCGTGTTTACCCTG ACCGTGCCTTCTGAGAGAGGACTTagctgcattagcgaagcggatgcgaccaccccggaaagcgcgaacctgggcg aagaaattctgagccagctgtatctttggccaagggtgacctaccattcccctagttatgcttaccaccaatttgaaagacgagccaaatat aaaagaCACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCTACCCTGTGTACGTGTTCGGC GATTGCGTGCAGGGCGATtgggatgcgattcgctttcgctattgcgcgccgccgggctatgcgctgctgcgctgcaacgatac caactatagcgctctgctggctgtgggggccctagaaggacccaggaatcaggactggcttggtgtcccaagacaacttgtaactCGG

ATGCAGGCTATTCAGAATGCCGGCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTT CTGGCTGCAAgcgtttctgatggcgctgaccgatagcggcccgaaaaccaacattattgtggatagccagtatgtgatgggcatta gcaaaccgagctttcaggaatttgtggattgggaaaacgtgagcccggaactgaacagcaccgatcagccgtttTGGCAAGCCG GAATCCTGGCCAGAAATCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTA CCAGggtcagtcactagtcatctctgcttctatcattgtcttcaacctgCtggaactggaaggtgattatcgagatgatggcaacgtgtggg tgcataccccgctgagcccgcgcaccctgaacgcgtgggtgaaagcggtggaagaaaaaaaaggtattccagttcacctagagctgg ccagtatgaccaacaTggagctcatgagcagtattgtgcatcagcaggtcAGAACATACGGCCCCGTGTTCATGTGT CTCGGCGGACTGCTTACAATGGTGGCTGGTGCTGTGTGGCTGACAGTGcgagtgctcgagctgttccg ggccgcgcagctggccaacgacgtggtcctccagatcatggagctttgtggtgcagcgtttcgccaggtgtgccataccaccgtgccgtg gccgaacgcgagcctgaccccgaaatggaacaacgaaaccacccagccccagatcgccaactgcagcgtgtatgacttttttgtgtgg ctccattattattctgttcgagacacactttggccaagggtgacctaccatatgaacaaatatgcgtatcatatgctggaaagacgagccaa atataaaagaGGACCAGGACCTGGCGCTAAATTTGTGGCCGCCTGGACACTGAAAGCCGCTGCT GGTCCTGGACCTGGCCAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTCG GACCCGGACCAGGCTGATGATTTCGAAATTTAAATAAGCTTGCGGCCGCTAGGGATAACAGG GTAATtatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtgg
Polipéptido de UbA76-25merPDTT (SEQ ID NO:33)
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLH LVLRLRGAMFQALSEGCTPYDINQMLNVLGDHQVSGLEQLESIINFEKLTEWTSSNVMPILSPLTK GILGFVFTLTVPSERGLSCISEADATTPESANLGEEILSQLYLWPRVTYHSPSYAYHQFERRAKYK RHFPGFGQSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYSALLAVGALEGPRN QDWLGVPRQLVTRMQAIQNAGLCTLVAMLEETIFWLQAFLMALTDSGPKTNIIVDSQYVMGISKP SFQEFVDWENVSPELNSTDQPFWQAGILARNLVPMVATVQGQNLKYQGQSLVISASIIVFNLLELE GDYRDDGNVWVHTPLSPRTLNAWVKAVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVF MCLGGLLTMVAGAVWLTVRVLELFRAAQLANDVVLQIMELCGAAFRQVCHTTVPWPNASLTPKW NNETTQPQIANCSVYDFFVWLHYYSVRDTLWPRVTYHMNKYAYHMLERRAKYKRGPGPGAKFV AAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG
Nucleótido de MAG-25merPDTT (SEQ ID NO:34)
ATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTGCACACCCTATGATATTAACCAGATGCT GAATGTCCTGGGAGACCACCAGGTCTCTGGCCTGGAGCAGCTGGAGAGCATCATCAACTTC GAGAAGCTGACCGAGTGGACAAGCTCCAATGTGATGCCTATCCTGTCCCCACTGACCAAGG GCATCCTGGGCTTCGTGTTTACCCTGACAGTGCCTTCTGAGCGGGGCCTGTCTTGCATCAGC GAGGCAGACGCAACCACACCAGAGTCCGCCAATCTGGGCGAGGAGATCCTGTCTCAGCTGT

ACCTGTGGCCCCGGGTGACATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGGAGA GCCAAGTACAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGCTGTTCGGCTACCCCGTGTA CGTGTTCGGCGATTGCGTGCAGGGCGACTGGGATGCCATCCGGTTTAGATACTGCGCACCA CCTGGATATGCACTGCTGAGGTGTAACGACACCAATTATTCCGCCCTGCTGGCAGTGGGCG CCCTGGAGGGCCCTCGCAATCAGGATTGGCTGGGCGTGCCAAGGCAGCTGGTGACACGCA TGCAGGCCATCCAGAACGCAGGCCTGTGCACCCTGGTGGCAATGCTGGAGGAGACAATCTT CTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCGGCCCCAAGACAAACATCATCGTG GATTCCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGTTTGTGGACTGGGAGAA CGTGAGCCCAGAGCTGAATTCCACCGATCAGCCATTCTGGCAGGCAGGAATCCTGGCAAGG AACCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCCAGAGCC TGGTCATCAGCGCCTCCATCATCGTGTTTAACCTGCTGGAGCTGGAGGGCGACTATCGGGA CGATGGCAACGTGTGGGTGCACACCCCACTGAGCCCCAGAACACTGAACGCCTGGGTGAAG GCCGTGGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGGCCTCCATGACCAATATGG AGCTGATGTCTAGCATCGTGCACCAGCAGGTGAGGACATACGGACCCGTGTTCATGTGCCT GGGAGGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGCT GTTCAGAGCCGCCCAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCGGAGCA GCCTTTCGCCAGGTGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGT GGAACAATGAGACAACACAGCCTCAGATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTGG CTGCACTACTATAGCGTGAGGGATACCCTGTGGCCCCGCGTGACATACCACATGAATAAGTA CGCCTATCACATGCTGGAGAGGCGCGCCAAGTATAAGAGAGGCCCTGGCCCAGGCGCAAAG TTTGTGGCAGCATGGACCCTGAAGGCCGCCGCCGGCCCCGGCCCCGGCCAGTATATCAAG GCTAACAGTAAGTTCATTGGAATCACAGAGCTGGGACCCGGACCTGGA
Polipéptido de MAG-25merPDTT (SEQ ID NO:35)
MAGMFQALSEGCTPYDINQMLNVLGDHQVSGLEQLESIINFEKLTEWTSSNVMPILSPLTKGILGF VFTLTVPSERGLSCISEADATTPESANLGEEILSQLYLWPRVTYHSPSYAYHQFERRAKYKRHFP GFGQSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYSALLAVGALEGPRNQDW LGVPRQLVTRMQAIQNAGLCTLVAMLEETIFWLQAFLMALTDSGPKTNIIVDSQYVMGISKPSFQE FVDWENVSPELNSTDQPFWQAGILARNLVPMVATVQGQNLKYQGQSLVISASIIVFNLLELEGDY RDDGNVWVHTPLSPRTLNAWVKAVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLG GLLTMVAGAVWLTVRVLELFRAAQLANDVVLQIMELCGAAFRQVCHTTVPWPNASLTPKWNNET TQPQIANCSVYDFFVWLHYYSVRDTLWPRVTYHMNKYAYHMLERRAKYKRGPGPGAKFVAAWT LKAAAGPGPGQYIKANSKFIGITELGPGPG
Nucleótido de Ub7625merPDTT_NoSFL (SEQ ID NO:36)

GCCCGGGCATTTAAATGCGATCGCATCGATtacgactctagaatagtctagtccgcaggccaccatgCAGATCT
TCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATCGA
GAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATC
TTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTC
AACCCTGCACCTGGTCCTTCGCCTGAGAGGTGcCatgtttcaggcgctgagcgaaggctgcaccccgtatgatatt
aaccagatgctgaacgtgctgggcgatcatcagtttaagcacatcaaagcctttgaccggacatttgctaacaacccaggtcccatggttg
tgtttgccacacctgggCCTATCCTGTCTCCTCTGACAAAGGGCATCCTGGGCTTCGTGTTTACCCTG
ACCGTGCCTTCTGAGAGAGGACTTagctgcattagcgaagcggatgcgaccaccccggaaagcgcgaacctgggcg
aagaaattctgagccagctgtatctttggccaagggtgacctaccattcccctagttatgcttaccaccaatttgaaagacgagccaaatat
aaaagaCACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCTACCCTGTGTACGTGTTCGGC
GATTGCGTGCAGGGCGATtgggatgcgattcgctttcgctattgcgcgccgccgggctatgcgctgctgcgctgcaacgatac
caactatagcgctctgctggctgtgggggccctagaaggacccaggaatcaggactggcttggtgtcccaagacaacttgtaactCGG
ATGCAGGCTATTCAGAATGCCGGCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTT
CTGGCTGCAAgcgtttctgatggcgctgaccgatagcggcccgaaaaccaacattattgtggatagccagtatgtgatgggcatta
gcaaaccgagctttcaggaatttgtggattgggaaaacgtgagcccggaactgaacagcaccgatcagccgtttTGGCAAGCCG
GAATCCTGGCCAGAAATCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTA
CCAGggtcagtcactagtcatctctgcttctatcattgtcttcaacctgCtggaactggaaggtgattatcgagatgatggcaacgtgtggg
tgcataccccgctgagcccgcgcaccctgaacgcgtgggtgaaagcggtggaagaaaaaaaaaggtattccagttcacctagagctgg
ccagtatgaccaacaTggagctcatgagcagtattgtgcatcagcaggtcAGAACATACGGCCCCGTGTTCATGTGT
CTCGGCGGACTGCTTACAATGGTGGCTGGTGCTGTGTGGCTGACAGTGcgagtgctcgagctgttccg
ggccgcgcagctggccaacgacgtggtcctccagatcatggagctttgtggtgcagcgtttcgccaggtgtgccataccaccgtgccgtg
gccgaacgcgagcctgaccccgaaatggaacaacgaaaccacccagccccagatcgccaactgcagcgtgtatgacttttttgtgtgg
ctccattattattctgttcgagacacactttggccaagggtgacctaccatatgaacaaatatgcgtatcatatgctggaaagacgagccaa
atataaaagaGGACCAGGACCTGGCGCTAAATTTGTGGCCGCCTGGACACTGAAAGCCGCTGCT
GGTCCTGGACCTGGCCAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTCG
GACCCGGACCAGGCTGATGATTTCGAAATTTAAATAAGCTTGCGGCCGCTAGGGATAACAGG
GTAATtatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtgg

Polipéptido de Ub7625merPDTT_NoSFL (SEQ ID NO:37)

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLH
LVLRLRGAMFQALSEGCTPYDINQMLNVLGDHQFKHIKAFDRTFANNPGPMVVFATPGPILSPLT
KGILGFVFTLTVPSERGLSCISEADATTPESANLGEEILSQLYLWPRVTYHSPSYAYHQFERRAKY
KRHFPGFGQSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYSALLAVGALEGPR
NQDWLGVPRQLVTRMQAIQNAGLCTLVAMLEETIFWLQAFLMALTDSGPKTNIIVDSQYVMGISK
PSFQEFVDWENVSPELNSTDQPFWQAGILARNLVPMVATVQGQNLKYQGQSLVISASIIVFNLLEL

EGDYRDDGNVWVHTPLSPRTLNAWVKAVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVF
MCLGGLLTMVAGAVWLTVRVLELFRAAQLANDVVLQIMELCGAAFRQVCHTTVPWPNASLTPKW
NNETTQPQIANCSVYDFFVWLHYYSVRDTLWPRVTYHMNKYAYHMLERRAKYKRGPGPGAKFV
AAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG

ChAdV68.5WTnt.MAG25mer (SEQ ID NO:2); AC_000011.1 con las secuencias E1 (de nt 577 a 3403) y E3 (nt 27,125- 31,825) eliminadas; nucleótidos de ATCC VR-594 correspondientes sustituidos en cinco posiciones; modelo de casete de neoantígeno bajo el control del promotor/potenciador del CMV insertado en lugar de la E1 eliminada; SV40 poliA 3' del casete

CCATCTTCAATAATATACCTCAAACTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATT
TGGGGAGGAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTG
ACGTTTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACG
TCAAACGAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCCGCGCTCTCTGACAGGAAAT
GAGGTGTTTCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACTGAATGAGG
AAGTGAAAATCTGAGTAATTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTA
GACTTTGACCGATTACGTGGGGGTTTCGATTACCGTGTTTTTCACCTAAATTTCCGCGTACGG
TGTCAAAGTCCGGTGTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTGCGCT
CTCCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAG
TCAGATCTACACTTTGAAAGTAGGGATAACAGGGTAATgacattgattattgactagttGttaaTAGTAATC
AATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTT
CCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACT
GCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATTGACGTCAATGA
CGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGgTGATGCGGTTTTGGCAGTACACCAATG
GGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGG
AGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTG
ACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAgcTCGTTTAGTGAA
CCGTCAGATCGCCTGGAACGCCATCCACGCTGTTTTGACCTCCATAGAAGACAGCGATCGCG
ccaccATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTGCACACCCTATGATATTAACCAGA
TGCTGAATGTCCTGGGAGACCACCAGGTCTCTGGCCTGGAGCAGCTGGAGAGCATCATCAA
CTTCGAGAAGCTGACCGAGTGGACAAGCTCCAATGTGATGCCTATCCTGTCCCCACTGACCA
AGGGCATCCTGGGCTTCGTGTTTACCCTGACAGTGCCTTCTGAGCGGGGCCTGTCTTGCATC
AGCGAGGCAGACGCAACCACACCAGAGTCCGCCAATCTGGGCGAGGAGATCCTGTCTCAGC

TGTACCTGTGGCCCCGGGTGACATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGG
AGAGCCAAGTACAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGCTGTTCGGCTACCCCGT
GTACGTGTTCGGCGATTGCGTGCAGGGCGACTGGGATGCCATCCGGTTTAGATACTGCGCA
CCACCTGGATATGCACTGCTGAGGTGTAACGACACCAATTATTCCGCCCTGCTGGCAGTGGG
CGCCCTGGAGGGCCCTCGCAATCAGGATTGGCTGGGCGTGCCAAGGCAGCTGGTGACACG
CATGCAGGCCATCCAGAACGCAGGCCTGTGCACCCTGGTGGCAATGCTGGAGGAGACAATC
TTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCGGCCCCAAGACAAACATCATCGT
GGATTCCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGTTTGTGGACTGGGAGA
ACGTGAGCCCAGAGCTGAATTCCACCGATCAGCCATTCTGGCAGGCAGGAATCCTGGCAAG
GAACCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCCAGAGC
CTGGTCATCAGCGCCTCCATCATCGTGTTTAACCTGCTGGAGCTGGAGGGCGACTATCGGG
ACGATGGCAACGTGTGGGTGCACACCCCACTGAGCCCCAGAACACTGAACGCCTGGGTGAA
GGCCGTGGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGGCCTCCATGACCAATATG
GAGCTGATGTCTAGCATCGTGCACCAGCAGGTGAGGACATACGGACCCGTGTTCATGTGCC
TGGGAGGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGC
TGTTCAGAGCCGCCCAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCGGAGC
AGCCTTTCGCCAGGTGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAG
TGGAACAATGAGACAACACAGCCTCAGATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTG
GCTGCACTACTATAGCGTGAGGGATACCCTGTGGCCCCGCGTGACATACCACATGAATAAGT
ACGCCTATCACATGCTGGAGAGGCGCGCCAAGTATAAGAGAGGCCCTGGCCCAGGCGCAAA
GTTTGTGGCAGCATGGACCCTGAAGGCCGCCGCCGGCCCCGGCCCCGGCCAGTATATCAA
GGCTAACAGTAAGTTCATTGGAATCACAGAGCTGGGACCCGGACCTGGATAATGAGTTTAAA
CTCCCATTTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTG
GACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGC
TTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGT
TTCAGGTTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTA
AAATAACTATAACGGTCCTAAGGTAGCGAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTG
CATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTTGCAGCAGCATGAGCGGAAG
CGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCCTCCTGGGCG
GGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGCGAACC
TCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGCAGC
TGCTGCATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCAC
TCTGGTGGCCAACTCGAGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGC
TGCTGATGGCCCAGCTCGAGGCCTTGACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGG

CTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGAAATCCAAATAAAAAATGAA
TCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTTATTTGATTTTTC
GCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCCAG
GACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTG
GAGGTAGCTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTTGTAAATCACCCAGTCATAG
CAGGGGCGCAGGGCATGGTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCA
GCCCTTTGGTGTAGGTGTTTACAAATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGAT
GAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTTACCGCCCAGATCCCGCCTGGGG
TTCATGTTGTGCAGGACCACCAGCACGGTGTATCCGGTGCACTTGGGGAATTTATCATGCAA
CTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCCAGGTTTTCCATG
CACTCATCCATGATGATGGCGATGGGCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGG
GGGTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTCATCATAGGCCATTTTAATGAATTT
GGGGCGGAGGGTGCCGGACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCC
CTCACAGATCTGCATCTCCCAGGCTTTGAGCTCGGAGGGGGGGATCATGTCCACCTGCGGG
GCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATGAGCTGGGCCGAAAGCAAGTTCCGG
AGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGACCGGCTGCAGG
TGGTAGTTGAGGGAGAGACAGCTGCCGTCCTCCCGGAGGAGGGGGGCCACCTCGTTCATCA
TCTCGCGCACGTGCATGTTCTCGCGCACCAGTTCCGCCAGGAGGCGCTCTCCCCCCAGGGA
TAGGAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCGGCCATGGGCATT
TTGGAGAGGGTTTGTTGCAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGG
CATCTCGATCCAGCAGACCTCCTCGTTTCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACC
AGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCTTCCAGGGTCGCAGCGTCCGCGTC
AGGGTGGTCTCCGTCACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGCGAGGGTGCG
CTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAG
GTAGCAATTGACCATGAGTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGC
TTACCTTTGGAAGTCTGCCCGCAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGG
GGGCGAGGAAGACGGACTCGGGGGCGTAGGCGTCCGCGCCGCAGTGGGCGCAGACGGTC
TCGCACTCCACGAGCCAGGTGAGGTCGGGCTGGTCGGGGTCAAAAACCAGTTTCCCGCCGT
TCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGTCCCCGCTGGGTGACAAAGA
GGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCTCGAGCGGTGTGCCGCGGTC
CTCCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAA
GGAGGCCACGTGGGACGGGTAGCGGTCGTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTA
TGCAAACACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCAC
GTGACCGGGGGTCCCGGCCGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCTCACT

GTCTTCCGGATCGCTGTCCAGGAGCGCCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCG
GGCATGACCTCGGCACTCAGGTTGTCAGTTTCTAGAAACGAGGAGGATTTGATATTGACGGT
GCCGGCGGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAGACGATCTTTTTGT
TGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGC
GCATGGTCTGGTTTTTTTCCTTGTCGGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTAC
TCGCGCGCCACGCACTTCCATTCGGGGAAGACGGTGGTCAGCTCGTCGGGCACGATTCTGA
CCTGCCAGCCCCGATTATGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAG
GGGCTCATTAGTCCAGCAGAGGCGTCCGCCCTTGCGCGAGCAGAAGGGGGGCAGGGGGTC
CAGCATGACCTCGTCGGGGGGGTCGGCATCGATGGTGAAGATGCCGGGCAGGAGGTCGGG
GTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTCGCGCACGGCC
AGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGA
GGCGTACATGCCGCAGATGTCGTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTG
GGGTAGCAGCGCCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGG
GCGAGGAGCCCCGGGCCCAGGTTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGG
CGGAAAATGGCATGCGAGTTGGAGGAGATGGTGGGCCTTTGGAAGATGTTGAAGTGGGCGT
GGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCTTGCAGCTTGGCGACGA
GCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGATGTCATA
CTTGAGCTGTCCCTTTTGTTTCCACAGCTCGCGGTTGAGAAGGAACTCTTCGCGGTCCTTCC
AGTACTCTTCGAGGGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTG
GTTGACGGCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGC
CTTGCGCAGGGAGGTGTGCGTGAGGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACTGG
TGCTTGAAGTCGATATCGTCGCAGCCCCCCTGCTCCCAGAGCTGGAAGTCCGTGCGCTTCTT
GTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGGATCTTGCCCGCGCGGGG
CATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGATGACCTGG
GCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGA
ATCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGG
GTCGCTGAGCCCGTGCTGCTCGAGCGCCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAA
GGAAGTCCAGAGATCCACGGCCAGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAACTG
CTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGTAGAAGGTGCGGGGGTCCCCGTGCCA
GCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACGAGCCGGTCGTCCCC
GGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCAGGTG
TAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGA
AGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGC
CGACGGCGCGCCGAACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCT

GCACGGGATGCACGTGCTGCACGAGCTGTACCTGAGTTCCTTTGACGAGGAATTTCAGTGG
GAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTACTACGTCGTGGTGGTCGGCCTGGCCC
TCTTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGGCAGGTCCAGACCTCGG
CGCGAGCGGGTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGTCCTG
AGACGCTGCGGAGTCAGGTCAGTGGGCAGCGGCGGCGCGCGGTTGACTTGCAGGAGTTTT
TCCAGGGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGT
CGATGGCTTGCAGGGTCCCGTGCCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGG
CGGCTGGGGCGACGGGGGCGGTGCCTCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGC
GCCGGGCGGCAGGGGCGGCTCGGGGCCCGGAGGCAGGGGCGGCAGGGGCACGTCGGCG
CCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGACGACGCGA
CGGTTGACGTCCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACC
TGAAAGAGAGTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGCAGGATCTCT
TGCACGTCGCCCGAGTTGTCCTGGTAGGCGATCTCGGTCATGAACTGCTCGATCTCCTCCTC
TTGAAGGTCTCCGCGGCCGGCGCGCTCCACGGTGGCCGCGAGGTCGTTGGAGATGCGGCC
CATGAGCTGCGAGAAGGCGTTCATGCCCGCCTCGTTCCAGACGCGGCTGTAGACCACGACG
CCCTCGGGATCGCgGGCGCGCATGACCACCTGGGCGAGGTTGAGCTCCACGTGGCGCGTG
AAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGCGATGTGCTCGG
TGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACGTCGCCCAGCGCCTC
CAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAAACTGGGAGTTGCGCGCC
GAGACGGTCAACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGC
GCTCGAAGGCCCCCGGGAGTTCCTCCACTTCCTCTTCTTCCTCCTCCACTAACATCTCTTCTA
CTTCCTCCTCAGGCGGCAGTGGTGGCGGGGGAGGGGGCCTGCGTCGCCGGCGGCGCACG
GGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCGGCGTCGCATGGTCTCGGTG
ACGGCGCGCCCGTCCTCGCGGGGCCGCAGCGTGAAGACGCCGCCGCGCATCTCCAGGTG
GCCGGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCC
CGTAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGA
ACGAAGGCTTCGAGCCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGT
CATGTTGGTTGGGAGCGGGGCGGGCGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCT
GAGACGGCGGATGGTGGCGAGGAGCACCAGGTCTTTGGGCCCGGCTTGCTGGATGCGCAG
ACGGTCGGCCATGCCCCAGGCGTGGTCCTGACACCTGGCCAGGTCCTTGTAGTAGTCCTGC
ATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGCATGCGCGTGAGCCCG
AAGCCGCGCTGGGGCTGGACGAGCGCCAGGTCGGCGACGACGCGCTCGGCGAGGATGGC
TTGCTGGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCT
CCGGTGTTGATGGTGTAGGAGCAGTTGGCCATGACGGACCAGTTGACGGTCTGGTGGCCCG

GACGCACGAGCTCGTGGTACTTGAGGCGCGAGTAGGCGCGCGTGTCGAAGATGTAGTCGTT
GCAGGTGCGCACCAGGTACTGGTAGCCGATGAGGAAGTGCGGCGGCGGCTGGCGGTAGAG
CGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCCTCGAGCATGGTGCGGTGGT
AGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGA
ACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTCATGGTGGGCACGGT
CTGGCCCGTGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTCAG
CGGCTCGACTCCGTGGCCTGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGT
TCGAATCTCGAATCAGGCTGGAGCCGCAGCTAACGTGGTATTGGCACTCCCGTCTCGACCCA
AGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCGTTTTGCAACTTTTTTTTGGAGGCCG
GATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCCGTAGTCTGGAG
AAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTCGAGGCCGGCCGGATTCCGCGGC
TAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTA
CGGAGCGAGCCCCTCTTTTGTTTTGTTTGTTTTTGCCAGATGCATCCCGTACTGCGGCAGAT
GCGCCCCCACCACCCTCCACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCG
CCCCAGCAGCAACTTCCAGCCACGACCGCCGCGGCCGCCGTGAGCGGGGCTGGACAGAGT
TATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTGGCGCGCCTGGGGGCGTCGTCG
CCGGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTACGTGCCCAAG
CAGAACCTGTTCAGAGACAGGAGCGGCGAGGAGCCCGAGGAGATGCGCGCGGCCCGGTTC
CACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGA
TTTCGAGGCGGACGAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAA
CCTGGTCACGGCGTACGAGCAGACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAAC
AACCACGTGCGCACCCTGATCGCGCGCGAGGAGGTGACCCTGGGCCTGATGCACCTGTGG
GACCTGCTGGAGGCCATCGTGCAGAACCCCACCAGCAAGCCGCTGACGGCGCAGCTGTTCC
TGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTCAGGGAGGCGCTGCTGAATATCACCGA
GCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGA
GCGCGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGC
AAGTACTACGCTAGGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGAT
CGACGGGTTTTACATGCGCATGACCCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTG
TACCGCAACGACAGGATGCACCGTGCGGTGAGCGCCAGCAGGCGGCGCGAGCTGAGCGAC
CAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGGGACCGAGGGGGAGAG
CTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCCTTGGAGGCGGC
GGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGA
AGACTGATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGC
GATGCGGGCGGCGCTGCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCA

GGCCATGCAACGCATCATGGCGCTGACGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCC
CAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGTGCCCTCGCGCTCCAACCCCACG
CACGAGAAGGTCCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCATCCGCGGCGAC
GAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAAC
GTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAG
CGGTTCCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCC
AGCCCGCCAACGTGCCCCGGGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCT
GATGGTGACCGAGGTGCCCCAGAGCGAGGTGTACCAGTCCGGGCCGGACTACTTCTTCCAG
ACCAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGCTTTCAAGAACTTGCAGGGCC
TGTGGGGCGTGCAGGCCCCGGTCGGGGACCGCGCGACGGTGTCGAGCCTGCTGACGCCG
AACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCAACCGCA
ACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGA
CGAGCAGACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGG
CAACCTGGAAGCCACCCTGAACTTTTTGCTGACCAACCGGTCGCAGAAGATCCCGCCCCAGT
ACGCGCTCAGCACCGAGGAGGAGCGCATCCTGCGTTACGTGCAGCAGAGCGTGGGCCTGT
TCCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACATGACCGCGCGCAACATGG
AGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGACTACTTGCATCGG
GCGGCCGCCATGAACTCTGACTATTTCACCAACGCCATCCTGAATCCCCACTGGCTCCCGCC
GCCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCCTGTGGGAC
GATGTGGACAGCAGCGTGTTCTCCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGA
AGGAAGGCAGCGACCGACGCCCGTCCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCG
GCGGTGCCCGAGGCCGCCAGTCCTTTCCCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCA
GCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGAAGAGGAGTACTTGAATG
ACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAAGCCTGGT
GGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTC
GCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGGTGGCACGACAGGCAGC
GGGGACAGATGTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGTTGGACTTGGGTG
GGAGTGGTAACCCGTTCGCTCACCTGCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACC
GAAAATAAATGATACTCACCAAGGCCATGGCGACCAGCGTGCGTTCGTTTCTTCTCTGTTGTT
GTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCCTCCTCCCTCGTACGAGAGCG
TGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTACGTGC
CCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCAC
CCTTGTACGATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCT
GAACTACCAGAACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCC

CCACGGAGGCCAGCACCCAGACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGC
TGAAAACCATCATGCACACCAACATGCCCAACGTGAACGAGTTCATGTACAGCAACAAGTTCA
AGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACAGTGACAGAGGATTATGA
TGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGGCAACTTCT
CGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGG
CGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGC
TGGGCTGGGACCCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCA
TCCCGATATTGTCTTGCTGCCCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAAC
CTGCTGGGCATTCGCAAGAGGCAGCCCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATC
TGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTCGACGCCTATGAGAAAAGCAAGGAGGA
TGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTCAGGGGCGATAAT
TTTGCAAGCGCCGCAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCA
TTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAAC
ACCGCCTACCGCAGCTGGTACCTAGCCTACAACTATGGCGACCCCGAGAAGGGCGTGCGCT
CCTGGACGCTGCTCACCACCTCGGACGTCACCTGCGGCGTGGAGCAAGTCTACTGGTCGCT
GCCCGACATGATGCAAGACCCGGTCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCG
GTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGAGCTTCTTCAACGAGCAGGCCGTCT
ACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACGTCTTCAACCGCTTCCCCGAGAAC
CAGATCCTCGTCCGCCCGCCCGCGCCCACCATTACCACCGTCAGTGAAAACGTTCCTGCTCT
CACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGT
TACTGACGCCAGACGCCGCACCTGCCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCG
CGCGTCCTCTCGAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCCCAGTAATAACACCGG
TTGGGGCCTGCGCGCGCCCAGCAAGATGTACGGAGGCGCTCGCCAACGCTCCACGCAACA
CCCCGTGCGCGTGCGCGGGCACTTCCGCGCTCCCTGGGGCGCCCTCAAGGGCCGCGTGC
GGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGCGCGCAACTACA
CCCCCGCCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTGGCcGACG
CGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGGAGC
ACCCCCGCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAG
GGCCATGCTCAGGGCGGCCAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGA
GACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCAGCATGTCCCGCCCGCGGCGAGGG
AACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTGCGCACCCGCCCC
CCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATGTCC
AAGCGCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGG
TGGTGAAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGA

AGAAAGTGATGTGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCCGGCGGCGCGT
GCAGTGGCGCGGGCGGAAGGTGCAACCGGTGCTGAGACCCGGCACCACCGTGGTCTTCAC
GCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTACGACGAGGTGTACGGGGATGA
TGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGCTTACGGCAAGCGCAGCCGT
TCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCCACGCCGAGC
CTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAAG
CGCGAGGGCGAGGATCTGTACCCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTG
GAAGACGTGCTGGAGACCATGAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGG
CCCATCAAGCAGGTGGCCCCGGGCCTGGGCGTGCAGACCGTGGACATCAAGATTCCCACG
GAGCCCATGGAAACGCAGACCGAGCCCATGATCAAGCCCAGCACCAGCACCATGGAGGTGC
AGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACGGCGCGG
CCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCG
CGGCACGCGCTTCTACCGCGGTCATACCAGCAGCCGCCGCCGCAAGACCACCACTCGCCG
CCGCCGTCGCCGCACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCG
CCGCGGCCGCGCACCTCTGACCCTGCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAA
ACTTTCGCCtGCTTTGCAGATCAATGGCCCTCACATGCCGCCTTCGCGTTCCCATTACGGGCT
ACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGGATGCGTCGCCACCACC
ACCGGCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCCTGCCCGCGCTGATCC
CCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCT
CTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAATAAACCaATGGACTCTGACGCTCCT
GGTCCTGTGATGTGTTTTCGTAGACAGATGGAAGACATCAATTTTTCGTCCCTGGCTCCGCG
ACACGGCACGCGGCCGTTCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGG
GGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGGCTTAAGAATTTCGGGTCCACGCTTAAAA
CCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCAGGCGCTGAGGGATAAGCTGAAAG
AGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGGGGTGGTGG
ACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCG
CCGGCTCCGTGGAGATGCCGCAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCG
AGAAGCGACCCCGCCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCC
CGTACGAGGAGGCGGTGAAACTGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCA
CCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCTGGACTTGCCTCCTCCCCAGCCTTCCCG
CCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCGCGCGCGACCCGGGGG
CACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAGTG
CAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTG
TGTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTC

GCCGAGTTGCAAGATGGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCC
GGACAGGACGCTTCGGAGTACCTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACA
CCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCCACGGTGGCGCCCACGCACGATGTGAC
CACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACCGCGAGGACAACAC
CTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACATGGCC
AGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCCTACTCCG
GCACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTCAGTGGACATATAAA
GCCGATGGTGAAACTGCCACAGAAAAAACCTATACATATGGAAATGCACCCGTGCAGGGCAT
TAACATCACAAAAGATGGTATTCAACTTGGAACTGACACCGATGATCAGCCAATCTACGCAGA
TAAAACCTATCAGCCTGAACCTCAAGTGGGTGATGCTGAATGGCATGACATCACTGGTACTG
ATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAATGAAGCCTTGTTATGGTTCTT
TTGCCAAGCCTACTAATAAAGAAGGAGGTCAGGCAAATGTGAAAACAGGAACAGGCACTACT
AAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCTAGC
TCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATAC
AAAGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTCAGCAAGCCATGCCCAACAG
ACCTAACTACATTGGTTTCAGAGACAACTTTATCGGGCTCATGTACTACAACAGCACTGGCAA
TATGGGGGTGCTGGCCGGTCAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGA
AACACCGAGCTGTCCTACCAGCTCTTGCTTGACTCTCTGGGTGACAGAACCCGGTATTTCAG
TATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTGCGCATTATTGAAAATCATGGTG
TGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAGAACAGATACTTATC
AGGGAATTAAGGCTAATGGAACTGATCAAACCACATGGACCAAAGATGACAGTGTCAATGAT
GCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACATCCAAGCCAACCTGTG
GAGGAACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGG
CCAATGTTACCCTGCCCACCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGC
GCCCTCGCTGGTGGACTCCTACATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATGGAC
AACGTGAACCCCTTCAACCACCACCGCAATGCGGGGCTGCGCTACCGCTCCATGCTCCTGG
GCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAAATTTTTCGCCATCAAGAGC
CTCCTGCTCCTGCCCGGGTCCTACACCTACGAGTGGAACTTCCGCAAGGACGTCAACATGAT
CCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTCACCAGC
ATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCAT
GCTGCGCAACGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCT
ACCCCATCCCGGCCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGC
CTTCCGCGGCTGGTCCTTCACGCGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGG
TTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTACCTCGACGGCACCTTCTACCTCAA

CCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCAGCTGGCCCGGCAACGACC
GGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTACAACGT
GGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATC
GGCTACCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCA
ACTTCCAGCCCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGT
CACCCTGGCCTACCAGCACAACAACTCGGGCTTCGTCGGCTACCTCGCGCCCACCATGCGC
CAGGGCCAGCCCTACCCCGCCAACTACCCCTACCCGCTCATCGGCAAGAGCGCCGTCACCA
GCGTCACCCAGAAAAAGTTCCTCTGCGACAGGGTCATGTGGCGCATCCCCTTCTCCAGCAAC
TTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAACTCCGCCCA
CGCGCTAGACATGAATTTCGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTT
CGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTAC
CTGCGCACCCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATG
GCCGCGGGCTCCGGCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCC
CTACTTCCTGGGCACCTTCGATAAGCGCTTCCCGGGATTCATGGCCCCGCACAAGCTGGCCT
GCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGAGCACTGGCTGGCCTTCGCCT
GGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGGACGAGCGCCT
CAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGCAGCGCCCTGGCCACCGA
GGACCGCTGCGTCACCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGC
CTGCGGGCTCTTCTGCTGCATGTTCCTGCACGCCTTCGTGCACTGGCCCGACCGCCCCATG
GACAAGAACCCCACCATGAACTTGCTGACGGGGGTGCCCAACGGCATGCTCCAGTCGCCCC
AGGTGGAACCCACCCTGCGCCGCAACCAGGAGGCGCTCTACCGCTTCCTCAACTCCCACTC
CGCCTACTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCCACCGCCTTCGACCGCATGAAT
CAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTTTCATGTTACACA
TGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCCGGGTCTCGGCATGGCCCGCG
GGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACTCGGGGATCAGCAGTT
TGGGCAGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGC
CCAGCAGGTCGGGCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGG
AGTTGCGGTACACGGGGTTGCAGCACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGC
CAGCACCGTCGCGTCGGTGATGCTCTCCACGTCGAGGTCCTCGGCGTTGGCCATCCCGAAG
GGGGTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCTTGTGGTTGCAAT
CGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTCGGCGTTCATCCCCGGGTACATGGC
CTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGA
CCCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAGC
GCGCGTCGTTGTTGGCCAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGC

CCGGTCGGGGTTCTCCTTCAGCGCGCGCTGCCCGTTCTCGCTCGCCACATCCATCTCGATC
ATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCACCGCAGCTTGCCCTCGGCCTCGG
TGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACTCCCAGTTCTTGTGGGCGATCTGGGA
ATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCTTGTTGCTA
GTGAAGGTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGT
ACACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTCAGGTCGGTCTCCACGCGGTA
GCGGTCCATCAGCATAGTCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGGGCAGGC
TCATAGGGTTCTTCACCATCATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTCGCTCTCGTC
CAGGGTCTCAAAGCTCCGCTTGCCGTCCTTCTCGGTGATCCGCACCGGGGGGTAGCTGAAG
CCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCCTCGCTGTCCTGGCTGACGTCCT
GCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGCGGCGGAGATGT
TGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCACTACTATCTCTTCCTCTTCTTGGT
CCGAGGCCACGCGGCGGTAGGTATGTCTCTTCGGGGGCAGAGGCGGAGGCGACGGGCTCT
CGCCGCCGCGACTTGGCGGATGGCTGGCAGAGCCCCTTCCGCGTTCGGGGGTGCGCTCCC
GGCGGCGCTCTGACTGACTTCCTCCGCGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAAC
AAGCATGGAGACTCAGCCATCGCCAACCTCGCCATCTGCCCCCACCGCCGACGAGAAGCAG
CAGCAGCAGAATGAAAGCTTAACCGCCCCGCCGCCCAGCCCCGCCACCTCCGACGCGGCC
GTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACGCCCG
CGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTCACAAGAAGAGATACACCAAGAACAGCC
AGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTC
CACCTGAGCGGGGGGGAGGACGCGCTCATCAAGCATCTGGCCCGGCAGGCCACCATCGTC
AAGGATGCGCTGCTCGACCGCACCGAGGTGCCCCTCAGCGTGGAGGAGCTCAGCCGCGCC
TACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCCAAGCGCCAGCCCAATGGCACCTGCG
AGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGGCCCTGGCCACCTA
CCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGCGCCG
ACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCCTACCTGATATCGCCTCCTTGGAAGAGGTT
CCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGA
GAAGGAGGAGAGCATGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGG
CTGGCGGTGCTCAAACGCACGGTCGAGCTGACCCATTTCGCCTACCCGGCTCTGAACCTGC
CCCCCAAAGTCATGAGCGCGGTCATGGACCAGGTGCTCATCAAGCGCGCGTCGCCCATCTC
CGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCAGCGACGAGCAGCT
GGCCCGGTGGCTGGGTCCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTCATG
ATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCG
GAGACCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCC

AGGCCTGCAAGATCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCAC
GAGAACCGCCTGGGGCAGAACGTGCTGCACACCACCCTGCGCGGGGAGGCCCGGCGCGA
CTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACACCTGGCAGACGGGCATGGGCGTG
TGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCCTGCAGAAGAACC
TCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTCAT
TTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGC
ATGTTGCAAAACTTTCGCTCTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTC
CGCGCTGCCCTCGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCCGCCGCTGTGGAG
CCACTGCTACCTGCTGCGCCTGGCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGAC
GTCAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTGCAACCTCTGCACGCCGCACCGC
TCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCACCTTCGAGTTGC
AAGGGCCCAGCGAAGGCGAGGGTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGGC
TGTGGACCTCGGCCTACTTGCGCAAGTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAG
GTTCTACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCGGCCTGCGTCATCACCCAG
GGGGCGATCCTGGCCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAA
GGGCCGCGGGGTCTACCTCGACCCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCA
GGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGCCGCCCGTGGAGGATTTGGAGG
AAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGACTGGGACAGCA
CTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGG
AGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCGTCCTCGGCGGGGGAGAAAGCAAGCA
GCACGGATACCATCTCCGCTCCGGGTCGGGGTCCCGCTCGACCACACAGTAGATGGGACGA
GACCGGACGATTCCCGAACCCCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAA
GTCCTGGCGGGGGCACAAAAACGCCATCGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATC
TCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACTTTCCCCGCAACATCTTGCA
TTACTACCGTCACCTCCACAGCCCCTACTACTTCCAAGAAGAGGCAGCAGCAGCAGAAAAAG
ACCAGCAGAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCG
CGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCCACCCTCTATGC
CATCTTCCAGCAGAGTCGGGGGCAGGAGCAGGAACTGAAAGTCAAGAACCGTTCTCTGCGC
TCGCTCACCCGCAGTTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGA
CGCCGAGGCTCTCTTCAACAAGTACTGCGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCC
CAGTCGCAGAAAAAGGCGGGAATTACGTCACCTGTGCCCTTCGCCCTAGCCGCCTCCACCC
ATCATCATGAGCAAAGAGATTCCCACGCCTTACATGTGGAGCTACCAGCCCCAGATGGGCCT
GGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCCCGC
GATGATCTCACGGGTGAATGACATCCGCGCCCACCGAAACCAGATACTCCTAGAACAGTCAG

CGCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTA
CCAGGAAATTCCCCAGCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAG
CTGACTAACTCAGGTGTCCAGCTGGCGGGCGGCGCCACCCTGTGTCGTCACCGCCCCGCTC
AGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCACACAGCTCAACGACGAGGTGGTGA
GCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACTCGCCGGATCGGGGAGATCTTC
CTTCACGCCTCGTCAGGCCGTCCTGACTTTGGAGAGTTCGTCCTCGCAGCCCCGCTCGGGT
GGCATCGGCACTCTCCAGTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTC
CGGCTCCCCCGGCCACTACCCGGACGAGTTCATCCCGAACTTCGACGCCATCAGCGAGTCG
GTGGACGGCTACGATTGAAACTAATCACCCCCTTATCCAGTGAAATAAAGATCATATTGATGA
TGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTT
GAGTTTAACAAAAAAATAAAGAATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTT
TCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTGCAGGCCCCGGCGGGCTG
CAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTAT
CTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCCGTCTACCCCT
ACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGTCTCTTCAGATGGA
TTCCAAGAGAAGCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCAAGA
ACGGGGAAATCACCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCTCGGGAAAACTCAT
CTCCAACACGGCCACCAAGGCCGCCGCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTA
ACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAA
TATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCG
TGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAA
GCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTG
GGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTT
TGGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATC
TGGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCA
CTTTGTGGACAACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAAAC
TAACACTTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTCAGTCTTAGTTGTAG
GAAGTGGAAACCTAAACCCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTT
GATGCAAACGGTGTTCTTTTAACAGAACATTCTACACTAAAAAAATACTGGGGGTATAGGCAG
GGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTCATGCCCAATTTAAAAGCT
TATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAG
ATGTTTCAAAACCTATGCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAGTACATA
TTCAATGTCATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAA
CTCTTATACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCC

TTCCCACCCCACTCTGTGGAACAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTAT
TGATTCAACAGTTTTACAGGATTCGAGCAGTTATTTTTCCTCCACCCTCCCAGGACATGGAAT
ACACCACCCTCTCCCCCCCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGACATGCTT
TTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGA
AACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAGGATTGTCCTCGGT
GGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCGCGA
ACGGGATCGGCCGGTGGTGTCGCATCAGGCCCCGCAGCAGTCGCTGCCGCCGCCGCTCCG
TCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGCCCT
CAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTCGCT
GCAGTACGTGCAACACAGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGC
CGAAACTCATCGCGGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAAT
CAAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGG
TTCACCACCTCCCGGTACCACATCACCCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCG
GAACCACAGGGCCAGCACCGCCCCGCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCA
ATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGT
TGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAA
ACCATATCCCAGGGCACGGGGAACTCTTGCAGGACAGCGAACCCCGCAGAACAGGGCAATC
CTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATCC
TCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATAC
GGGTGATGGCGGGACGCGGCTGATCGTGTTCGCGACCGTGTCATGATGCAGTTGCTTTCGG
ACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGCGGC
GGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGTGC
AGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATC
GACCACCGTGGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTTGGGTTTCGGTGACG
GCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCGGAGTAC
TTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAG
CCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACG
CGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATCAT
GTTACACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATTCGAACTAGT
TCcTGAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCAT
TCTTAAGCACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACCTGCAGCAGATTGACA
AGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTA
CTCTTTCATATCCTCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGC
CACAGTACAGATAAACCGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAG

CATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTT
TTAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGAT
GAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAA
TGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAGGC
AGGCCACGGGGTCTCCGGCGCGACCCTCGTAAAAATTGTCGCTATGATTGAAAACCATCACA
GAGAGACGTTCCCGGTGGCCGGCGTGAATGATTCGACAAGATGAATACACCCCCGGAACAT
TGGCGTCCGCGAGTGAAAAAAAGCGCCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCT
CAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGCGTACAAAATGT
AATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTCCAGGTACACATACAAAGCC
TCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGGCTG
AGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGG
CCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAACCGGTG
ACACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCAAAACTGCCGTCATTTCCGGGTTCC
CACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAAAACGTCACCCGCCC
CGCCCCTAACGGTCGCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACC
TCATTTGCATATTAACGCGCACAAAAAGTTTGAGG

Virus de la encefalitis equina venezolana [VEE] (SEQ ID NO:3) GenBank: L01442.2

atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa

gtgaacagcg atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag gcaagtctgg catcattaaa agcgcagtca ccaaaaaaga tctagtggtg agcgccaaga aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg

tgcggagcgc tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct caccgaaagt gcctgcgtat attcatccaa ggaagtatct cgtggaaaca ccaccggtag acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac

gcagagctgt tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc ggccccgcgc aggccctggt tccccagaac cgacccttt ctggcgatgc aggtgcagga attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa cagaaagggg gaggccaagg gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag aaaaccagca gagactttgg ccatgctcag cgttaacgtt gacaacccgg gctacgatga gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag acttcagact tcctcgcagt atggcctgga ttcctccggc aacttaaagg gcaggaccat gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg agcctcacta cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg aaatccacat

gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc taggatacca ttttgtctgg ctgtgctttg ctgcgcccgc actgcccggg ccgagaccac ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt gccttttta gtcatggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa ttaactgcag gtccgctttc cacagcttgg acacccttg atcgcaaaat cgtgcagtat gccggggaga tctataatta tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttgtt caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac cgagcaaggg tcggcgacta tccatttctc gaccgcaaat atccacccgg agttcaggct ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga aagaccatat tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt ttatttttc ttttcttttc cgaatcggat tttgttttta atatttc

VEE-MAG25mer (SEQ ID NO:4); contiene el nucleótido de MAG-25merPDTT (bases 30-1755)

atgggcggcgcatgagagaagcccagaccaattacctacccaaaatggagaaagttcacgttgacatcgaggaagacagcccattcc tcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttcgca tctggcttcaaaactgatcgaaacggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattctaa gcacaagtatcattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactgtaa ggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgagcgaccctgacctggaaactgagactatgtg cctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccgacaagtctctatca ccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgtttaagaacttggctggagcatatccatcata ctctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtagagggat gtccattcttagaaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttactgag

gagctggcacctgccgtctgtatttcacttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacgtcgt
taaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgcaaagtg
acagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattgtgtgaccaaatgactggcatactggcaa
cagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgcacccagagaaacacca
ataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatgaaagg
ccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccggatac
ccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgagaacaa
gaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgcagccgat
gaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccacctttggcagctgatgttgaggagcccactctggaagccg
atgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataaaggttaccagctacgctggcgaggac
aagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtcatagtg
ataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatacccgtcc
aggactttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatgg
aggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgacagga
aacagtgcgtcaagaaagaactagtcactgggctagggctcacaggcgagctggtggatcctcccttccatgaattcgcctacgagagt
ctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtgccaggatcaggcaagtctggcatcattaaaa
gcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaaagg
gctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagcttttgctt
gtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggttttttaac
atgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcg
gtcgtctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaa
acctaagcaggacgatctcattctcacttgtttcagagggtgggtgaagcagttgcaaatagattacaaaggcaacgaaataatgacgg
cagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgtacgcacccacctcagaac
atgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactgactgcca
agtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccc
taccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgaccactga
acaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttggactcg
atctggactccggtctattttctgcacccactgttccgttatccattaggaataatcactgggataactccccgtcgcctaacatgtacgggctg
aataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttgccactggaagagtctatgacatgaacactggt
acactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccaccataatgaacacccac
agagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaaatggttgac
tggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataatatttgttaat
gtgaggaccccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgtctgcatct
gaatcccggcggaacctgtgtcagcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcggcagttca

agttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtacgcac
aatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggatgtgcaccctcatatcatgtggtgcga
ggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcggaggggtgtgcggagcgctg
tataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaacatatcattcat
gccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaagattgtcaac
gataacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcattgaaccattt
gctgacagctttagacaccactgatgcagatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaagcagtggct
aggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaacctgatgcagagctggtgagggtgcatccgaag
agttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttggaagggaccaagtttcaccaggcggccaa
ggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaaagcatgagc
agtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgccatgactcca
gaaagagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcactggtgtgc
agaagatccaatgctcccagcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccaccggt
agacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccgaggatgagac
caggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttgctgtcagatggcccgacccaccaggt
gctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcatccgactttgatgtggacagttta
tccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtatggagttt
ctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttgcaccca
gcagggcctgctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttacccc
gtcacgcactcctagcaggtcggtctcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagaggagtttg
aggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcctccgacaccggtcaagggcatttacaacaaaaatc
agtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgtatgccccgcgcctcgaccaagaaaaagaa
gaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtggagaacatgaaagcc
ataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatcctgttcctt
tgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttccgactgtgg
cttcttactgtattattccagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccctgcaaagct
gcgcagctttccaaagaaacactcctatttggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctccagaacgtcc
tggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattggattcggcggcctttaatgtggaatgcttcaa
gaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaacgtggtaaattacattaccaa
attaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggtttgtaatggactt
aaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccgatccgcta
gcaacagcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgtttgatatgtc
ggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaaagtgagg
acgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtggacgcagagctgttgacgctgattgaggcggctttcggcga

aatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctggaatgttcctcacactgtttgtgaacacagtc
attaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcattggagatgacaatatcgtgaaa
ggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgggcgagaa
agcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctgtttaagctt
ggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtgg
gtattctttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactctagctagc
agtgttaaatcattcagctacctgagaggggcccctataactctctacggctaacctgaatggactacgactctagaatagtctttaatTAA
GCCACCATGGCAGGCATGTTTCAGGCGCTGAGCGAAGGCTGCACCCCGTATGATATTAACC
AGATGCTGAACGTGCTGGGCGATCATCAGGTCTCAGGCCTTGAGCAGCTTGAGAGTATAATC
AACTTTGAAAAACTGACTGAATGGACCAGTTCTAATGTTATGCCTATCCTGTCTCCTCTGACAA
AGGGCATCCTGGGCTTCGTGTTTACCCTGACCGTGCCTTCTGAGAGAGGACTTAGCTGCATT
AGCGAAGCGGATGCGACCACCCCGGAAAGCGCGAACCTGGGCGAAGAAATTCTGAGCCAG
CTGTATCTTTGGCCAAGGGTGACCTACCATTCCCCTAGTTATGCTTACCACCAATTTGAAAGA
CGAGCCAAATATAAAAGACACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCTACCCTGT
GTACGTGTTCGGCGATTGCGTGCAGGGCGATTGGGATGCGATTCGCTTTCGCTATTGCGCG
CCGCCGGGCTATGCGCTGCTGCGCTGCAACGATACCAACTATAGCGCTCTGCTGGCTGTGG
GGGCCCTAGAAGGACCCAGGAATCAGGACTGGCTTGGTGTCCCAAGACAACTTGTAACTCG
GATGCAGGCTATTCAGAATGCCGGCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCT
TCTGGCTGCAAGCGTTTCTGATGGCGCTGACCGATAGCGGCCCGAAAACCAACATTATTGTG
GATAGCCAGTATGTGATGGGCATTAGCAAACCGAGCTTTCAGGAATTTGTGGATTGGGAAAA
CGTGAGCCCGGAACTGAACAGCACCGATCAGCCGTTTTGGCAAGCCGGAATCCTGGCCAGA
AATCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTACCAGGGTCAGTCAC
TAGTCATCTCTGCTTCTATCATTGTCTTCAACCTGCTGGAACTGGAAGGTGATTATCGAGATG
ATGGCAACGTGTGGGTGCATACCCCGCTGAGCCCGCGCACCCTGAACGCGTGGGTGAAAG
CGGTGGAAGAAAAAAAAGGTATTCCAGTTCACCTAGAGCTGGCCAGTATGACCAACATGGAG
CTCATGAGCAGTATTGTGCATCAGCAGGTCAGAACATACGGCCCCGTGTTCATGTGTCTCGG
CGGACTGCTTACAATGGTGGCTGGTGCTGTGTGGCTGACAGTGCGAGTGCTCGAGCTGTTC
CGGGCCGCGCAGCTGGCCAACGACGTGGTCCTCCAGATCATGGAGCTTTGTGGTGCAGCGT
TTCGCCAGGTGTGCCATACCACCGTGCCGTGGCCGAACGCGAGCCTGACCCCGAAATGGAA
CAACGAAACCACCCAGCCCCAGATCGCCAACTGCAGCGTGTATGACTTTTTTGTGTGGCTCC
ATTATTATTCTGTTCGAGACACACTTTGGCCAAGGGTGACCTACCATATGAACAAATATGCGT
ATCATATGCTGGAAAGACGAGCCAAATATAAAAGAGGACCAGGACCTGGCGCTAAATTTGTG
GCCGCCTGGACACTGAAAGCCGCTGCTGGTCCTGGACCTGGCCAGTACATCAAGGCCAACA
GCAAGTTCATCGGCATCACCGAACTCGGACCCGGACCAGGCTGATGATTcgaacggccgtatcacgc

ccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagccgtaattattataattggcttg gtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgcttacatagaactc gcggcgattggcatgccgccttaaaatttttattttatttttcttttcttttccgaatcggattttgtttttaatatttcaaaaaaaaaaaaaaaaaaa aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
Virus de la encefalitis equina venezolana, cepa TC-83 [TC-83] (SEQ ID NO:5) GenBank: L01443.1
atgggcggcg catgagagaa gcccagacca attacctacc caaaatggag aaagttcacg ttgacatcga ggaagacagc ccattcctca gagctttgca gcggagcttc ccgcagtttg aggtagaagc caagcaggtc actgataatg accatgctaa tgccagagcg ttttcgcatc tggcttcaaa actgatcgaa acggaggtgg acccatccga cacgatcctt gacattggaa gtgcgcccgc ccgcagaatg tattctaagc acaagtatca ttgtatctgt ccgatgagat gtgcggaaga tccggacaga ttgtataagt atgcaactaa gctgaagaaa aactgtaagg aaataactga taaggaattg gacaagaaaa tgaaggagct cgccgccgtc atgagcgacc ctgacctgga aactgagact atgtgcctcc acgacgacga gtcgtgtcgc tacgaagggc aagtcgctgt ttaccaggat gtatacgcgg ttgacggacc gacaagtctc tatcaccaag ccaataaggg agttagagtc gcctactgga taggctttga caccacccct tttatgttta agaacttggc tggagcatat ccatcatact ctaccaactg ggccgacgaa accgtgttaa cggctcgtaa cataggccta tgcagctctg acgttatgga gcggtcacgt agagggatgt ccattcttag aaagaagtat ttgaaaccat ccaacaatgt tctattctct gttggctcga ccatctacca cgagaagagg gacttactga ggagctggca cctgccgtct gtatttcact tacgtggcaa gcaaaattac acatgtcggt gtgagactat agttagttgc gacgggtacg tcgttaaaag aatagctatc agtccaggcc tgtatgggaa gccttcaggc tatgctgcta cgatgcaccg cgagggattc ttgtgctgca aagtgacaga cacattgaac ggggagaggg tctcttttcc cgtgtgcacg tatgtgccag ctacattgtg tgaccaaatg actggcatac tggcaacaga tgtcagtgcg gacgacgcgc aaaaactgct ggttgggctc aaccagcgta tagtcgtcaa cggtcgcacc cagagaaaca ccaataccat gaaaaattac cttttgcccg tagtggccca ggcatttgct aggtgggcaa aggaatataa ggaagatcaa gaagatgaaa ggccactagg actacgagat agacagttag tcatggggtg ttgttgggct tttagaaggc acaagataac atctatttat aagcgcccgg atacccaaac catcatcaaa gtgaacagcg atttccactc attcgtgctg cccaggatag gcagtaacac attggagatc gggctgagaa caagaatcag gaaaatgtta gaggagcaca aggagccgtc acctctcatt accgccgagg acgtacaaga agctaagtgc gcagccgatg aggctaagga ggtgcgtgaa gccgaggagt tgcgcgcagc tctaccacct ttggcagctg atgttgagga gcccactctg gaagccgatg tcgacttgat gttacaagag gctggggccg gctcagtgga gacacctcgt ggcttgataa aggttaccag ctacgctggc gaggacaaga tcggctctta cgctgtgctt tctccgcagg ctgtactcaa gagtgaaaaa ttatcttgca tccaccctct cgctgaacaa gtcatagtga taacacactc tggccgaaaa gggcgttatg ccgtggaacc ataccatggt aaagtagtgg tgccagaggg acatgcaata cccgtccagg actttcaagc tctgagtgaa agtgccacca ttgtgtacaa cgaacgtgag ttcgtaaaca ggtacctgca ccatattgcc acacatggag gagcgctgaa cactgatgaa gaatattaca aaactgtcaa gcccagcgag cacgacggcg aatacctgta cgacatcgac aggaaacagt gcgtcaagaa agaactagtc actgggctag ggctcacagg cgagctggtg gatcctccct tccatgaatt cgcctacgag agtctgagaa cacgaccagc cgctccttac caagtaccaa ccataggggt gtatggcgtg ccaggatcag gcaagtctgg catcattaaa agcgcagtca

ccaaaaaaga tctagtggtg agcgccaaga aagaaaactg tgcagaaatt ataagggacg tcaagaaaat gaaagggctg gacgtcaatg ccagaactgt ggactcagtg ctcttgaatg gatgcaaaca ccccgtagag accctgtata ttgacgaagc ttttgcttgt catgcaggta ctctcagagc gctcatagcc attataagac ctaaaaaggc agtgctctgc ggggatccca aacagtgcgg ttttttaac atgatgtgcc tgaaagtgca ttttaaccac gagatttgca cacaagtctt ccacaaaagc atctctcgcc gttgcactaa atctgtgact tcggtcgtct caaccttgtt ttacgacaaa aaaatgagaa cgacgaatcc gaaagagact aagattgtga ttgacactac cggcagtacc aaacctaagc aggacgatct cattctcact tgtttcagag ggtgggtgaa gcagttgcaa atagattaca aaggcaacga aataatgacg gcagctgcct ctcaagggct gacccgtaaa ggtgtgtatg ccgttcggta caaggtgaat gaaaatcctc tgtacgcacc cacctcagaa catgtgaacg tcctactgac ccgcacggag gaccgcatcg tgtggaaaac actagccggc gacccatgga taaaaacact gactgccaag taccctggga atttcactgc cacgatagag gagtggcaag cagagcatga tgccatcatg aggcacatct tggagagacc ggaccctacc gacgtcttcc agaataaggc aaacgtgtgt tgggccaagg ctttagtgcc ggtgctgaag accgctggca tagacatgac cactgaacaa tggaacactg tggattattt tgaaacggac aaagctcact cagcagagat agtattgaac caactatgcg tgaggttctt tggactcgat ctggactccg gtctattttc tgcacccact gttccgttat ccattaggaa taatcactgg gataactccc cgtcgcctaa catgtacggg ctgaataaag aagtggtccg tcagctctct cgcaggtacc cacaactgcc tcgggcagtt gccactggaa gagtctatga catgaacact ggtacactgc gcaattatga tccgcgcata aacctagtac ctgtaaacag aagactgcct catgctttag tcctccacca taatgaacac ccacagagtg acttttcttc attcgtcagc aaattgaagg gcagaactgt cctggtggtc ggggaaaagt tgtccgtccc aggcaaaatg gttgactggt tgtcagaccg gcctgaggct accttcagag ctcggctgga tttaggcatc ccaggtgatg tgcccaaata tgacataata tttgttaatg tgaggacccc atataaatac catcactatc agcagtgtga agaccatgcc attaagctta gcatgttgac caagaaagct tgtctgcatc tgaatcccgg cggaacctgt gtcagcatag gttatggtta cgctgacagg gccagcgaaa gcatcattgg tgctatagcg cggcagttca agttttcccg ggtatgcaaa ccgaaatcct cacttgaaga gacggaagtt ctgtttgtat tcattgggta cgatcgcaag gcccgtacgc acaatcctta caagctttca tcaaccttga ccaacattta tacaggttcc agactccacg aagccggatg tgcaccctca tatcatgtgg tgcgagggga tattgccacg gccaccgaag gagtgattat aaatgctgct aacagcaaag gacaacctgg cggaggggtg tgcggagcgc tgtataagaa attcccggaa agcttcgatt tacagccgat cgaagtagga aaagcgcgac tggtcaaagg tgcagctaaa catatcattc atgccgtagg accaaacttc aacaaagttt cggaggttga aggtgacaaa cagttggcag aggcttatga gtccatcgct aagattgtca acgataacaa ttacaagtca gtagcgattc cactgttgtc caccggcatc ttttccggga acaaagatcg actaacccaa tcattgaacc atttgctgac agctttagac accactgatg cagatgtagc catatactgc agggacaaga aatgggaaat gactctcaag gaagcagtgg ctaggagaga agcagtggag gagatatgca tatccgacga ctcttcagtg acagaacctg atgcagagct ggtgagggtg catccgaaga gttctttggc tggaaggaag ggctacagca caagcgatgg caaaactttc tcatatttgg aagggaccaa gtttcaccag gcggccaagg atatagcaga aattaatgcc atgtggcccg ttgcaacgga ggccaatgag caggtatgca tgtatatcct cggagaaagc atgagcagta ttaggtcgaa atgccccgtc gaagagtcgg aagcctccac accacctagc acgctgcctt gcttgtgcat ccatgccatg actccagaaa gagtacagcg cctaaaagcc tcacgtccag aacaaattac tgtgtgctca tcctttccat tgccgaagta tagaatcact ggtgtgcaga agatccaatg ctcccagcct atattgttct caccgaaagt gcctgcgtat attcatccaa ggaagtatct

cgtggaaaca ccaccggtag acgagactcc ggagccatcg gcagagaacc aatccacaga ggggacacct gaacaaccac cacttataac cgaggatgag accaggacta gaacgcctga gccgatcatc atcgaagagg aagaagagga tagcataagt ttgctgtcag atggcccgac ccaccaggtg ctgcaagtcg aggcagacat tcacgggccg ccctctgtat ctagctcatc ctggtccatt cctcatgcat ccgactttga tgtggacagt ttatccatac ttgacaccct ggagggagct agcgtgacca gcggggcaac gtcagccgag actaactctt acttcgcaaa gagtatggag tttctggcgc gaccggtgcc tgcgcctcga acagtattca ggaaccctcc acatcccgct ccgcgcacaa gaacaccgtc acttgcaccc agcagggcct gctcgagaac cagcctagtt tccaccccgc caggcgtgaa tagggtgatc actagagagg agctcgaggc gcttaccccg tcacgcactc ctagcaggtc ggtctcgaga accagcctgg tctccaaccc gccaggcgta aatagggtga ttacaagaga ggagtttgag gcgttcgtag cacaacaaca atgacggttt gatgcgggtg catacatctt ttcctccgac accggtcaag ggcatttaca acaaaaatca gtaaggcaaa cggtgctatc cgaagtggtg ttggagagga ccgaattgga gatttcgtat gccccgcgcc tcgaccaaga aaaagaagaa ttactacgca agaaattaca gttaaatccc acacctgcta acagaagcag ataccagtcc aggaaggtgg agaacatgaa agccataaca gctagacgta ttctgcaagg cctagggcat tatttgaagg cagaaggaaa agtggagtgc taccgaaccc tgcatcctgt tcctttgtat tcatctagtg tgaaccgtgc cttttcaagc cccaaggtcg cagtggaagc ctgtaacgcc atgttgaaag agaactttcc gactgtggct tcttactgta ttattccaga gtacgatgcc tatttggaca tggttgacgg agcttcatgc tgcttagaca ctgccagttt ttgccctgca aagctgcgca gctttccaaa gaaacactcc tatttggaac ccacaatacg atcggcagtg ccttcagcga tccagaacac gctccagaac gtcctggcag ctgccacaaa aagaaattgc aatgtcacgc aaatgagaga attgcccgta ttggattcgg cggcctttaa tgtggaatgc ttcaagaaat atgcgtgtaa taatgaatat tgggaaacgt ttaaagaaaa ccccatcagg cttactgaag aaaacgtggt aaattacatt accaaattaa aaggaccaaa agctgctgct ctttttgcga agacacataa tttgaatatg ttgcaggaca taccaatgga caggtttgta atggacttaa agagagacgt gaaagtgact ccaggaacaa aacatactga agaacggccc aaggtacagg tgatccaggc tgccgatccg ctagcaacag cgtatctgtg cggaatccac cgagagctgg ttaggagatt aaatgcggtc ctgcttccga acattcatac actgtttgat atgtcggctg aagactttga cgctattata gccgagcact tccagcctgg ggattgtgtt ctggaaactg acatcgcgtc gtttgataaa agtgaggacg acgccatggc tctgaccgcg ttaatgattc tggaagactt aggtgtggac gcagagctgt tgacgctgat tgaggcggct ttcggcgaaa tttcatcaat acatttgccc actaaaacta aatttaaatt cggagccatg atgaaatctg gaatgttcct cacactgttt gtgaacacag tcattaacat tgtaatcgca agcagagtgt tgagagaacg gctaaccgga tcaccatgtg cagcattcat tggagatgac aatatcgtga aaggagtcaa atcggacaaa ttaatggcag acaggtgcgc cacctggttg aatatggaag tcaagattat agatgctgtg gtgggcgaga aagcgcctta tttctgtgga gggtttattt tgtgtgactc cgtgaccggc acagcgtgcc gtgtggcaga ccccctaaaa aggctgttta agcttggcaa acctctggca gcagacgatg aacatgatga tgacaggaga agggcattgc atgaagagtc aacacgctgg aaccgagtgg gtattctttc agagctgtgc aaggcagtag aatcaaggta tgaaaccgta ggaacttcca tcatagttat ggccatgact actctagcta gcagtgttaa atcattcagc tacctgagag gggcccctat aactctctac ggctaacctg aatggactac gacatagtct agtccgccaa gatgttcccg ttccagccaa tgtatccgat gcagccaatg ccctatcgca acccgttcgc ggccccgcgc aggccctggt tccccagaac cgacccttttt ctggcgatgc aggtgcagga attaacccgc tcgatggcta acctgacgtt caagcaacgc cgggacgcgc cacctgaggg gccatccgct aagaaaccga agaaggaggc ctcgcaaaaa

cagaaagggg gaggccaagg gaagaagaag aagaaccaag ggaagaagaa ggctaagaca gggccgccta atccgaaggc acagaatgga aacaagaaga agaccaacaa gaaaccaggc aagagacagc gcatggtcat gaaattggaa tctgacaaga cgttcccaat catgttggaa gggaagataa acggctacgc ttgtgtggtc ggagggaagt tattcaggcc gatgcatgtg gaaggcaaga tcgacaacga cgttctggcc gcgcttaaga cgaagaaagc atccaaatac gatcttgagt atgcagatgt gccacagaac atgcgggccg atacattcaa atacacccat gagaaacccc aaggctatta cagctggcat catggagcag tccaatatga aaatgggcgt ttcacggtgc cgaaaggagt tggggccaag ggagacagcg gacgacccat tctggataac cagggacggg tggtcgctat tgtgctggga ggtgtgaatg aaggatctag gacagccctt tcagtcgtca tgtggaacga gaagggagtt accgtgaagt atactccgga gaactgcgag caatggtcac tagtgaccac catgtgtctg ctcgccaatg tgacgttccc atgtgctcaa ccaccaattt gctacgacag aaaaccagca gagactttgg ccatgctcag cgttaacgtt gacaacccgg gctacgatga gctgctggaa gcagctgtta agtgccccgg aaggaaaagg agatccaccg aggagctgtt taaggagtat aagctaacgc gcccttacat ggccagatgc atcagatgtg cagttgggag ctgccatagt ccaatagcaa tcgaggcagt aaagagcgac gggcacgacg gttatgttag acttcagact tcctcgcagt atggcctgga ttcctccggc aacttaaagg gcaggaccat gcggtatgac atgcacggga ccattaaaga gataccacta catcaagtgt cactccatac atctcgcccg tgtcacattg tggatgggca cggttatttc ctgcttgcca ggtgcccggc aggggactcc atcaccatgg aatttaagaa agattccgtc acacactcct gctcggtgcc gtatgaagtg aaatttaatc ctgtaggcag agaactctat actcatcccc cagaacacgg agtagagcaa gcgtgccaag tctacgcaca tgatgcacag aacagaggag cttatgtcga gatgcacctc ccgggctcag aagtggacag cagtttggtt tccttgagcg gcagttcagt caccgtgaca cctcctgttg ggactagcgc cctggtggaa tgcgagtgtg gcggcacaaa gatctccgag accatcaaca agacaaaaca gttcagccag tgcacaaaga aggagcagtg cagagcatat cggctgcaga acgataagtg ggtgtataat tctgacaaac tgcccaaagc agcgggagcc accttaaaag gaaaactgca tgtcccattc ttgctggcag acggcaaatg caccgtgcct ctagcaccag aacctatgat aacctttggt ttcagatcag tgtcactgaa actgcaccct aagaatccca catatctaac cacccgccaa cttgctgatg agcctcacta cacgcacgag ctcatatctg aaccagctgt taggaatttt accgtcaccg aaaaagggtg ggagtttgta tggggaaacc acccgccgaa aaggttttgg gcacaggaaa cagcacccgg aaatccacat gggctaccgc acgaggtgat aactcattat taccacagat accctatgtc caccatcctg ggtttgtcaa tttgtgccgc cattgcaacc gtttccgttg cagcgtctac ctggctgttt tgcagatcta gagttgcgtg cctaactcct taccggctaa cacctaacgc taggatacca ttttgtctgg ctgtgctttg ctgcgcccgc actgcccggg ccgagaccac ctgggagtcc ttggatcacc tatggaacaa taaccaacag atgttctgga ttcaattgct gatccctctg gccgccttga tcgtagtgac tcgcctgctc aggtgcgtgt gctgtgtcgt gccttttta gtcatggccg gcgccgcagg cgccggcgcc tacgagcacg cgaccacgat gccgagccaa gcgggaatct cgtataacac tatagtcaac agagcaggct acgcaccact ccctatcagc ataacaccaa caaagatcaa gctgatacct acagtgaact tggagtacgt cacctgccac tacaaaacag gaatggattc accagccatc aaatgctgcg gatctcagga atgcactcca acttacaggc ctgatgaaca gtgcaaagtc ttcacagggg tttacccgtt catgtggggt ggtgcatatt gcttttgcga cactgagaac acccaagtca gcaaggccta cgtaatgaaa tctgacgact gccttgcgga tcatgctgaa gcatataaag cgcacacagc ctcagtgcag gcgttcctca acatcacagt gggagaacac tctattgtga ctaccgtgta tgtgaatgga gaaactcctg tgaatttcaa tggggtcaaa ttaactgcag gtccgctttc cacagcttgg acacccttg atcgcaaaat cgtgcagtat

gccggggaga tctataatta tgattttcct gagtatgggg caggacaacc aggagcattt ggagatatac aatccagaac agtctcaagc tcagatctgt atgccaatac caacctagtg ctgcagagac ccaaagcagg agcgatccac gtgccataca ctcaggcacc ttcgggtttt gagcaatgga agaaagataa agctccatca ttgaaattta ccgccccttt cggatgcgaa atatatacaa accccattcg cgccgaaaac tgtgctgtag ggtcaattcc attagccttt gacattcccg acgccttgtt caccagggtg tcagaaacac cgacactttc agcggccgaa tgcactctta acgagtgcgt gtattcttcc gactttggtg ggatcgccac ggtcaagtac tcggccagca agtcaggcaa gtgcgcagtc catgtgccat cagggactgc taccctaaaa gaagcagcag tcgagctaac cgagcaaggg tcggcgacta tccatttctc gaccgcaaat atccacccgg agttcaggct ccaaatatgc acatcatatg ttacgtgcaa aggtgattgt caccccccga aagaccatat tgtgacacac cctcagtatc acgcccaaac atttacagcc gcggtgtcaa aaaccgcgtg gacgtggtta acatccctgc tgggaggatc agccgtaatt attataattg gcttggtgct ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca attggcaagc tgcttacata gaactcgcgg cgattggcat gccgccttaa aatttttatt ttatttttc ttttcttttc cgaatcggat tttgttttta atatttc
Vector de suministro del VEE (SEQ ID NO:6); genoma del VEE con los nucleótidos 7544-11175 eliminados [proteínas estructurales del alfavirus removidas]
ATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcgaggaagacagccca ttcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttc gcatctggcttcaaaactgatcgaaacggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtatt ctaagcacaagtatcattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaact gtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgagcgaccctgacctggaaactgagact atgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccgacaagtctc tatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgtttaagaacttggctggagcatatcca tcatactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtaga gggatgtccattcttagaaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggactta ctgaggagctggcacctgccgtctgtatttcacttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggta cgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgca aagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattgtgtgaccaaatgactggcatact ggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgcacccagagaaa caccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatg aaaggccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgccc ggatacccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctga gaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgca gccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccacctttggcagctgatgttgaggagcccactctgga agccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataaaggttaccagctacgctggcg aggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtc

atagtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatac
ccgtccaggactttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccac
acatggaggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcg
acaggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacaggcgagctggtggatcctcccttccatgaattcgcctac
gagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtgccaggatcaggcaagtctggcatca
ttaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatg
aaagggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagc
ttttgcttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggttt
ttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtg
acttcggtcgtctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagt
accaaacctaagcaggacgatctcattctcacttgtttcagagggtgggtgaagcagttgcaaatagattacaaaggcaacgaaataat
gacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgtacgcacccacctc
agaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactga
ctgccaagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagacc
ggaccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatga
ccactgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttcttt
ggactcgatctggactccggtctattttctgcacccactgttccgttatccattaggaataatcactgggataactccccgtcgcctaacatgta
cgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttgccactggaagagtctatgacatga
acactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccaccataatga
acacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaa
atggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataat
atttgttaatgtgaggaccccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagctt
gtctgcatctgaatcccggcggaacctgtgtcagcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcg
gcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggccc
gtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggatgtgcaccctcatatcat
gtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcggaggggtgtgc
ggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaac
atatcattcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaa
gattgtcaacgataacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcat
tgaaccatttgctgacagctttagacaccactgatgcagatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaa
gcagtggctaggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaacctgatgcagagctggtgagggtg
catccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttggaagggaccaagtttcaccag
gcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaaa
gcatgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgcca

tgactccagaaagagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcac
tggtgtgcagaagatccaatgctcccagcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacac
caccggtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccgagga
tgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttgctgtcagatggcccgaccca
ccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcatccgactttgatgtgga
cagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtat
ggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttg
cacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgc
ttaccccgtcacgcactcctagcaggtcggtctcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagag
gagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcctccgacaccggtcaagggcatttacaaca
aaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgtatgccccgcgcctcgaccaagaa
aaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtggagaacatg
aaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatc
ctgttcctttgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttcc
gactgtggcttcttactgtattattccagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccct
gcaaagctgcgcagctttccaaagaaacactcctatttggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctcca
gaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattggattcggcggcctttaatgtgga
atgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaacgtggtaaattac
attaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggtttgt
aatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccg
atccgctagcaacagcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgttt
gatatgtcggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaa
agtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtggacgcagagctgttgacgctgattgaggcggctt
tcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctggaatgttcctcacactgtttgtgaa
cacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcattggagatgacaatat
cgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgg
gcgagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctg
tttaagcttggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaacc
gagtgggtattctttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactcta
gctagcagtgttaaatcattcagctacctgagaggggcccctataactctctacggcTAAcctgaatggactacgactatcacgcccaa
acatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagccgtaattattataattggcttggtgct
ggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgcttacatagaactcgcgg
cgattggcatgccgccttaaaatttttattttatttttctttttctttttccgaatcggattttgtttttaatatttcAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Vector de suministro de la TC-83 (SEQ ID NO:7); genoma de la TC-83 con los nucleótidos 7544-11175 eliminados [proteínas estructurales del alfavirus removidas]
ATAGGCGGCGCATGAGAGAAGCCCAGACCAATTACCTACCCAAAATGGAGAAAGTTCACGTT GACATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGG TAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGAGCGTTTTCGCATCTGGCT TCAAAACTGATCGAAACGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCC CGCCCGCAGAATGTATTCTAAGCACAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAG ATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAAACTGTAAGGAAATAACTGATA AGGAATTGGACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGCGACCCTGACCTGGAAAC TGAGACTATGTGCCTCCACGACGACGAGTCGTGTCGCTACGAAGGGCAAGTCGCTGTTTACC AGGATGTATACGCGGTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGA GTCGCCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATAT CCATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATG CAGCTCTGACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGA AACCATCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTAC TGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGCAAAATTACACATGTCGG TGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAGCTATCAGTCCAGGCCT GTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGGATTCTTGTGCTGCAAAG TGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCAGCTACA TTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCGCAAAAACT GCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAACACCAATACCA TGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCAAAGGAATATAAG GAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAGTTAGTCATGGGGTGTTG TTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCGGATACCCAAACCATCAT CAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATAGGCAGTAACACATTGGAGA TCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCACAAGGAGCCGTCACCTCTCATT ACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGATGAGGCTAAGGAGGTGCGTGAA GCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCCACTCTGG AAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGGAGACACCTCGTGG CTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTC CGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCCTCTCGCTGAACAAGTCATAG TGATAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATACCATGGTAAAGTAGTG GTGCCAGAGGGACATGCAATACCCGTCCAGGACTTTCAAGCTCTGAGTGAAAGTGCCACCAT

TGTGTACAACGAACGTGAGTTCGTAAACAGGTACCTGCACCATATTGCCACACATGGAGGAG
CGCTGAACACTGATGAAGAATATTACAAAACTGTCAAGCCCAGCGAGCACGACGGCGAATAC
CTGTACGACATCGACAGGAAACAGTGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCAC
AGGCGAGCTGGTGGATCCTCCCTTCCATGAATTCGCCTACGAGAGTCTGAGAACACGACCA
GCCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTG
GCATCATTAAAAGCGCAGTCACCAAAAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACTGT
GCAGAAATTATAAGGGACGTCAAGAAAATGAAAGGGCTGGACGTCAATGCCAGAACTGTGGA
CTCAGTGCTCTTGAATGGATGCAAACACCCCGTAGAGACCCTGTATATTGACGAAGCTTTTGC
TTGTCATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAAGGCAGTGCTCTG
CGGGGATCCCAAACAGTGCGGTTTTTTTAACATGATGTGCCTGAAAGTGCATTTTAACCACGA
GATTTGCACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGT
CGTCTCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGT
GATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCATTCTCACTTGTTTCAGAG
GGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCT
CAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGTGAATGAAAATCCTCTGTA
CGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGGAGGACCGCATCGTGTGG
AAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAAGTACCCTGGGAATTTCAC
TGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAGAGA
CCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCTTTAGTGCC
GGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTTG
AAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGA
CTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTATCCATTAGGAATAATCAC
TGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAAGAAGTGGTCCGTCAGCTCTC
TCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACACTG
GTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATG
CTTTAGTCCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTCGTCAGCAAATTGAA
GGGCAGAACTGTCCTGGTGGTCGGGGAAAAGTTGTCCGTCCCAGGCAAAATGGTTGACTGG
TTGTCAGACCGGCCTGAGGCTACCTTCAGAGCTCGGCTGGATTTAGGCATCCCAGGTGATGT
GCCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATATAAATACCATCACTATCAGCAG
TGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCC
GGCGGAACCTGTGTCAGCATAGGTTATGGTTACGCTGACAGGGCCAGCGAAAGCATCATTG
GTGCTATAGCGCGGCAGTTCAAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTTGAAGAG
ACGGAAGTTCTGTTTGTATTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAA
GCTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACC

CTCATATCATGTGGTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTG
CTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGA
AAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAAC
ATATCATTCATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGGTTGAAGGTGACAAACAGT
TGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAATTACAAGTCAGTAGCGA
TTCCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATCATTGAACC
ATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAAT
GGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATC
CGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGAGTTCTT
TGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTTCTCATATTTGGAAGGGACC
AAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGA
GGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCATGAGCAGTATTAGGTCGAAAT
GCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACGCTGCCTTGCTTGTGCATCCA
TGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCAGAACAAATTACTGTGT
GCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAGAAGATCCAATGCTCCCAG
CCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCAAGGAAGTATCTCGTGGAAACA
CCACCGGTAGACGAGACTCCGGAGCCATCGGCAGAGAACCAATCCACAGAGGGGACACCTG
AACAACCACCACTTATAACCGAGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATCATC
GAAGAGGAAGAAGAGGATAGCATAAGTTTGCTGTCAGATGGCCCGACCCACCAGGTGCTGC
AAGTCGAGGCAGACATTCACGGGCCGCCCTCTGTATCTAGCTCATCCTGGTCCATTCCTCAT
GCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCTGGAGGGAGCTAGCGTGAC
CAGCGGGGCAACGTCAGCCGAGACTAACTCTTACTTCGCAAAGAGTATGGAGTTTCTGGCGC
GACCGGTGCCTGCGCCTCGAACAGTATTCAGGAACCCTCCACATCCCGCTCCGCGCACAAG
AACACCGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCCGCCA
GGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTA
GCAGGTCGGTCTCGAGAACCAGCCTGGTCTCCAACCCGCCAGGCGTAAATAGGGTGATTAC
AAGAGAGGAGTTTGAGGCGTTCGTAGCACAACAACAATGACGGTTTGATGCGGGTGCATACA
TCTTTTCCTCCGACACCGGTCAAGGGCATTTACAACAAAAATCAGTAAGGCAAACGGTGCTAT
CCGAAGTGGTGTTGGAGAGGACCGAATTGGAGATTTCGTATGCCCCGCGCCTCGACCAAGA
AAAAGAAGAATTACTACGCAAGAAATTACAGTTAAATCCCACACCTGCTAACAGAAGCAGATA
CCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAG
GGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAACCCTGCATCCTGTTCCTTTG
TATTCATCTAGTGTGAACCGTGCCTTTTCAAGCCCCAAGGTCGCAGTGGAAGCCTGTAACGC
CATGTTGAAAGAGAACTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTA

TTTGGACATGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGCCCTGCAAAGCT
GCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAG
CGATCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAAGAAATTGCAATGTCACG
CAAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAATGTGGAATGCTTCAAGAAATAT
GCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAGAAAAC
GTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGCTCTTTTTGCGAAGACACAT
AATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAATGGACTTAAAGAGAGACGTG
AAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGGTACAGGTGATCCAGGCTGC
CGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGAGATTAAATG
CGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGCTGAAGACTTTGACGCTATTA
TAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTGACATCGCGTCGTTTGATAAA
AGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTCTGGAAGACTTAGGTGTGGACG
CAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTCATCAATACATTTGCCCACTA
AAACTAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCCTCACACTGTTTGTGAACAC
AGTCATTAACATTGTAATCGCAAGCAGAGTGTTGAGAGAACGGCTAACCGGATCACCATGTG
CAGCATTCATTGGAGATGACAATATCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACA
GGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGC
GCCTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGG
CAGACCCCCTAAAAAGGCTGTTTAAGCTTGGCAAACCTCTGGCAGCAGACGATGAACATGAT
GATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTATTCTTTC
AGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAAACCGTAGGAACTTCCATCATAGTTATGG
CCATGACTACTCTAGCTAGCAGTGTTAAATCATTCAGCTACCTGAGAGGGGCCCCTATAACTC
TCTACGGCTAACCTGAATGGACTACGACTATCACGCCCAAACATTTACAGCCGCGGTGTCAA
AAACCGCGTGGACGTGGTTAACATCCCTGCTGGGAGGATCAGCCGTAATTATTATAATTGGC
TTGGTGCTGGCTACTATTGTGGCCATGTACGTGCTGACCAACCAGAAACATAATTGAATACAG
CAGCAATTGGCAAGCTGCTTACATAGAACTCGCGGCGATTGGCATGCCGCCTTAAAATTTTTA
TTTTATTTTTCTTTTCTTTTCCGAATCGGATTTTGTTTTTAATATTTCAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Vector de producción del VEE (SEQ ID NO:8); genoma del VEE con los nucleótidos 7544-11175 eliminados, más el promotor 5’ de T7, más sitios de restricción 3’

TAATACGACTCACTATAGGATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttc
acgttgacatcgaggaagacagcccattcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgat
aatgaccatgctaatgccagagcgttttcgcatctggcttcaaaactgatcgaaacggaggtggacccatccgacacgatccttgacattg

gaagtgcgcccgcccgcagaatgtattctaagcacaagtatcattgtatctgtccgatgagatgtgcggaagatccggacagattgtataa
gtatgcaactaagctgaagaaaaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgag
cgaccctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtata
cgcggttgacggaccgacaagtctctatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatg
tttaagaacttggctggagcatatccatcatactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagct
ctgacgttatggagcggtcacgtagagggatgtccattcttagaaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgac
catctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcacttacgtggcaagcaaaattacacatgtcggtgt
gagactatagttagttgcgacgggtacgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatgctgctacgat
gcaccgcgagggattcttgtgctgcaaagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctaca
ttgtgtgaccaaatgactggcatactggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagt
cgtcaacggtcgcacccagagaaacaccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaagg
aatataaggaagatcaagaagatgaaaggccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcac
aagataacatctatttataagcgcccggatacccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggc
agtaacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgag
gacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccacctttggca
gctgatgttgaggagcccactctggaagccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttg
ataaaggttaccagctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttg
catccaccctctcgctgaacaagtcatagtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagt
ggtgccagagggacatgcaatacccgtccaggactttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaa
acaggtacctgcaccatattgccacacatggaggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacg
acggcgaatacctgtacgacatcgacaggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacaggcgagctggtg
gatcctcccttccatgaattcgcctacgagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtg
ccaggatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcaga
aattataagggacgtcaagaaaatgaaagggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacacccc
gtagagaccctgtatattgacgaagcttttgcttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgct
ctgcggggatcccaaacagtgcggtttttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaa
gcatctctcgccgttgcactaaatctgtgacttcggtcgtctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagact
aagattgtgattgacactaccggcagtaccaaacctaagcaggacgatctcattctcacttgtttcagagggtgggtgaagcagttgcaaa
tagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaat
gaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccg
gcgacccatggataaaaacactgactgccaagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccat
catgaggcacatcttggagagaccggaccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgct
gaagaccgctggcatagacatgaccactgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtat
tgaaccaactatgcgtgaggttctttggactcgatctggactccggtctattttctgcacccactgttccgttatccattaggaataatcactggg

ataactccccgtcgcctaacatgtacgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttg
ccactggaagagtctatgacatgaacactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctc
atgctttagtcctccaccataatgaacacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcgggg
aaaagttgtccgtcccaggcaaaatggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccagg
tgatgtgcccaaatatgacataatatttgttaatgtgaggaccccatataaataccatcactatcagcagtgtgaagaccatgccattaagct
tagcatgttgaccaagaaagcttgtctgcatctgaatcccggcggaacctgtgtcagcataggttatggttacgctgacagggccagcga
aagcatcattggtgctatagcgcggcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagttctgtttgtatt
cattgggtacgatcgcaaggcccgtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaa
gccggatgtgcaccctcatatcatgtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaagg
acaacctggcggaggggtgtgcggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcg
actggtcaaaggtgcagctaaacatatcattcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggc
agaggcttatgagtccatcgctaagattgtcaacgataacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaa
caaagatcgactaacccaatcattgaaccatttgctgacagctttagacaccactgatgcagatgtagccatatactgcagggacaagaa
atgggaaatgactctcaaggaagcagtggctaggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaacc
tgatgcagagctggtgagggtgcatccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatattt
ggaagggaccaagtttcaccaggcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggt
atgcatgtatatcctcggagaaagcatgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacg
ctgccttgcttgtgcatccatgccatgactccagaaagagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatccttt
ccattgccgaagtatagaatcactggtgtgcagaagatccaatgctcccagcctatattgttctcaccgaaagtgcctgcgtatattcatcca
aggaagtatctcgtggaaacaccaccggtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaac
aaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagt
ttgctgtcagatggcccgacccaccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcc
tcatgcatccgactttgatgtggacagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagac
taactcttacttcgcaaagagtatggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccg
cgcacaagaacaccgtcacttgcacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatc
actagagaggagctcgaggcgcttaccccgtcacgcactcctagcaggtcggtctcgagaaccagcctggtctccaacccgccaggcg
taaatagggtgattacaagagaggagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcctccgac
accggtcaagggcatttacaacaaaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgta
tgccccgcgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagatacca
gtccaggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagt
ggagtgctaccgaaccctgcatcctgttcctttgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaac
gccatgttgaaagagaactttccgactgtggcttcttactgtattattccagagtacgatgcctatttggacatggttgacggagcttcatgctg
cttagacactgccagtttttgccctgcaaagctgcgcagctttccaaagaaacactcctatttggaacccacaatacgatcggcagtgcctt
cagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtatt

ggattcggcggcctttaatgtggaatgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggctta
ctgaagaaaacgtggtaaattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcag
gacataccaatggacaggtttgtaatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaa
ggtacaggtgatccaggctgccgatccgctagcaacagcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcct
gcttccgaacattcatacactgtttgatatgtcggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaa
actgacatcgcgtcgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtggacgcagag
ctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctg
gaatgttcctcacactgtttgtgaacacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgc
agcattcattggagatgacaatatcgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaag
tcaagattatagatgctgtggtgggcgagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgt
ggcagacccccctaaaaaggctgtttaagcttggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggcattgcat
gaagagtcaacacgctggaaccgagtgggtattctttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtaggaacttccat
catagttatggccatgactactctagctagcagtgttaaatcattcagctacctgagaggggcccctataactctctacggcTAAcctgaat
ggactacgactatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagcc
gtaattattataattggcttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggca
agctgcttacatagaactcgcggcgattggcatgccgccttaaaatttttattttatttttctttctttccgaatcggattttgtttttaatatttcAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAA tacgtagtttaaac

Vector de producción de la TC-83 (SEQ ID NO:9); genoma de la TC-83 con los nucleótidos 7544-11175 eliminados, más el promotor 5’ de T7, más sitios de restricción 3’

TAATACGACTCACTATAGGATAGGCGGCGCATGAGAGAAGCCCAGACCAATTACCTACCCAA
AATGGAGAAAGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGGA
GCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTCACTGATAATGACCATGCTAATGCCAGA
GCGTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTGGACCCATCCGACACGATCCT
TGACATTGGAAGTGCGCCCGCCCGCAGAATGTATTCTAAGCACAAGTATCATTGTATCTGTCC
GATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAAACT
GTAAGGAAATAACTGATAAGGAATTGGACAAGAAAATGAAGGAGCTCGCCGCCGTCATGAGC
GACCCTGACCTGGAAACTGAGACTATGTGCCTCCACGACGACGAGTCGTGTCGCTACGAAG
GGCAAGTCGCTGTTTACCAGGATGTATACGCGGTTGACGGACCGACAAGTCTCTATCACCAA
GCCAATAAGGGAGTTAGAGTCGCCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAG
AACTTGGCTGGAGCATATCCATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGC
TCGTAACATAGGCCTATGCAGCTCTGACGTTATGGAGCGGTCACGTAGAGGGATGTCCATTC
TTAGAAAGAAGTATTTGAAACCATCCAACAATGTTCTATTCTCTGTTGGCTCGACCATCTACCA

CGAGAAGAGGGACTTACTGAGGAGCTGGCACCTGCCGTCTGTATTTCACTTACGTGGCAAGC
AAAATTACACATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCGTTAAAAGAATAG
CTATCAGTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGGG
ATTCTTGTGCTGCAAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCA
CGTATGTGCCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCG
GACGACGCGCAAAAACTGCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCC
AGAGAAACACCAATACCATGAAAAATTACCTTTTGCCCGTAGTGGCCCAGGCATTTGCTAGGT
GGGCAAAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAGATAGACAG
TTAGTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG
GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTCGTGCTGCCCAGGATAGG
CAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCACAAGG
AGCCGTCACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGATGAGGC
TAAGGAGGTGCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTT
GAGGAGCCCACTCTGGAAGCCGATGTCGACTTGATGTTACAAGAGGCTGGGGCCGGCTCAG
TGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGGACAAGATCGGCTCT
TACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCACCCTCTC
GCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATA
CCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAGGACTTTCAAGCTCTGA
GTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTCGTAAACAGGTACCTGCACCATATT
GCCACACATGGAGGAGCGCTGAACACTGATGAAGAATATTACAAAACTGTCAAGCCCAGCGA
GCACGACGGCGAATACCTGTACGACATCGACAGGAAACAGTGCGTCAAGAAAGAACTAGTCA
CTGGGCTAGGGCTCACAGGCGAGCTGGTGGATCCTCCCTTCCATGAATTCGCCTACGAGAG
TCTGAGAACACGACCAGCCGCTCCTTACCAAGTACCAACCATAGGGGTGTATGGCGTGCCA
GGATCAGGCAAGTCTGGCATCATTAAAAGCGCAGTCACCAAAAAAGATCTAGTGGTGAGCGC
CAAGAAAGAAAACTGTGCAGAAATTATAAGGGACGTCAAGAAAATGAAAGGGCTGGACGTCA
ATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCCGTAGAGACCCTGTAT
ATTGACGAAGCTTTTGCTTGTCATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCT
AAAAAGGCAGTGCTCTGCGGGGATCCCAAACAGTGCGGTTTTTTTAACATGATGTGCCTGAA
AGTGCATTTTAACCACGAGATTTGCACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCAC
TAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTTACGACAAAAAAATGAGAACGACGAATCC
GAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAAACCTAAGCAGGACGATCTCA
TTCTCACTTGTTTCAGAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACGAAATAA
TGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCGGTACAAGGT
GAATGAAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCCGCACGG

AGGACCGCATCGTGTGGAAAACACTAGCCGGCGACCCATGGATAAAAACACTGACTGCCAA
GTACCCTGGGAATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGA
GGCACATCTTGGAGAGACCGGACCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTG
GGCCAAGGCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGACATGACCACTGAACAATGG
AACACTGTGGATTATTTTGAAACGGACAAAGCTCACTCAGCAGAGATAGTATTGAACCAACTA
TGCGTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATTTTCTGCACCCACTGTTCCGTTA
TCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACGGGCTGAATAAAGAA
GTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACTGCCTCGGGCAGTTGCCACTGGAAGAG
TCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAA
ACAGAAGACTGCCTCATGCTTTAGTCCTCCACCATAATGAACACCCACAGAGTGACTTTTCTT
CATTCGTCAGCAAATTGAAGGGCAGAACTGTCCTGGTGGTCGGGGAAAAGTTGTCCGTCCCA
GGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTACCTTCAGAGCTCGGCTGGATTT
AGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAATGTGAGGACCCCATATAA
ATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAAGAAAGC
TTGTCTGCATCTGAATCCCGGCGGAACCTGTGTCAGCATAGGTTATGGTTACGCTGACAGGG
CCAGCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTCAAGTTTTCCCGGGTATGCAAACCG
AAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTATTCATTGGGTACGATCGCAAGGCCCGT
ACGCACAATCCTTACAAGCTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCAC
GAAGCCGGATGTGCACCCTCATATCATGTGGTGCGAGGGGATATTGCCACGGCCACCGAAG
GAGTGATTATAAATGCTGCTAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCT
GTATAAGAAATTCCCGGAAAGCTTCGATTTACAGCCGATCGAAGTAGGAAAAGCGCGACTGG
TCAAAGGTGCAGCTAAACATATCATTCATGCCGTAGGACCAAACTTCAACAAAGTTTCGGAGG
TTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAAC
AATTACAAGTCAGTAGCGATTCCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGA
CTAACCCAATCATTGAACCATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATA
TACTGCAGGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAG
TGGAGGAGATATGCATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGG
GTGCATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTACAGCACAAGCGATGGCAAAACTTT
CTCATATTTGGAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAGCAGAAATTAATGCCA
TGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAAAGCAT
GAGCAGTATTAGGTCGAAATGCCCCGTCGAAGAGTCGGAAGCCTCCACACCACCTAGCACG
CTGCCTTGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACG
TCCAGAACAAATTACTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCA
GAAGATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTCATCCAAG

GAAGTATCTCGTGGAAACACCACCGGTAGACGAGACTCCGGAGCCATCGGCAGAGAACCAA
TCCACAGAGGGGACACCTGAACAACCACCACTTATAACCGAGGATGAGACCAGGACTAGAAC
GCCTGAGCCGATCATCATCGAAGAGGAAGAAGAGGATAGCATAAGTTTGCTGTCAGATGGCC
CGACCCACCAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCCGCCCTCTGTATCTAGCTC
ATCCTGGTCCATTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCT
GGAGGGAGCTAGCGTGACCAGCGGGGCAACGTCAGCCGAGACTAACTCTTACTTCGCAAAG
AGTATGGAGTTTCTGGCGCGACCGGTGCCTGCGCCTCGAACAGTATTCAGGAACCCTCCAC
ATCCCGCTCCGCGCACAAGAACACCGTCACTTGCACCCAGCAGGGCCTGCTCGAGAACCAG
CCTAGTTTCCACCCCGCCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTT
ACCCCGTCACGCACTCCTAGCAGGTCGGTCTCGAGAACCAGCCTGGTCTCCAACCCGCCAG
GCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCGTAGCACAACAACAATGACGG
TTTGATGCGGGTGCATACATCTTTTCCTCCGACACCGGTCAAGGGCATTTACAACAAAAATCA
GTAAGGCAAACGGTGCTATCCGAAGTGGTGTTGGAGAGGACCGAATTGGAGATTTCGTATGC
CCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGAAATTACAGTTAAATCCCACAC
CTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGA
CGTATTCTGCAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAAC
CCTGCATCCTGTTCCTTTGTATTCATCTAGTGTGAACCGTGCCTTTTCAAGCCCCAAGGTCGC
AGTGGAAGCCTGTAACGCCATGTTGAAAGAGAACTTTCCGACTGTGGCTTCTTACTGTATTAT
TCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTCATGCTGCTTAGACACTGCCA
GTTTTTGCCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATAC
GATCGGCAGTGCCTTCAGCGATCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAA
AAGAAATTGCAATGTCACGCAAATGAGAGAATTGCCCGTATTGGATTCGGCGGCCTTTAATGT
GGAATGCTTCAAGAAATATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCAT
CAGGCTTACTGAAGAAAACGTGGTAAATTACATTACCAAATTAAAAGGACCAAAAGCTGCTGC
TCTTTTTGCGAAGACACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAAT
GGACTTAAAGAGAGACGTGAAAGTGACTCCAGGAACAAAACATACTGAAGAACGGCCCAAGG
TACAGGTGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGAATCCACCGAGA
GCTGGTTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTCATACACTGTTTGATATGTCGGC
TGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAAACTG
ACATCGCGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGATTCTG
GAAGACTTAGGTGTGGACGCAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTC
ATCAATACATTTGCCCACTAAAACTAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTC
CTCACACTGTTTGTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTTGAGAGAACGG
CTAACCGGATCACCATGTGCAGCATTCATTGGAGATGACAATATCGTGAAAGGAGTCAAATC

GGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAGTCAAGATTATAGATG CTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCGTGACC GGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAGGCTGTTTAAGCTTGGCAAACCTCTGG CAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTG GAACCGAGTGGGTATTCTTTCAGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAAACCGTAG GAACTTCCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTCAGCTACC TGAGAGGGGCCCCTATAACTCTCTACGGCTAACCTGAATGGACTACGACTATCACGCCCAAA CATTTACAGCCGCGGTGTCAAAAACCGCGTGGACGTGGTTAACATCCCTGCTGGGAGGATCA GCCGTAATTATTATAATTGGCTTGGTGCTGGCTACTATTGTGGCCATGTACGTGCTGACCAAC CAGAAACATAATTGAATACAGCAGCAATTGGCAAGCTGCTTACATAGAACTCGCGGCGATTG GCATGCCGCCTTAAAATTTTTATTTTATTTTTCTTTTCTTTTCCGAATCGGATTTTGTTTTTAATA TTTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAAAtacgtagtttaaac
VEE-UbAAY (SEQ ID NO:14); vector de suministro del VEE con los epítopos tumorales de ratón del MHC clase I SIINFEKL y AH1-A5 insertados
ATGggcggcgcatgagagaagcccagaccaattacctacccaaaatggagaaagttcacgttgacatcgaggaagacagcccatt cctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttcg catctggcttcaaaactgatcgaaacggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtattct aagcacaagtatcattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactgt aaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgagcgaccctgacctggaaactgagactat gtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccgacaagtctcta tcaccaagccaataagggagttagagtcgcctactggataggctttgacaccaccccttttatgtttaagaacttggctggagcatatccatc atactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtagagg gatgtccattcttagaaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggacttact gaggagctggcacctgccgtctgtatttcacttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggtacg tcgttaaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgcaa agtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattgtgtgaccaaatgactggcatactg gcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgcacccagagaaac accaataccatgaaaaattaccttttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatga aaggccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgcccg gatacccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctgag aacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgcag ccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccacctttggcagctgatgttgaggagcccactctggaa gccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataaaggttaccagctacgctggcga

ggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtcat
agtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaataccc
gtccaggactttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacac
atggaggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcgac
aggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacaggcgagctggtggatcctcccttccatgaattcgcctacga
gagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtgccaggatcaggcaagtctggcatcatta
aaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaa
agggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagctttt
gcttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggttttttt
aacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgact
tcggtcgtctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtacc
aaacctaagcaggacgatctcattctcacttgtttcagagggtgggtgaagcagttgcaaatagattacaaaggcaacgaaataatgac
ggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgtacgcacccacctcaga
acatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactgactgc
caagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccgga
ccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatgacca
ctgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttctttgg
actcgatctggactccggtctattttctgcacccactgttccgttatccattaggaataatcactgggataactccccgtcgcctaacatgtacg
ggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttgccactggaagagtctatgacatgaac
actggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccaccataatgaac
acccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaaat
ggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataatat
ttgttaatgtgaggaccccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagcttgt
ctgcatctgaatcccggcggaacctgtgtcagcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcgg
cagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgt
acgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggatgtgcaccctcatatcatgt
ggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcggaggggtgtgcg
gagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaaca
tatcattcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaag
attgtcaacgataacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcatt
gaaccatttgctgacagctttagacaccactgatgcagatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaag
cagtggctaggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaacctgatgcagagctggtgagggtgc
atccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttggaagggaccaagtttcaccagg
cggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaaag

catgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgccat
gactccagaaagagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcact
ggtgtgcagaagatccaatgctcccagcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacacc
accggtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccgaggat
gagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttgctgtcagatggcccgaccca
ccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcatccgactttgatgtgga
cagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtat
ggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttg
cacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgc
ttaccccgtcacgcactcctagcaggtcggtctcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagag
gagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcctccgacaccggtcaagggcatttacaaca
aaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgtatgccccgcgcctcgaccaagaa
aaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtggagaacatg
aaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatc
ctgttcctttgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttcc
gactgtggcttcttactgtattattccagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccct
gcaaagctgcgcagctttccaaagaaacactcctatttggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctcca
gaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattggattcggcggcctttaatgtgga
atgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaacgtggtaaattac
attaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggtttgt
aatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccg
atccgctagcaacagcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgttt
gatatgtcggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaa
agtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtggacgcagagctgttgacgctgattgaggcggctt
tcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctggaatgttcctcacactgtttgtgaa
cacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcattggagatgacaatat
cgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgg
gcgagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctg
tttaagcttggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaacc
gagtgggtattctttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactcta
gctagcagtgttaaatcattcagctacctgagaggggcccctataactctctacggctaacctgaatggactacgactctagaatagtcttta
attaaagtccgccatatgaggccaccatgCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTA
GAGGTGGAGCCCAGTGACACCATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCA
TCCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCT

CTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCTGAGAGGTGGCG
CTGCTTACAGTATAATCAACTTTGAAAAACTGGCTGCTTACGGCATCCTGGGCTTTGTGTTTA
CACTGGCTGCCTACCTGCTGTTTGGCTATCCTGTGTACGTGGCCGCTTATGGACTGTGTACC
CTGGTGGCCATGCTGGCTGCTTACAATCTGGTGCCTATGGTGGCCACAGTGGCCGCCTATT
GTCTTGGCGGACTGCTGACAATGGTGGCAGCCTACAgcccgagctatgcgtatcatcagtttGCAGCCTA
CGGCCCAGGACCAGGCgCTAAATTTGTGGCTGCCTGGACACTGAAAGCCGCCGCTGGACCA
GGTCCTGGACAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTCGGCCCAG
GACCAGGCTATCCCTACGATGTGCCTGATTACGCCTGATagTGATGATTCGAACGGCCGtatca
cgcccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagccgtaattattataattggc
ttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgcttacatagaa
ctcgcggcgattggcatgccgccttaaaatttttattttatttttttctttttctttttccgaatcggattttgtttttaatatttcAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
A

VEE-Luciferasa (SEQ ID NO:15); vector de suministro del VEE con el gen de luciferasa insertado en 7545

ATGggcggcgcatgagagaagcccagaccaattacctacccaaaATGGagaaagttcacgttgacatcgaggaagacagccca
ttcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcactgataatgaccatgctaatgccagagcgttttc
gcatctggcttcaaaactgatcgaaacggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcccgcagaatgtatt
ctaagcacaagtatcattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaact
gtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgccgtcatgagcgaccctgacctggaaactgagact
atgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccaggatgtatacgcggttgacggaccgacaagtctc
tatcaccaagccaataagggagttagagtcgcctactggataggctttgacaccacccctttatgtttaagaacttggctggagcatatcca
tcatactctaccaactgggccgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgtaga
gggatgtccattcttagaaagaagtatttgaaaccatccaacaatgttctattctctgttggctcgaccatctaccacgagaagagggactta
ctgaggagctggcacctgccgtctgtatttcacttacgtggcaagcaaaattacacatgtcggtgtgagactatagttagttgcgacgggta
cgtcgttaaaagaatagctatcagtccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgca
aagtgacagacacattgaacggggagagggtctcttttcccgtgtgcacgtatgtgccagctacattgtgtgaccaaatgactggcatact
ggcaacagatgtcagtgcggacgacgcgcaaaaactgctggttgggctcaaccagcgtatagtcgtcaacggtcgcacccagagaaa
caccaataccatgaaaaattacctttgcccgtagtggcccaggcatttgctaggtgggcaaaggaatataaggaagatcaagaagatg
aaaggccactaggactacgagatagacagttagtcatggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgccc
ggatacccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgcccaggataggcagtaacacattggagatcgggctga
gaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcattaccgccgaggacgtacaagaagctaagtgcgca

gccgatgaggctaaggaggtgcgtgaagccgaggagttgcgcgcagctctaccacctttggcagctgatgttgaggagcccactctgga
agccgatgtcgacttgatgttacaagaggctggggccggctcagtggagacacctcgtggcttgataaaggttaccagctacgctggcg
aggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaattatcttgcatccaccctctcgctgaacaagtc
atagtgataacacactctggccgaaaagggcgttatgccgtggaaccataccatggtaaagtagtggtgccagagggacatgcaatac
ccgtccaggactttcaagctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccac
acatggaggagcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgagcacgacggcgaatacctgtacgacatcg
acaggaaacagtgcgtcaagaaagaactagtcactgggctagggctcacaggcgagctggtggatcctcccttccatgaattcgcctac
gagagtctgagaacacgaccagccgctccttaccaagtaccaaccataggggtgtatggcgtgccaggatcaggcaagtctggcatca
ttaaaagcgcagtcaccaaaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatg
aaagggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaagc
ttttgcttgtcatgcaggtactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcggggatcccaaacagtgcggttt
ttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtg
acttcggtcgtctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagt
accaaacctaagcaggacgatctcattctcacttgtttcagagggtgggtgaagcagttgcaaatagattacaaaggcaacgaaataat
gacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcggtacaaggtgaatgaaaatcctctgtacgcacccacctc
agaacatgtgaacgtcctactgacccgcacggaggaccgcatcgtgtggaaaacactagccggcgacccatggataaaaacactga
ctgccaagtaccctgggaatttcactgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagacc
ggaccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctggcatagacatga
ccactgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtattgaaccaactatgcgtgaggttcttt
ggactcgatctggactccggtctattttctgcacccactgttccgttatccattaggaataatcactgggataactccccgtcgcctaacatgta
cgggctgaataaagaagtggtccgtcagctctctcgcaggtacccacaactgcctcgggcagttgccactggaagagtctatgacatga
acactggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgctttagtcctccaccataatga
acacccacagagtgacttttcttcattcgtcagcaaattgaagggcagaactgtcctggtggtcggggaaaagttgtccgtcccaggcaaa
atggttgactggttgtcagaccggcctgaggctaccttcagagctcggctggatttaggcatcccaggtgatgtgcccaaatatgacataat
atttgttaatgtgaggaccccatataaataccatcactatcagcagtgtgaagaccatgccattaagcttagcatgttgaccaagaaagctt
gtctgcatctgaatcccggcggaacctgtgtcagcataggttatggttacgctgacagggccagcgaaagcatcattggtgctatagcgcg
gcagttcaagttttcccgggtatgcaaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggccc
gtacgcacaatccttacaagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggatgtgcaccctcatatcat
gtggtgcgaggggatattgccacggccaccgaaggagtgattataaatgctgctaacagcaaaggacaacctggcggaggggtgtgc
ggagcgctgtataagaaattcccggaaagcttcgatttacagccgatcgaagtaggaaaagcgcgactggtcaaaggtgcagctaaac
atatcattcatgccgtaggaccaaacttcaacaaagtttcggaggttgaaggtgacaaacagttggcagaggcttatgagtccatcgctaa
gattgtcaacgataacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatcat
tgaaccatttgctgacagctttagacaccactgatgcagatgtagccatatactgcagggacaagaaatgggaaatgactctcaaggaa
gcagtggctaggagagaagcagtggaggagatatgcatatccgacgactcttcagtgacagaacctgatgcagagctggtgagggtg

catccgaagagttctttggctggaaggaagggctacagcacaagcgatggcaaaactttctcatatttggaagggaccaagtttcaccag
gcggccaaggatatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaaa
gcatgagcagtattaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcacgctgccttgcttgtgcatccatgcca
tgactccagaaagagtacagcgcctaaaagcctcacgtccagaacaaattactgtgtgctcatcctttccattgccgaagtatagaatcac
tggtgtgcagaagatccaatgctcccagcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacac
caccggtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccgagga
tgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttgctgtcagatggcccgaccca
ccaggtgctgcaagtcgaggcagacattcacgggccgccctctgtatctagctcatcctggtccattcctcatgcatccgactttgatgtgga
cagtttatccatacttgacaccctggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtat
ggagtttctggcgcgaccggtgcctgcgcctcgaacagtattcaggaaccctccacatcccgctccgcgcacaagaacaccgtcacttg
cacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcgtgaatagggtgatcactagagaggagctcgaggcgc
ttaccccgtcacgcactcctagcaggtcggtctcgagaaccagcctggtctccaacccgccaggcgtaaatagggtgattacaagagag
gagtttgaggcgttcgtagcacaacaacaatgacggtttgatgcgggtgcatacatcttttcctccgacaccggtcaagggcatttacaaca
aaaatcagtaaggcaaacggtgctatccgaagtggtgttggagaggaccgaattggagatttcgtatgccccgcgcctcgaccaagaa
aaagaagaattactacgcaagaaattacagttaaatcccacacctgctaacagaagcagataccagtccaggaaggtggagaacatg
aaagccataacagctagacgtattctgcaaggcctagggcattatttgaaggcagaaggaaaagtggagtgctaccgaaccctgcatc
ctgttcctttgtattcatctagtgtgaaccgtgccttttcaagccccaaggtcgcagtggaagcctgtaacgccatgttgaaagagaactttcc
gactgtggcttcttactgtattattccagagtacgatgcctatttggacatggttgacggagcttcatgctgcttagacactgccagtttttgccct
gcaaagctgcgcagctttccaaagaaacactcctatttggaacccacaatacgatcggcagtgccttcagcgatccagaacacgctcca
gaacgtcctggcagctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattggattcggcggcctttaatgtgga
atgcttcaagaaatatgcgtgtaataatgaatattgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaacgtggtaaattac
attaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaatatgttgcaggacataccaatggacaggtttgt
aatggacttaaagagagacgtgaaagtgactccaggaacaaaacatactgaagaacggcccaaggtacaggtgatccaggctgccg
atccgctagcaacagcgtatctgtgcggaatccaccgagagctggttaggagattaaatgcggtcctgcttccgaacattcatacactgttt
gatatgtcggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactgacatcgcgtcgtttgataaa
agtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtggacgcagagctgttgacgctgattgaggcggctt
tcggcgaaatttcatcaatacatttgcccactaaaactaaatttaaattcggagccatgatgaaatctggaatgttcctcacactgtttgtgaa
cacagtcattaacattgtaatcgcaagcagagtgttgagagaacggctaaccggatcaccatgtgcagcattcattggagatgacaatat
cgtgaaaggagtcaaatcggacaaattaatggcagacaggtgcgccacctggttgaatatggaagtcaagattatagatgctgtggtgg
gcgagaaagcgccttatttctgtggagggtttattttgtgtgactccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctg
tttaagcttggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaacc
gagtgggtattctttcagagctgtgcaaggcagtagaatcaaggtatgaaaccgtaggaacttccatcatagttatggccatgactactcta
gctagcagtgttaaatcattcagctacctgagaggggcccctataactctctacggcTAAcctgaatggactacgactctagaatagtcttt
aattaaagtccgccatatgagatggaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaagacgggaccgcc

ggcgagcagctgcacaaagccatgaagcgctacgccctggtgcccggcaccatcgcctttaccgacgcacatatcgaggtggacatta
cctacgccgagtacttcgagatgagcgttcggctggcagaagctatgaagcgctatgggctgaatacaaaccatcggatcgtggtgtgc
agcgagaatagcttgcagttcttcatgcccgtgttgggtgccctgttcatcggtgtggctgtggccccagctaacgacatctacaacgagcg
cgagctgctgaacagcatgggcatcagccagcccaccgtcgtattcgtgagcaagaaagggctgcaaaagatcctcaacgtgcaaaa
gaagctaccgatcatacaaaagatcatcatcatggatagcaagaccgactaccagggcttccaaagcatgtacaccttcgtgacttccc
atttgccacccggcttcaacgagtacgacttcgtgcccgagagcttcgaccgggacaaaaccatcgccctgatcatgaacagtagtggc
agtaccggattgcccaagggcgtagccctaccgcaccgcaccgcttgtgtccgattcagtcatgcccgcgaccccatcttcggcaacca
gatcatccccgacaccgctatcctcagcgtggtgccatttcaccacggcttcggcatgttcaccacgctgggctacttgatctgcggctttcg
ggtcgtgctcatgtaccgcttcgaggaggagctattcttgcgcagcttgcaagactataagattcaatctgccctgctggtgcccacactattt
agcttcttcgctaagagcactctcatcgacaagtacgacctaagcaacttgcacgagatcgccagcggcggggcgccgctcagcaagg
aggtaggtgaggccgtggccaaacgcttccacctaccaggcatccgccagggctacggcctgacagaaacaaccagcgccattctga
tcacccccgaaggggacgacaagcctggcgcagtaggcaaggtggtgcccttcttcgaggctaaggtggtggacttggacaccggtaa
gacactgggtgtgaaccagcgcggcgagctgtgcgtccgtggccccatgatcatgagcggctacgttaacaaccccgaggctacaaac
gctctcatcgacaaggacggctggctgcacagcggcgacatcgcctactgggacgaggacgagcacttcttcatcgtggaccggctga
agagcctgatcaaatacaagggctaccaggtagccccagccgaactggagagcatcctgctgcaacaccccaacatcttcgacgccg
gggtcgccggcctgcccgacgacgatgccggcgagctgcccgccgcagtcgtcgtgctggaacacggtaaaaccatgaccgagaag
gagatcgtggactatgtggccagccaggttacaaccgccaagaagctgcgcggtggtgttgtgttcgtggacgaggtgcctaaaggact
gaccggcaagttggacgcccgcaagatccgcgagattctcattaaggccaagaagggcggcaagatcgccgtgtaaTTCGAAC
GGCCGtatcacgcccaaacatttacagccgcggtgtcaaaaaccgcgtggacgtggttaacatccctgctgggaggatcagccgta
attattataattggcttggtgctggctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagc
tgcttacatagaactcgcggcgattggcatgccgccttaaaatttttattttattttttctttttctttttccgaatcggattttgtttttaatatttcAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAA

ubiquitina (SEQ ID NO:38)

UbG76 0-228

ATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACAC
CATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTG
ATCTTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTC
AACCCTGCACCTGGTCCTTCGCCTGAGAGGTGGC

Ubiquitina A76 (SEQ ID NO:39)

UbA76 0-228

ATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACAC

CATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTG
ATCTTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTC
AACCCTGCACCTGGTCCTTCGCCTGAGAGGTGCC

Péptido señal de HLA-A2 (MHC clase I) (SEQ ID NO:40)

>MHC PepSeñal 0-78

atggccgtcatggcgccccgaaccctcgtcctgctactctcgggggctctggccctgacccagacctgggcgggctct

Dominio transmembrana de HLA-A2 (MHC clase I) (SEQ ID NO:41)

>Dominio TM de HLA A2 0-201

CCGtcttcccagcccaccatccCCATCGTGGGCAtcattgctggcctggttctctttggagctgtgatcactggagctgtggtcgctgctgt
gatgtggaggaggaagagctcagatagaaaaggagggagctactctcaggctgcaagcagtgacagtgcccagggctctgatgtgtctctca
cagcttgtaaagtgtga

Sec. líder de IgK (SEQ ID NO:42)

>Sec. líder de IgK 0-60

atggagaccgatacactgctgctgtgggtgctgctcctgtgggtgccaggaagcacaggc

DC-Lamp humana (SEQ ID NO:43)

>DCLAMPHumana 0-3178

ggcaccgattcggggcctgcccggacttcgccgcacgctgcagaacctcgcccagcgcccaccatgccccggcagctcagcgcggcggccg
cgctcttcgcgtccctggccgtaattttgcacgatggcagtcaaatgagagcaaaagcatttccagaaaccagagattattctcaacctactgcag
cagcaacagtacaggacataaaaaaaacctgtccagcaaccagctaagcaagcacctcaccaaactttagcagcaagattcatggatggtcat
atcacctttcaaacagcggccacagtaaaaattccaacaactaccccagcaactacaaaaaacactgcaaccaccagcccaattacctacac
cctggtcacaacccaggccacacccaacaactcacacacagctcctccagttactgaagttacagtcggccctagcttagccccttattcactgc
cacccaccatcaccccaccagctcatacagctggaaccagttcatcaaccgtcagccacacaactgggaacaccactcaacccagtaacca
gaccacccttccagcaactttatcgatagcactgcacaaaagcacaaccggtcagaagcctgatcaacccacccatgccccaggaacaacg
gcagctgcccacaataccacccgcacagctgcacctgcctccacggttcctgggcccacccttgcacctcagccatcgtcagtcaagactgga
atttatcaggttctaaacggaagcagactctgtataaaagcagagatggggatacagctgattgttcaagacaaggagtcggttttttcacctcgga
gatacttcaacatcgaccccaacgcaacgcaagcctctgggaactgtggcacccgaaaatccaaccttctgttgaattttcagggcggatttgtga
atctcacatttaccaaggatgaagaatcatattatatcagtgaagtgggagcctatttgaccgtctcagatccagagacagtttaccaaggaatca
aacatgcggtggtgatgttccagacagcagtcgggcattccttcaagtgcgtgagtgaacagagcctccagttgtcagcccacctgcaggtgaa
aacaaccgatgtccaacttcaagcctttgattttgaagatgaccactttggaaatgtggatgagtgctcgtctgactacacaattgtgcttcctgtgatt
ggggccatcgtggttggtctctgccttatgggtatgggtgtctataaaatccgcctaaggtgtcaatcatctggataccagagaatctaattgttgccc

gggggggaatgaaaataatggaatttagagaactctttcatcccttccaggatggatgttgggaaattccctcagagtgtgggtccttcaaacaatgt
aaaccaccatcttctattcaaatgaagtgagtcatgtgtgatttaagttcaggcagcacatcaatttctaaatacttttttgtttatttttatgaaagatatagt
gagctgtttatttttctagtttcctttagaatattttagccactcaaagtcaacatttgagatatgttgaattaacataatatatgtaaagtagaataagcctt
caaattataaaccaagggtcaattgtaactaatactactgtgtgtgcattgaagattttatttttaccccttgatcttaacaaagcctttgctttgttatcaaat
ggactttcagtgcttttactatctgtgttttatggtttcatgtaacatacatattcctggtgtagcacttaactccttttccacttttaaatttgttttttgttttttgaga
cggagtttcactcttgtcacccaggctggagtacagtggcacgatctcggcttatggcaacctccgcctcccgggttcaagtgattctcctgcttcag
cttcccgagtagctgggattacaggcacacactaccacgcctggctaatttttgtatttttattatagacgggtttcaccatgttggccagactggtcttg
aactcttgacctcaggtgatccacccacctcagcctcccaaagtgctgggattacaggcatgagccattgcgcccggccttaaatgtttttttttaatc
atcaaaaagaacaacatatctcaggttgtctaagtgtttttatgtaaaaccaacaaaaagaacaaatcagcttatatttttatcttgatgactcctgct
ccagaattgctagactaagaattaggtggctacagatggtagaactaaacaataagcaagagacaataataatggcccttaattattaacaaag
tgccagagtctaggctaagcactttatctatatctcatttcattctcacaacttataagtgaatgagtaaactgagacttaagggaactgaatcactta
aatgtcacctggctaactgatggcagagccagagcttgaattcatgttggtctgacatcaaggtctttggtcttctccctacaccaagttacctacaa
gaacaatgacaccacactctgcctgaaggctcacacctcataccagcatacgctcaccttacagggaaatgggtttatccaggatcatgagaca
ttagggtagatgaaaggagagctttgcagataacaaaatagcctatccttaataaatcctccactctctggaaggagactgaggggctttgtaaaa
cattagtcagttgctcatttttatgggattgcttagctgggctgtaaagatgaaggcatcaaataaactcaaagtatttttaaatttttttgataatagaga
aacttcgctaaccaactgttctttcttgagtgtatagccccatcttgtggtaacttgctgcttctgcacttcatatccatatttcctattgttcactttattctgta
gagcagcctgccaagaattttatttctgctgttttttttgctgctaaagaaaggaactaagtcaggatgttaacagaaaagtccacataaccctagaa
ttcttagtcaaggaataattcaagtcagcctagagaccatgttgactttcctcatgtgtttccttatgactcagtaagttggcaaggtcctgactttagtctt
aataaaacattgaattgtagtaaaggtttttgcaataaaaacttactttgg

LAMP1 de ratón (SEQ ID NO:44)

>LampRatón1 0-1858

attccggaggtgaaaaacaatggcacaacgtgtataatggccagcttctctgcctcctttctgaccacctacgagactgcgaatggttctcagatcg
tgaacatttccctgccagcctctgcagaagtactgaaaaatggcagttcttgtggtaaagaaaatgtttctgaccccagcctcacaattacttttgga
agaggatatttactgacactcaacttcacaaaaaatacaacacgttacagtgtccagcatatgtattttacatataacttgtcagatacagaacatttt
cccaatgccatcagcaaagagatctacaccatggattccacaactgacatcaaggcagacatcaacaaagcataccggtgtgtcagtgatatc
cgggtctacatgaagaatgtgaccgttgtgctccgggatgccactatccaggcctacctgtcgagtggcaacttcagcaaggaagagacacact
gcacacaggatggaccttccccaaccactgggccacccagcccctcaccaccacttgtgcccacaaaccccactgtatccaagtacaatgtta
ctggtaacaacggaacctgcctgctggcctctatggcactgcaactgaatatcacctacctgaaaaaggacaacaagacggtgaccagagcgt
tcaacatcagcccaaatgacacatctagtgggagttgcggtatcaacttggtgaccctgaaagtggagaacaagaacagagccctggaattgc
agtttgggatgaatgccagctctagcctgtttttcttgcaaggagtgcgcttgaatatgactcttcctgatgccctagtgcccacattcagcatctccaa
ccattcactgaaagctcttcaggccactgtgggaaactcatacaagtgcaacactgaggaacacatctttgtcagcaagatgctctccctcaatgt
cttcagtgtgcaggtccaggctttcaaggtggacagtgacaggtttgggtctgtggaagagtgtgttcaggatggtaacaacatgttgatccccattg
ctgtgggcggtgccctggcagggctgatcctcatcgtcctcattgcctacctcattggcaggaagaggagtcacgccggctatcagaccatctag

cctggtgggcaggtgcaccagagatgcacaggggcctgttctcacatccccaagcttagataggtgtggaagggaggcacactttctggcaaa
ctgttttaaaatctgctttatcaaatgtgaagttcatcttgcaacatttactatgcacaaaggaataactattgaaatgacggtgttaattttgctaactgg
gttaaatattgatgagaaggctccactgatttgacttttaagacttggtgtttggttcttcattcttttactcagatttaagcctatcaaagggatactctggt
ccagaccttggcctggcaagggtggctgatggttaggctgcacacacttaagaagcaacgggagcagggaaggcttgcacacaggcacgca
cagggtcaacctctggacacttggcttgggctacctggccttgggggggctgaactctggcatctggctgggtacacacccccccaatttctgtgct
ctgccacccgtgagctgccactttcctaaatagaaaatggcattatttttatttacttttttgtaaagtgatttccagtcttgtgttggcgttcagggtggccc
tgtctctgcactgtgtacaataatagattcacactgctgacgtgtcttgcagcgtaggtgggttgtacactgggcatcagctcacgtaatgcattgcct
gtaacgatgctaataaaaa

ADNc de Lamp1 humana (SEQ ID NO:45)

>Lamp1 humana 0-2339

ggcccaaccgccgcccgcgccccgctctccgcaccgtacccggccgcctcgcgccatggcggcccccggcagcgcccggcgacccctgct
gctgctactgctgttgctgctgctcggcctcatgcattgtgcgtcagcagcaatgtttatggtgaaaaatggcaacgggaccgcgtgcataatggcc
aacttctctgctgccttctcagtgaactacgacaccaagagtggccctaagaacatgacctttgacctgccatcagatgccacagtggtgctcaac
cgcagctcctgtggaaaagagaacacttctgaccccagtctcgtgattgcttttggaagaggacatacactcactctcaatttcacgagaaatgca
acacgttacagcgtccagctcatgagttttgtttataacttgtcagacacacacctttttccccaatgcgagctccaaagaaatcaagactgtggaatc
tataactgacatcagggcagatatagataaaaaatacagatgtgttagtggcacccaggtccacatgaacaacgtgaccgtaacgctccatgat
gccaccatccaggcgtacctttccaacagcagcttcagcaggggagagacacgctgtgaacaagacaggccttccccaaccacagcgcccc
ctgcgccacccagcccctcgccctcacccgtgcccaagagcccctctgtggacaagtacaacgtgagcggcaccaacgggacctgcctgctg
gccagcatggggctgcagctgaacctcacctatgagaggaaggacaacacgacggtgacaaggcttctcaacatcaaccccaacaagacct
cggccagcgggagctgcggcgcccacctggtgactctggagctgcacagcgagggcaccaccgtcctgctcttccagttcgggatgaatgcaa
gttctagccggtttttcctacaaggaatccagttgaatacaattcttcctgacgccagagaccctgcctttaaagctgccaacggctccctgcgagc
gctgcaggccacagtcggcaattcctacaagtgcaacgcggaggagcacgtccgtgtcacgaaggcgttttcagtcaatatattcaaagtgtgg
gtccaggctttcaaggtggaaggtggccagtttggctctgtggaggagtgtctgctggacgagaacagcatgctgatccccatcgctgtgggtggt
gccctggcggggctggtcctcatcgtcctcatcgcctacctcgtcggcaggaagaggagtcacgcaggctaccagactatctagcctggtgcac
gcaggcacagcagctgcaggggcctctgttcctttctctgggcttagggtcctgtcgaaggggaggcacactttctggcaaacgtttctcaaatctg
cttcatccaatgtgaagttcatcttgcagcatttactatgcacaacagagtaactatcgaaatgacggtgttaattttgctaactgggttaaatattttgct
aactggttaaacattaatatttaccaaagtaggattttgagggtgggggtgctctctctgaggggtgggggtgccgctgtctctgaggggtggggg
tgccgctgtctctgaggggtgggggtgccgctctctctgaggggtgggggtgccgctttctctgaggggtgggggtgccgctctctctgagggg
gtgggggtgctgctctctccgaggggtggaatgccgctgtctctgaggggtgggggtgccgctctaaattggctccatatcatttgagtttagggttct
ggtgtttggtttcttcattctttactgcactcagatttaagccttacaaagggaaagcctctggccgtcacacgtaggacgcatgaaggtcactcgtgg
tgaggctgacatgctcacacattacaacagtagagagggaaaatcctaagacagaggaactccagagatgagtgtctggagcgcttcagttca
gctttaaaggccaggacgggccacacgtggctggcggcctcgttccagtggcggcacgtccttgggcgtctctaatgtctgcagctcaagggctg
gcacttttttaaatataaaaatgggtgttatttttatttttttttgtaaagtgatttttggtcttctgttgacattcggggtgatcctgttctgcgctgtgtacaatgt

gagatcggtgcgttctcctgatgttttgccgtggcttggggattgtacacgggaccagctcacgtaatgcattgcctgtaacaatgtaataaaaagcc
tctttcttttaaaaaaaaaaaaaaaaaaaaaaa

Secuencia de ácidos nucleicos de toxoide tetánico (SEQ ID NO:46)
CAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTC

Secuencia de aminoácidos de toxoide tetánico (SEQ ID NO:47)
QYIKANSKFIGITEL

Secuencia de nucleótidos de PADRE (SEQ ID NO:48)
GCTAAATTTGTGGCTGCCTGGACACTGAAAGCCGCCGCT

Secuencia de aminoácidos de PADRE (SEQ ID NO:49)
AKFVAAWTLKAAA

WPRE (SEQ ID NO:50)
>WPRE 0-593
aatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgctgctttaatgcctttgtatca
tgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggttgctgtctctttatgaggagttgtggcccgttgtcaggcaacgtggcgtggt
gtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagctcctttccgggactttcgctttcccccctccctattgccacg
gcggaactcatcgccgcctgccttgcccgctgctggacaggggctcggctgttgggcactgacaattccgtggtgttgtcggggaagctgacgtc
ctttccatggctgctcgcctgtgttgccacctggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccttcccgc
ggcctgctgccggctctgcggcctcttccgcgtcttcgccttcgccctcagacgagtcggatctccctttgggccgcctccccgcctgt

IRES (SEQ ID NO:51)
>eGFP_IRES_SEAP_Inserto 1746-2335
tctccccccccccctctccctcccccccccctaacgttactggccgaagccgcttggaataaggccggtgtgcgtttgtctatatgttattttccacca
tattgccgtcttttggcaatgtgagggcccggaaacctggccctgtcttcttgacgagcattcctaggggtctttcccctctcgccaaaggaatgcaa
ggtctgttgaatgtcgtgaaggaagcagttcctctggaagcttcttgaagacaaacaacgtctgtagcgacccttgcaggcagcggaacccccc
acctggcgacaggtgcctctgcggccaaaagccacgtgtataagatacacctgcaaaggcggcacaaccccagtgccacgttgtgagttggat
agttgtggaaagagtcaaatggctctcctcaagcgtattcaacaaggggctgaaggatgcccagaaggtaccccattgtatgggatctgatctgg
ggcctcggtgcacatgctttacatgtgtttagtcgaggttaaaaaaacgtctaggccccccgaaccacggggacgtggttttcctttgaaaaacac
gatgataatatg

GFP (SEQ ID NO:52)

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtc
cggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcg
tgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggcta
cgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgc
atcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcat
ggccgacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactacc
agcagaacaccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacg
agaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagtag

SEAP (SEQ ID NO:53)

atgctgctgctgctgctgctgctgggcctgaggctacagctctccctgggcatcatcccagttgaggaggagaacccggacttctggaaccgcga
ggcagccgaggccctgggtgccgccaagaagctgcagcctgcacagacagccgccaagaacctcatcatcttcctgggcgatgggatgggg
gtgtctacggtgacagctgccaggatcctaaaagggcagaagaaggacaaactggggcctgagataccctggccatggaccgcttcccata
tgtggctctgtccaagacatacaatgtagacaaacatgtgccagacagtggagccacagccacggcctacctgtgcggggtcaagggcaactt
ccagaccattggcttgagtgcagccgcccgctttaaccagtgcaacacgacacgcggcaacgaggtcatctccgtgatgaatcgggccaaga
aagcagggaagtcagtgggagtggtaaccaccacacgagtgcagcacgcctcgccagccggcacctacgcccacacggtgaaccgcaact
ggtactcggacgccgacgtgcctgcctcggcccgccaggaggggtgccaggacatcgctacgcagctcatctccaacatggacattgacgtga
tcctaggtggaggccgaaagtacatgtttcgcatgggaaccccagaccctgagtacccagatgactacagccaaggtgggaccaggctggac
gggaagaatctggtgcaggaatggctggcgaagcgccagggtgcccggtatgtgtggaaccgcactgagctcatgcaggcttccctggaccc
gtctgtgacccatctcatgggtctctttgagcctggagacatgaaatacgagatccaccgagactccacactggacccctccctgatggagatga
cagaggctgccctgcgcctgctgagcaggaacccccgcggcttcttcctcttcgtggagggtggtcgcatcgaccatggtcatcatgaaagcag
ggcttaccgggcactgactgagacgatcatgttcgacgacgccattgagagggcgggccagctcaccagcgaggaggacacgctgagcctc
gtcactgccgaccactcccacgtcttctccttcggaggctacccccctgcgagggagctccatcttcgggctggcccctggcaaggcccgggaca
ggaaggcctacacggtcctcctatacggaaacggtccaggctatgtgctcaaggacggcgcccggccggatgttaccgagagcgagagcgg
gagccccgagtatcggcagcagtcagcagtgcccctggacgaagagacccacgcaggcgaggacgtggcggtgttcgcgcgcggcccgc
aggcgcacctggttcacggcgtgcaggagcagaccttcatagcgcacgtcatggccttcgccgcctgcctggagccctacaccgcctgcgacct
ggcgcccccccgccggcaccaccgacgccgcgcacccgggttactctagagtcggggcggccggccgcttcgagcagacatgataa

Luciferasa Firefly (SEQ ID NO:54)

atggaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaagacgggaccgccggcgagcagctgcacaaagccatg
aagcgctacgccctggtgcccggcaccatcgcctttaccgacgcacatatcgaggtggacattacctacgccgagtacttcgagatgagcgttcg
gctggcagaagctatgaagcgctatgggctgaatacaaaccatcggatcgtggtgtgcagcgagaatagcttgcagttcttcatgcccgtgttgg
gtgccctgttcatcggtgtggctgtggccccagctaacgacatctacaacgagcgcgagctgctgaacagcatgggcatcagccagcccaccgt

cgtattcgtgagcaagaaagggctgcaaaagatcctcaacgtgcaaaagaagctaccgatcatacaaaagatcatcatcatggatagcaaga
ccgactaccagggcttccaaagcatgtacaccttcgtgacttcccatttgccacccggcttcaacgagtacgacttcgtgcccgagagcttcgacc
gggacaaaaccatcgccctgatcatgaacagtagtggcagtaccggattgcccaagggcgtagccctaccgcaccgcaccgcttgtgtccgat
tcagtcatgcccgcgaccccatcttcggcaaccagatcatccccgacaccgctatcctcagcgtggtgccatttcaccacggcttcggcatgttca
ccacgctgggctacttgatctgcggctttcgggtcgtgctcatgtaccgcttcgaggaggagctattcttgcgcagcttgcaagactataagattcaa
tctgccctgctggtgcccacactatttagcttcttcgctaagagcactctcatcgacaagtacgacctaagcaacttgcacgagatcgccagcggc
ggggcgccgctcagcaaggaggtaggtgaggccgtggccaaacgcttccacctaccaggcatccgccagggctacggcctgacagaaaca
accagcgccattctgatcacccccgaaggggacgacaagcctggcgcagtaggcaaggtggtgcccttcttcgaggctaaggtggtggacttg
gacaccggtaagacactgggtgtgaaccagcgcggcgagctgtgcgtccgtggccccatgatcatgagcggctacgttaacaaccccgaggc
tacaaacgctctcatcgacaaggacggctggctgcacagcggcgacatcgcctactgggacgaggacgagcacttcttcatcgtggaccggct
gaagagcctgatcaaatacaagggctaccaggtagccccagccgaactggagagcatcctgctgcaacaccccaacatcttcgacgccggg
gtcgccggcctgcccgacgacgatgccggcgagctgcccgccgcagtcgtcgtgctggaacacggtaaaaccatgaccgagaaggagatc
gtggactatgtggccagccaggttacaaccgccaagaagctgcgcggtggtgttgtgttcgtggacgaggtgcctaaaggactgaccggcaagt
tggacgcccgcaagatccgcgagattctcattaaggccaagaagggcggcaagatcgccgtgtaa

FMDV 2A (SEQ ID NO:55)

GTAAAGCAAACACTGAACTTTGACCTTCTCAAGTTGGCTGGAGACGTTGAGTCCAATCCTGGGCCC

Referencias

1. Desrichard, A., Snyder, A. y Chan, T. A. Cancer Neoantigens and Applications for Immunotherapy. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* (2015). doi:10.1158/1078-0432.CCR-14-3175

2. Schumacher, T. N. y Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348,** 69–74 (2015).

3. Gubin, M. M., Artyomov, M. N., Mardis, E. R. y Schreiber, R. D. Tumor neoantigens: building a framework for personalized cancer immunotherapy. *J. Clin. Invest.* **125,** 3413-3421 (2015).

4. Rizvi, N. A. et ál. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348,** 124-128 (2015).

5. Snyder, A. et ál. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371,** 2189-2199 (2014).

6. Carreno, B. M. et ál. Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348,** 803-808 (2015).

7. Tran, E. et ál. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344,** 641-645 (2014).

8. Hacohen, N. y Wu, C. J.-Y. Solciitud de patente estadounidense: 0110293637 - COMPOSITIONS AND METHODS OF IDENTIFYING TUMOR SPECIFIC NEOANTIGENS. (A1). en <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20110293637.PGNR.>

9. Lundegaard, C., Hoof, I., Lund, O. y Nielsen, M. State of the art and challenges in sequence based T-cell epitope prediction. *Immunome Res.* **6 Supl 2,** S3 (2010).

10. Yadav, M. et ál. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515,** 572–576 (2014).

11. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J. y Mann, M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics MCP* **14,** 658–673 (2015).

12. Van Allen, E. M. et ál. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350,** 207-211 (2015).

13. Yoshida, K. y Ogawa, S. Splicing factor mutations and cancer. *Wiley Interdiscip. Rev. RNA* **5,** 445–459 (2014).

14. Red de Investigación del Atlas del Genoma del Cáncer. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543-550 (2014).

15. Rajasagi, M. et ál. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124,** 453–462 (2014).

16. Downing, S. R. et ál. Solicitud de patente de Estados Unidos 0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). en <http://appft1.uspto.gov/netacgi/nph-Parser?Sect1=PTO1&Sect2=HITOFF&d=PG01&p=1&u=/netahtml/PTO/srchnum.html&r=1&f=G&l=50&s1=20120208706.PGNR.>

17. Target Capture for NextGen Sequencing - IDT. en <http://www.idtdna.com/pages/products/nextgen/target-capture>

18. Shukla, S. A. et ál. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33,** 1152-1158 (2015).

19. Cieslik, M. et ál. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25,** 1372-1381 (2015).

20. Bodini, M. et ál. The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* **125,** 600-605 (2015).

21. Saunders, C. T. et ál. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* **28,** 1811–1817 (2012).

22. Cibulskis, K. et ál. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31,** 213-219 (2013).

23. Wilkerson, M. D. et ál. Integrated RNA and DNA sequencing improves mutation detection in low purity tumors. *Nucleic Acids Res.* **42,** e107 (2014).

24. Mose, L. E., Wilkerson, M. D., Hayes, D. N., Perou, C. M. y Parker, J. S. ABRA: improved coding indel detection via assembly-based realignment. *Bioinforma. Oxf. Engl.* **30,** 2813-2815 (2014).

25. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. y Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinforma. Oxf. Engl.* **25,** 2865-2871 (2009).

26. Lam, H. Y. K. et ál. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* **28,** 47-55 (2010).

27. Frampton, G. M. et ál. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat. Biotechnol.* **31,** 1023-1031 (2013).

28. Boegel, S. et ál. HLA typing from RNA-Seq sequence reads. *Genome Med.* **4,** 102 (2012).

29. Liu, C. et ál. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* **41,** e142 (2013).

30. Mayor, N. P. et ál. HLA Typing for the Next Generation. *PloS One* **10,** e0127153 (2015).

31. Roy, C. K., Olson, S., Graveley, B. R., Zamore, P. D. y Moore, M. J. Assessing long-distance RNA

sequence connectivity via RNA-templated DNA-DNA ligation. *eLife* **4,** (2015).

32. Song, L. y Florea, L. CLASS: constrained transcript assembly of RNA-seq reads. *BMC Bioinformatics* **14 Supl 5,** S14 (2013).

33. Maretty, L., Sibbesen, J. A. y Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* **15,** 501 (2014).

34. Pertea, M. et ál. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33,** 290-295 (2015).

35. Roberts, A., Pimentel, H., Trapnell, C. y Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinforma. Oxf. Engl.* (2011). doi:10.1093/bioinformatics/btr355

36. Vitting-Seerup, K., Porse, B. T., Sandelin, A. y Waage, J. spliceR: an R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics* **15,** 81 (2014).

37. Rivas, M. A. et ál. Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348,** 666-669 (2015).

38. Skelly, D. A., Johansson, M., Madeoy, J., Wakefield, J. y Akey, J. M. A powerful and flexible statistical framework for testing hypotheses of allele-specific gene expression from RNA-seq data. *Genome Res.* **21,** 1728-1737 (2011).

39. Anders, S., Pyl, P. T. y Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinforma. Oxf. Engl.* **31,** 166-169 (2015).

40. Furney, S. J. et ál. SF3B1 mutations are associated with alternative splicing in uveal melanoma. *Cancer Discov.* (2013). doi:10.1158/2159-8290.CD-13-0330

41. Zhou, Q. et ál. A chemical genetics approach for the functional assessment of novel cancer genes. *Cancer Res.* (2015). doi:10.1158/0008-5472.CAN-14-2930

42. Maguire, S. L. et ál. SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235,** 571-580 (2015).

43. Carithers, L. J. et ál. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* **13,** 311–319 (2015).

44. Xu, G. et ál. RNA CoMPASS: a dual approach for pathogen and host transcriptome analysis of RNA-seq datasets. *PloS One* **9,** e89445 (2014).

45. Andreatta, M. y Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinforma. Oxf. Engl.* (2015). doi:10.1093/bioinformatics/btv639

46. Jørgensen, K. W., Rasmussen, M., Buus, S. y Nielsen, M. NetMHCstab - predicting stability of peptide-MHC-I complexes; impacts for cytotoxic T lymphocyte epitope discovery. *Immunology* **141,** 18–26 (2014).

47. Larsen, M. V. et ál. An integrative approach to CTL epitope prediction: a combined algorithm integrating MHC class I binding, TAP transport efficiency, and proteasomal cleavage predictions. *Eur. J. Immunol.* **35,** 2295-2303 (2005).

48. Nielsen, M., Lundegaard, C., Lund, O. y Keşmir, C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57,** 33–41 (2005).

49. Boisvert, F.-M. et ál. A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* **11,** M111.011429–M111.011429 (2012).

50. Duan, F. et ál. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211,** 2231-2248 (2014).

51. Janeway's Immunobiology: 9780815345312: Medicine & Health Science Books en Amazon.com. en <http://www.amazon.com/Janeways-Immunobiology-Kenneth-Murphy/dp/0815345313>

52. Calis, J. J. A. et ál. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* **9,** e1003266 (2013).

53. Zhang, J. et ál. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346,** 256–259 (2014)

54. Walter, M. J. et ál. Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* 366, 1090-1098 (2012).

55. Hunt DF, Henderson RA, Shabanowitz J, Sakaguchi K, Michel H, Sevilir N, Cox AL, Appella E, Engelhard VH. Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. Science 1992. 255: 1261-1263.

56. Zarling AL, Polefrone JM, Evans AM, Mikesh LM, Shabanowitz J, Lewis ST, Engelhard VH, Hunt DF. Identification of class I MHC-associated phosphopeptides as targets for cancer immunotherapy. Proc Natl Acad Sci EUA. 3 de oct. de 2006;103(40):14889-94.

57. Bassani-Sternberg M, Pletscher-Frankild S, Jensen LJ, Mann M. Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. Mol Cell Proteomics. Mar. de 2015; 14(3):658-73. doi: 10.1074/mcp.M114.042812.

58. Abelin JG, Trantham PD, Penny SA, Patterson AM, Ward ST, Hildebrand WH, Cobbold M, Bai DL, Shabanowitz J, Hunt DF. Complementary IMAC enrichment methods for HLA-associated phosphopeptide identification by mass spectrometry. Nat Protoc. Sep. de 2015;10(9):1308-18. doi: 10.1038/nprot.2015.086. Epub 6 de ago. de 2015

59. Barnstable CJ, Bodmer WF, Brown G, Galfre G, Milstein C, Williams AF, Ziegler A. Production of monoclonal antibodies to group A erythrocytes, HLA and other human cell surface antigens-new tools for genetic analysis. Cell. May. de 1978;14(1):9-20.

60. Goldman JM, Hibbin J, Kearney L, Orchard K, Th'ng KH. HLA-DR monoclonal antibodies inhibit the proliferation of normal and chronic granulocytic leukaemia myeloid progenitor cells. Br J Haematol. Nov. de 1982;52(3):411-20.

61. Eng JK, Jahan TA, Hoopmann MR. Comet: an open-source MS/MS sequence database search tool. Proteomics. Ene. de 2013;13(1):22-4. doi: 10.1002/pmic.201200439. Publicación electrónica 4 dic 2012.

62. Eng JK, Hoopmann MR, Jahan TA, Egertson JD, Noble WS, MacCoss MJ. A deeper look into Comet--implementation and features. J Am Soc Mass Spectrom. Nov. de 2015;26(11):1865-74. doi: 10.1007/s13361-015-1179-x. Epub 27 de jun. de 2015.

63. Lukas Käll, Jesse Canterbury, Jason Weston, William Stafford Noble y Michael J. MacCoss. Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature Methods 4:923 – 925, noviembre de 2007

64. Lukas Käll, John D. Storey, Michael J. MacCoss y William Stafford Noble. Assigning confidence measures to peptides identified by tandem mass spectrometry. Journal of Proteome Research, 7(1):29-34, enero de 2008

65. Lukas Käll, John D. Storey y William Stafford Noble. Nonparametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry. Bioinformatics, 24(16):i42-i48, agosto de 2008

66. Kinney RM, BJ Johnson, VL Brown, DW Trent. Nucleotide Sequence of the 26 S mRNA of the Virulent Trinidad Donkey Strain of Venezuelan Equine Encephalitis Virus and Deduced Sequence of the Encoded Structural Proteins. Virology 152 (2), 400-413. 30 de jul. de 1986.

67. Jill E Slansky, Frédérique M Rattis, Lisa F Boyd, Tarek Fahmy, Elizabeth M Jaffee, Jonathan P Schneck, David H Margulies, Drew M Pardoll. Enhanced Antigen-Specific Antitumor Immunity with Altered Peptide Ligands that Stabilize the MHC-Peptide-TCR Complex. Immunity, Tomo 13, Número 4, 1.º de octubre de 2000, Páginas 529-538.

68. A Y Huang, P H Gulden, A S Woods, M C Thomas, C D Tong, W Wang, V H Engelhard, G Pasternack, R Cotter, D Hunt, D M Pardoll y E M Jaffee. The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. Proc Natl Acad Sci EUA.; 93(18): 9730–9735, 3 de sep. de 1996.

69. JOHNSON, BARBARA J. B., RICHARD M. KINNEY, CRYSTLE L. KOST Y DENNIS W. TRENT. Molecular Determinants of Alphavirus Neurovirulence: Nucleotide and Deduced Protein Sequence Changes during Attenuation of Venezuelan Equine Encephalitis Virus. J Gen Virol 67:1951-1960, 1986.

70. Aarnoudse, C.A., Krüse, M., Konopitzky, R., Brouwenstijn, N. y Schrier, P.I. (2002). TCR reconstitution in Jurkat reporter cells facilitates the identification of novel tumor antigens by cDNA expression cloning. Int J Cancer 99, 7–13.

71. Alexander, J., Sidney, J., Southwood, S., Ruppert, J., Oseroff, C., Maewal, A., Snoke, K., Serra, H.M., Kubo, R.T. y Sette, A. (1994). Development of high potency universal DR-restricted helper epitopes by modification of high affinity DR-blocking peptides. Immunity 1, 751–761.

72. Banu, N., Chia, A., Ho, Z.Z., Garcia, A.T., Paravasivam, K., Grotenbreg, G.M., Bertoletti, A. y Gehring, A.J. (2014). Building and optimizing a virus-specific T cell receptor library for targeted immunotherapy in viral infections. Scientific Reports 4, 4166.

73. Cornet, S., Miconnet, I., Menez, J., Lemonnier, F. y Kosmatopoulos, K. (2006). Optimal organization of a polypeptide-based candidate cancer vaccine composed of cryptic tumor peptides with enhanced immunogenicity. Vaccine 24, 2102–2109.

74. Depla, E., van der Aa, A., Livingston, B.D., Crimi, C., Allosery, K., de Brabandere, V., Krakover, J., Murthy, S., Huang, M., Power, S., et ál. (2008). Rational design of a multiepitope vaccine encoding T-lymphocyte epitopes for treatment of chronic hepatitis B virus infections. Journal of Virology 82, 435–450.

75. Ishioka, G.Y., Fikes, J., Hermanson, G., Livingston, B., Crimi, C., Qin, M., del Guercio, M.F., Oseroff, C., Dahlberg, C., Alexander, J., et ál. (1999). Utilization of MHC class I transgenic mice for development of minigene DNA vaccines encoding multiple HLA-restricted CTL epitopes. J Immunol 162, 3915–3925.

76. Janetzki, S., Price, L., Schroeder, H., Britten, C.M., Welters, M.J.P. y Hoos, A. (2015). Guidelines for the automated evaluation of Elispot assays. Nat Protoc 10, 1098–1115.

77. Lyons, G.E., Moore, T., Brasic, N., Li, M., Roszkowski, J.J. y Nishimura, M.I. (2006). Influence of human CD8 on antigen recognition by T-cell receptor-transduced cells. Cancer Res 66, 11455–11461.

78. Nagai, K., Ochi, T., Fujiwara, H., An, J., Shirakata, T., Mineno, J., Kuzushima, K., Shiku, H., Melenhorst, J.J., Gostick, E., et ál. (2012). Aurora kinase A-specific T-cell receptor gene transfer redirects T lymphocytes to display effective antileukemia reactivity. Blood 119, 368-376.

79. Panina-Bordignon, P., Tan, A., Termijtelen, A., Demotz, S., Corradin, G. y Lanzavecchia, A. (1989). Universally immunogenic T cell epitopes: promiscuous binding to human MHC class II and promiscuous recognition by T cells. Eur J Immunol 19, 2237–2242.

80. Vitiello, A., Marchesini, D., Furze, J., Sherman, L.A. y Chesnut, R.W. (1991). Analysis of the HLA-restricted influenza-specific cytotoxic T lymphocyte response in transgenic mice carrying a chimeric human-mouse class I major histocompatibility complex. J Exp Med 173, 1007–1015.

81. Yachi, P.P., Ampudia, J., Zal, T. y Gascoigne, N.R.J. (2006). Altered peptide ligands induce delayed CD8-T cell receptor interaction--a role for CD8 in distinguishing antigen quality. Immunity 25, 203-211.

82. Pushko P, Parker M, Ludwig GV, Davis NL, Johnston RE, Smith JF. Replicon-helper systems from attenuated Venezuelan equine encephalitis virus: expression of heterologous genes in vitro and immunization against heterologous pathogens in vivo. Virology. 22 de dic. de 1997;239(2):389-401.

83. Strauss, JH y E G Strauss. The alphaviruses: gene expression, replication, and evolution. Microbiol

Rev. Sep. de 1994; 58(3): 491-562.

84. Rhême C, Ehrengruber MU, Grandgirard D. Alphaviral cytotoxicity and its implication in vector development. Exp Physiol. Ene. de 2005;90(1):45-52. Epub 12 de nov. de 2004.

85. Riley, Michael K. II, y Wilfred Vermerris. Recent Advances in Nanomaterials for Gene Delivery—A Review. Nanomaterials 2017, 7(5), 94.

86. Frolov I, Hardy R, Rice CM. Cis-acting RNA elements at the 5' end of Sindbis virus genome RNA regulate minus- and plus-strand RNA synthesis. RNA. Nov. de 2001;7(11):1638-51.

87. Jose J, Snyder JE, Kuhn RJ. A structural and functional perspective of alphavirus replication and assembly. Future Microbiol. Sep. de 2009;4(7):837-56.

88. Bo Li and C. olin N. Dewey. RSEM: accurate transcript quantification from RNA-Seq data with or without a referenfe genome. BMC Bioinformatics, 12:323, agosto de 2011

89. Hillary Pearson, Tariq Daouda, Diana Paola Granados, Chantal Durette, Eric Bonneil, Mathieu Courcelles, Anja Rodenbrock, Jean-Philippe Laverdure, Caroline Côté, Sylvie Mader, Sébastien Lemieux, Pierre Thibault, and Claude Perreault. MHC class I-associated peptides derive from selective regions of the human genome. The Journal of Clinical Investigation, 2016,

90. Juliane Liepe, Fabio Marino, John Sidney, Anita Jeko, Daniel E. Bunting, Alessandro Sette, Peter M. Kloetzel, Michael P. H. Stumpf, Albert J. R. Heck, Michele Mishto. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. Science, 21, octubre de 2016.

91. Mommen GP., Marino, F., Meiring HD., Poelen, MC., van Gaans-van den Brink, JA., Mohammed S., Heck AJ., and van Els CA. Sampling From the Proteome to the Human Leukocyte Antigen-DR (HLA-DR) Ligandome Proceeds Via High Specificity. Mol Cell Proteomics 15(4): 1412-1423, abril 2016.

92. Sebastian Kreiter, Mathias Vormehr, Niels van de Roemer, Mustafa Diken, Martin Löwer, Jan Diekmann, Sebastian Boegel, Barbara Schrörs, Fulvia Vascotto, John C. Castle, Arbel D. Tadmor, Stephen P. Schoenberger, Christoph Huber, Özlem Türeci, and Ugur Sahin. Mutant MHC class II epitopes drive therapeutic immune responses to caner. Nature 520, 692-696, abril de 2015.

93. Tran E., Turcotte S., Gros A., Robbins P.F., Lu Y.C., Dudley M.E., Wunderlich J.R., Somerville R.P., Hogan K., Hinrichs C.S., Parkhurst M.R., Yang J.C., Rosenberg S.A. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science 344(6184) 641-645, mayo de 2014.

94. Andreatta M., Karosiene E., Rasmussen M., Stryhn A., Buus S., Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics 67(11-12) 641-650, noviembre de 2015.

95. Nielsen, M., Lund, O. NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. BMC Bioinformatics 10:296, septiembre de 2009.

96. Nielsen, M., Lundegaard, C., Lund, O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. BMC Bioinformatics 8:238, julio de 2007.

97. Zhang, J., et ál. PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Molecular & Cellular Proteomics. 11(4):1-8. 1/2/2012.

98. Jensen, Kamilla Kjaergaard, et ál. “Improved Methods for Prediting Peptide Binding Affinity to MHC Class II Molecules.” Immunology, 2018, doi:10.1111/imm.12889.

99. Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et ál. (2012). Absolute quantification of somatic DNA alterations in human cancer. Nat. Biotechnol. 30, 413–421

100. McGranahan, N., Rosenthal, R., Hiley, C.T., Rowan, A.J., Watkins, T.B.K., Wilson, G.A., Birkbak, N.J., Veeriah, S., Van Loo, P., Herrero, J., et ál. (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell 171, 1259–1271.e11.

101. Shukla, S.A., Rooney, M.S., Rajasagi, M., Tiao, G., Dixon, P.M., Lawrence, M.S., Stevens, J., Lane, W.J., Dellagatta, J.L., Steelman, S., et ál. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat. Biotechnol. 33, 1152–1158.

102. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et ál. (2010). Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. U. S. A. 107, 16910–16915.

103. Van Loo, P., Nordgard, S.H., Lingjærde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B., et ál. (2010). Allele-specific copy number analysis of tumors. Proc. Natl. Acad. Sci. U. S. A. 107, 16910–16915.

DIVERSAS MODALIDADES

1. En la presente, se describe un vector viral que comprende un neoantígeno o múltiples neoantígenos. En determinadas modalidades, se identifica un neoantígeno mediante el uso de un método descrito en la presente, p. ej., más adelante. En determinadas modalidades, un neoantígeno tiene al menos una característica o propiedad descrita en la presente, p. ej., más adelante.

2. En la presente, se describe un método para identificar uno o más neoantígenos a partir de una célula tumoral de un sujeto que sea probable que se presenten en la superficie de la célula tumoral, que comprende las siguientes etapas:

obtener datos de secuenciación de nucleótidos tumorales del genoma completo, transcriptoma y/o exoma de la célula tumoral del sujeto, donde los datos de secuenciación de nucleótidos tumorales se usan para obtener datos que representen secuencias peptídicas de cada uno de un conjunto de neoantígenos, y donde la secuencia peptídica de cada neoantígeno comprende al menos una alteración

que la distingue con respecto a la secuencia peptídica original de tipo salvaje correspondiente;

ingresar la secuencia peptídica de cada neoantígeno en uno o más modelos de presentación para generar un conjunto de probabilidades numéricas de que uno o más alelos del MHC presenten cada uno de los neoantígenos en la superficie celular tumoral de la célula tumoral del sujeto, donde el conjunto de probabilidades numéricas se ha identificado al menos en función de datos de espectrometría de masas; y

seleccionar un subconjunto del conjunto de neoantígenos en función del conjunto de probabilidades numéricas para generar un conjunto de neoantígenos seleccionados.

3. En determinadas modalidades, una cantidad del conjunto de neoantígenos seleccionados es 20.

4. En determinadas modalidades, el modelo de presentación representa una dependencia entre:

la presencia de un par de alelos de MHC de un tipo particular y un aminoácido particular en una posición particular de una secuencia peptídica; y

la probabilidad de la presentación sobre la superficie de la célula tumoral, por parte del alelo del MHC particular del par, de esa secuencia peptídica que comprende el aminoácido particular en la posición particular.

5. En determinadas modalidades, el ingreso de la secuencia peptídica comprende:

aplicar el o los modelos de presentación a la secuencia peptídica del neoantígeno correspondiente para generar un puntaje de dependencia para cada uno del o de los alelos del MHC que indique si el alelo del MHC presentará el neoantígeno correspondiente en función de al menos posiciones de los aminoácidos de la secuencia peptídica del neoantígeno correspondiente.

6. En determinadas modalidades, el método comprende además:

transformar los puntajes de dependencia para generar una probabilidad correspondiente por alelo para cada alelo del MHC que indique la probabilidad de que el alelo del MHC correspondiente presente el neoantígeno correspondiente; y

combinar las probabilidades por alelo para generar la probabilidad numérica.

7. En determinadas modalidades, la transformación de los puntajes de dependencia puede modelar la presentación de la secuencia peptídica del neoantígeno correspondiente como mutuamente excluyente.

8. En determinadas modalidades, el método comprende además:

transformar una combinación de los puntajes de dependencia para generar la probabilidad numérica.

9. En determinadas modalidades, la transformación de la combinación de los puntajes de dependencia modela la presentación de la secuencia peptídica del neoantígeno correspondiente como una interferencia entre los alelos del MHC.

10. En determinadas modalidades, el conjunto de probabilidades numéricas se identifica de manera

adicional mediante al menos una característica de no interacción con el alelo, y comprende además:

aplicar un modelo de no interacción con el alelo del o de los modelos de presentación a las características de no interacción con el alelo para generar un puntaje de dependencia para las características de no interacción con el alelo que indique si se presentará la secuencia peptídica del neoantígeno correspondiente en función de las características de no interacción con el alelo.

11. En determinadas modalidades, el método comprende además:

combinar el puntaje de dependencia para cada alelo del MHC en el o los alelos del MHC con el puntaje de dependencia para la característica de no interacción con el alelo;

transformar los puntajes de dependencia combinados para cada alelo del MHC para generar una probabilidad correspondiente por alelo para el alelo del MHC que indique la probabilidad de que el alelo del MHC correspondiente presente el neoantígeno correspondiente; y

combinar las probabilidades por alelo para generar la probabilidad numérica.

12. En determinadas modalidades, el método comprende además:

transformar una combinación de los puntajes de dependencia para cada uno de los alelos del MHC y el puntaje de dependencia para las características de no interacción con el alelo para generar la probabilidad numérica.

13. En determinadas modalidades, se entrena un conjunto de parámetros numéricos para el modelo de presentación en función de un conjunto de datos de entrenamiento que incluye al menos un conjunto de secuencias peptídicas de entrenamiento identificadas como presentes en múltiples muestras y uno o más alelos del MHC asociados a cada secuencia peptídica de entrenamiento, donde las secuencias peptídicas de entrenamiento se identifican mediante una espectrometría de masas de los péptidos aislados eluidos de los alelos del MHC derivados de las múltiples muestras.

14. En determinadas modalidades, el conjunto de datos de entrenamiento incluye además datos sobre los niveles de expresión del ARNm de la célula tumoral.

15. En determinadas modalidades, las muestras comprenden líneas celulares modificadas genéticamente para expresar un solo alelo del MHC clase I o clase II.

16. En determinadas modalidades, las muestras comprenden líneas celulares modificadas genéticamente para expresar múltiples alelos del MHC clase I o clase II.

17. En determinadas modalidades, las muestras comprenden líneas celulares humanas obtenidas o derivadas de múltiples pacientes.

18. En determinadas modalidades, las muestras comprenden muestras de tumores frescas o congeladas obtenidas de múltiples pacientes.

19. En determinadas modalidades, las muestras comprenden muestras de tejidos frescas o congeladas obtenidas de múltiples pacientes.

20. En determinadas modalidades, las muestras comprenden péptidos identificados mediante el uso de ensayos de linfocitos T.

21. En determinadas modalidades, el conjunto de datos de entrenamiento comprende datos asociados a:

la abundancia de péptidos del conjunto de péptidos de entrenamiento presentes en las muestras;
la longitud de los péptidos del conjunto de péptidos de entrenamiento en las muestras;

22. En determinadas modalidades, se genera el conjunto de datos de entrenamiento al comparar el conjunto de secuencias peptídicas de entrenamiento mediante la alineación a una base de datos que comprende un conjunto de secuencias proteicas conocidas, donde el conjunto de secuencias proteicas de entrenamiento es más largo que las secuencias peptídicas de entrenamiento y las incluye.

23. En determinadas modalidades, el conjunto de datos de entrenamiento se genera en función de la realización o de haberse realizado una espectrometría de masas en una línea celular para obtener datos de secuencias peptídicas del genoma completo, transcriptoma y/o exoma de la línea celular, donde los datos de secuenciación incluyen al menos una secuencia proteica que incluye una alteración.

24. En determinadas modalidades, se genera el conjunto de datos de entrenamiento en función de la obtención de datos de secuenciación de nucleótidos normales del genoma completo, transcriptoma y/o exoma de muestras de tejidos normales.

25. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a secuencias del proteoma asociadas a las muestras.

26. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a secuencias del peptidoma del MHC asociadas a las muestras.

27. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a mediciones de afinidad de unión entre péptido-MHC para al menos uno de los péptidos aislados.

28. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a mediciones de estabilidad de unión entre péptido-MHC para al menos uno de los péptidos aislados.

29. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a transcriptomas asociados a las muestras.

30. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a genomas asociados a las muestras.

31. En determinadas modalidades, las secuencias peptídicas de entrenamiento tienen longitudes dentro de un intervalo de k-meros donde k es entre 8-15, inclusive.

32. En determinadas modalidades, el método comprende además codificar la secuencia peptídica mediante el uso de un esquema de codificación one-hot.

33. En determinadas modalidades, el método comprende además codificar la secuencia peptídica de entrenamiento mediante el uso de un esquema de codificación one-hot con left-pad.

34. En la presente, también se describe un método de tratamiento de un sujeto con un tumor, que comprende realizar cualquiera de las etapas de los métodos que se describen en la presente, y que comprende además obtener una vacuna tumoral que comprenda el conjunto de neoantígenos seleccionados, y administrar la vacuna tumoral al sujeto.

35. En la presente, también se describe un método de fabricación de una vacuna tumoral, que comprende realizar cualquiera de las etapas descritas en la presente, y que comprende además producir o haber producido una vacuna tumoral que comprenda el conjunto de neoantígenos seleccionados.

36. En la presente, también se describe una vacuna tumoral que comprende un conjunto de neoantígenos seleccionados, que se seleccionan mediante la realización de un método que se describe en la presente.

37. En determinadas modalidades, la vacuna tumoral comprende una secuencia de nucleótidos, una secuencia de polipéptidos, un ARN, un ADN, una célula, un plásmido y/o un vector.

38. En determinadas modalidades, la vacuna tumoral comprende uno o más neoantígenos que se presentan en la superficie de la célula tumoral.

39. En determinadas modalidades, la vacuna tumoral comprende uno o más neoantígenos que son inmunógenos en el sujeto.

40. En determinadas modalidades, la vacuna tumoral no comprende uno o más neoantígenos que induzcan una respuesta autoinmunitaria contra el tejido normal en el sujeto.

41. En determinadas modalidades, la vacuna tumoral comprende además un adyuvante.

42. En determinadas modalidades, la vacuna tumoral comprende además un excipiente.

43. En determinadas modalidades, la selección del conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con una mayor probabilidad de presentarse sobre la superficie de la célula tumoral con respecto a neoantígenos no seleccionados en función del modelo de presentación.

44. En determinadas modalidades, la selección del conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con mayor probabilidad de poder inducir una respuesta inmunitaria específica para el tumor en el sujeto con respecto a neoantígenos no seleccionados en función del modelo de presentación.

45. En determinadas modalidades, la selección del conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con una mayor probabilidad de poder ser presentados a linfocitos

T no activados por parte de células presentadoras de antígenos (APC) con respecto a neoantígenos no seleccionados en función del modelo de presentación, donde opcionalmente la APC es una célula dendrítica (DC).

46. En determinadas modalidades, la selección del conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con una menor probabilidad de someterse a inhibición a través de una tolerancia central o periférica con respecto a neoantígenos no seleccionados en función del modelo de presentación.

47. En determinadas modalidades, la selección del conjunto de neoantígenos seleccionados comprende seleccionar neoantígenos con menor probabilidad de poder inducir una respuesta autoinmunitaria al tejido normal en el sujeto con respecto a neoantígenos no seleccionados en función del modelo de presentación.

48. En determinadas modalidades, se obtienen datos de secuenciación de nucleótidos del exoma o transcriptoma mediante la realización de una secuenciación en el tejido tumoral.

49. En determinadas modalidades, la secuenciación es secuenciación de nueva generación (NGS) o cualquier estrategia de secuenciación masiva en paralelo.

50. En determinadas modalidades, el conjunto de probabilidades numéricas se identifica además mediate al menos las características de interacción MHC-alelo que comprenden al menos una de las siguientes:

a. La afinidad prevista para la unión entre el alelo del MHC y el péptido codificado por el neoantígeno.

b. La estabilidad prevista del complejo péptido-MHC codificado por el neoantígeno.

c. La secuencia y longitud del péptido codificado por el neoantígeno.

d. La probabilidad de presentación de péptidos codificados por el neoantígeno con una secuencia similar en células de otros individuos que expresen el alelo del MHC particular tal como se evalúa mediante un análisis proteómico por espectrometría de masas u otro medio.

e. Los niveles de expresión del alelo del MHC particular en el sujeto en cuestión (p. ej., tal como se mide mediante secuenciación de ARN o espectrometría de masas).

f. La probabilidad con independencia de las secuencias peptídicas codificadas por neoantígenos a nivel general de presentación por parte del alelo del MHC particular en otros sujetos distintos que expresan el alelo del MHC particular.

g. La probabilidad con independencia de las secuencias peptídicas codificadas por neoantígenos a nivel general de presentación por parte de alelos del MHC en la misma familia de moléculas (p. ej., HLA-A, HLA-B, HLA-C, HLA-DQ, HLA-DR, HLA-DP) en otros sujetos distintos.

51. En determinadas modalidades, el conjunto de probabilidades numéricas se identifica además

mediante al menos características de no interacción MHC-alelo que comprenden al menos una de las siguientes:

a. Las secuencias de extremo C y N que flanquean el péptido codificado por el neoantígeno dentro de su secuencia proteica de origen.

b. La presencia de motivos de escisión de la proteasa en el péptido codificado por el neoantígeno, ponderada opcionalmente de acuerdo con la expresión de proteasas correspondientes en las células tumorales (tal como se mide mediante secuenciación de ARN o espectrometría de masas).

c. La tasa de renovación de la proteína de origen tal como se mide en el tipo de célula adecuado.

d. La longitud de la proteína de origen, considerando opcionalmente las variantes de corte y empalme específicas («isoformas») de expresión más alta en las células tumorales tal como se mide mediante secuenciación de ARN o espectrometría de masas del proteoma, o según la predicción a partir de la anotación de mutaciones de corte y empalme somáticas o de la línea germinal detectadas en los datos de secuencias de ADN o ARN.

e. El nivel de expresión del proteasoma, inmunoproteasoma, timoproteasoma u otras proteasas en las células tumorales (que se puede medir mediante secuenciación de ARN, espectrometría de masas del proteoma o inmunohistoquímica).

f. La expresión del gen de origen del péptido codificado por el neoantígeno (p. ej., tal como se mide mediante secuenciación de ARN o espectrometría de masas).

g. La expresión específica con respecto al tejido habitual del gen de origen del péptido codificado por el neoantígeno durante diversas etapas del ciclo celular.

h. Un catálogo abarcativo de características de la proteína de origen y/o sus dominios tal como puede encontrarse en, p. ej., uniProt o PDB http://www.rcsb.org/pdb/home/home.do.

i. Las características que describen las propiedades del dominio de la proteína de origen que contiene el péptido, por ejemplo: una estructura secundaria o terciaria (p. ej., hélice alfa con respecto a lámina beta); corte y empalme alternativo.

j. La probabilidad de presentación de péptidos de la proteína de origen del péptido codificado por el neoantígeno en cuestión en otros sujetos distintos.

k. La probabilidad de que el péptido no sea detectado o esté representado en exceso en la espectrometría de masas debido a sesgos técnicos.

l. La expresión de diversos módulos/vías de genes tal como se mide mediante secuenciación de ARN (sin la necesidad de contener la proteína de fuente del péptido) que son indicativos del estado de las células tumorales, el estroma o linfocitos de infiltración (TIL).

m. La cantidad de copias del gen de origen del péptido codificado por el neoantígeno en las células tumorales.

n. La probabilidad de que el péptido se una a la TAP o la afinidad de unión medida o prevista del péptido con respecto a la TAP.

o. El nivel de expresión de la TAP en las células tumorales (que se puede medir mediante secuenciación de ARN, espectrometría de masas del proteoma, inmunohistoquímica).

p. La presencia o ausencia de mutaciones tumorales, que incluyen, de modo no taxativo:

i. Mutaciones impulsoras en genes impulsores de cáncer conocidos tales como EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3.

ii. En genes que codifican las proteínas involucradas en la maquinaria de presentación de antígenos (p. ej., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 o cualquiera de los genes que codifican componentes del proteasoma o inmunoproteasoma). Los péptidos cuya presentación depende de un componente de la maquinaria de presentación de antígenos que se somete a una mutación de pérdida de función en el tumor tienen una probabilidad de presentación reducida.

q. La presencia o ausencia de polimorfismos de línea germinal funcionales, que incluyen, de modo no taxativo:

i. En genes que codifican las proteínas involucradas en la maquinaria de presentación de antígenos (p. ej., B2M, HLA-A, HLA-B, HLA-C, TAP-1, TAP-2, TAPBP, CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOB, HLA-DP, HLA-DPA1, HLA-DPB1, HLA-DQ, HLA-DQA1, HLA-DQA2, HLA-DQB1, HLA-DQB2, HLA-DR, HLA-DRA, HLA-DRB1, HLA-DRB3, HLA-DRB4, HLA-DRB5 o cualquiera de los genes que codifican componentes del proteasoma o inmunoproteasoma).

r. Tipo de tumor (p. ej., NSCLC, melanoma).

s. Subtipo de tumor clínico (p. ej., cáncer pulmonar escamoso con respecto a no escamoso).

t. Historial de tabaquismo.

u. La expresión típica del gen de origen del péptido en el tipo de tumor o subtipo clínico pertinente, opcionalmente estratificado mediante la mutación impulsora.

52. En determinadas modalidades, la o las mutaciones son un indel de desplazamiento de marco o sin desplazamiento de marco, una sustitución de cambio de sentido o sin sentido, una alteración en el sitio de corte y empalme, una reorganización genómica o fusión genética, o cualquier alteración genómica o de expresión que conduzca a un neoORF.

53. En determinadas modalidades, la célula tumoral se selecciona dentro del grupo que consiste en: cáncer pulmonar, melanoma, cáncer de mama, cáncer de ovario, cáncer de próstata, cáncer renal, cáncer gástrico, cáncer de colon, cáncer testicular, cáncer de cabeza y cuello, cáncer pancreático,

cáncer cerebral, linfoma de linfocitos B, leucemia mielógena aguda, leucemia mielógena crónica, leucemia linfocítica crónica y leucemia linfocítica de linfocitos T, cáncer pulmonar no microcítico y cáncer pulmonar microcítico.

54. En determinadas modalidades, el método comprende además obtener una vacuna tumoral que comprende el conjunto de neoantígenos seleccionados o un subconjunto de estos, que opcionalmente comprende además administrar la vacuna tumoral al sujeto.

55. En determinadas modalidades, al menos uno de los neoantígenos en el conjunto de neoantígenos seleccionados, cuando se encuentra en forma de polipéptido, comprende al menos una de las siguientes: una afinidad de unión con el MHC con un valor de IC50 de menos de 1000 nM; para polipéptidos de MHC Clase 1, una longitud de 8-15, 8, 9, 10, 11, 12, 13, 14 o 15 aminoácidos; la presencia de motivos de secuencias dentro o cerca del polipéptido en la secuencia proteica original que promueve la escisión del proteasoma; y la presencia de motivos de secuencias que promueven el transporte de TAP.

56. En la presente, también se describe un método para generar un modelo para identificar uno o más neoantígenos con probabilidad de presentarse sobre una superficie celular tumoral de una célula tumoral, que comprende ejecutar las siguientes etapas:

recibir datos de espectrometría de masas que comprendan datos asociados a múltiples péptidos asilados eluidos de un complejo principal de histocompatibilidad (MHC) derivado de múltiples muestras;

obtener un conjunto de datos de entrenamiento mediante al menos la identificación de un conjunto de secuencias peptídicas de entrenamiento presentes en las muestras y uno o más MHC asociados a cada secuencia peptídica de entrenamiento;

entrenar un conjunto de parámetros numéricos de un modelo de presentación mediante el uso del conjunto de datos de entrenamiento que comprende las secuencias peptídicas de entrenamiento, donde el modelo de presentación proporciona múltiples probabilidades numéricas de que uno o más alelos del MHC sobre la superficie celular tumoral presenten secuencias peptídicas de la célula tumoral.

57. En determinadas modalidades, el modelo de presentación representa una dependencia entre:

presencia de un aminoácido particular en una posición particular de una secuencia peptídica; y

la probabilidad de la presentación por parte de uno de los alelos del MHC en la célula tumoral, de la secuencia peptídica que contiene el aminoácido particular en la posición particular.

58. En determinadas modalidades, las muestras comprenden líneas celulares modificadas genéticamente para expresar un solo alelo del MHC clase I o clase II.

59. En determinadas modalidades, las muestras comprenden líneas celulares modificadas genéticamente para expresar múltiples alelos del MHC clase I o clase II.

60. En determinadas modalidades, las muestras comprenden líneas celulares humanas obtenidas o

derivadas de múltiples pacientes.

61. En determinadas modalidades, las muestras comprenden muestras de tumores frescas o congeladas obtenidas de múltiples pacientes.

62. En determinadas modalidades, las muestras comprenden péptidos identificados mediante el uso de ensayos de linfocitos T.

63. En determinadas modalidades, el conjunto de datos de entrenamiento comprende datos asociados a:

la abundancia de péptidos del conjunto de péptidos de entrenamiento presentes en las muestras;

la longitud de los péptidos del conjunto de péptidos de entrenamiento en las muestras;

64. En determinadas modalidades, la obtención del conjunto de datos de entrenamiento comprende: obtener un conjunto de secuencias proteicas de entrenamiento en función de las secuencias peptídicas de entrenamiento al comparar el conjunto de secuencias peptídicas de entrenamiento a través de la alineación con respecto a una base de datos que comprende un conjunto de secuencias proteicas conocidas, donde el conjunto de secuencias proteicas de entrenamiento es más largo que las secuencias peptídicas de entrenamiento y las incluye.

65. En determinadas modalidades, la obtención del conjunto de datos de entrenamiento comprende: realizar o haber realizado una espectrometría de masas en una línea celular para obtener datos de secuenciación de nucleótidos del genoma completo, transcriptoma y/o exoma de la línea celular, donde los datos de secuenciación de nucleótidos incluyen al menos una secuencia proteica que incluye una mutación.

66. En determinadas modalidades, entrenar el conjunto de parámetros del modelo de presentación comprende:

codificar las secuencias peptídicas de entremaniento mediante el uso de un esquema de codificación one-hot.

67. En determinadas modalidades, el método comprende además:

obtener datos de secuenciación de nucleótidos normales del genoma completo, transcriptoma y/o exoma de muestras de tejidos normales; y

entrenar el conjunto de parámetros del modelo de presentación mediante el uso de los datos de secuenciación de nucleótidos normales.

68. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a secuencias del proteoma asociadas a las muestras.

69. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a secuencias del peptidoma del MHC asociadas a las muestras.

70. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos

asociados a mediciones de afinidad de unión entre péptido-MHC para al menos uno de los péptidos aislados.

71. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a mediciones de estabilidad de unión entre péptido-MHC para al menos uno de los péptidos aislados.

72. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a transcriptomas asociados a las muestras.

73. En determinadas modalidades, el conjunto de datos de entrenamiento comprende además datos asociados a genomas asociados a las muestras.

74. En determinadas modalidades, entrenar el conjunto de parámetros numéricos comprende además:

realizar una regresión logística del conjunto de parámetros.

75. En determinadas modalidades, las secuencias peptídicas de entrenamiento tienen longitudes dentro de un intervalo de k-meros donde k es entre 8-15, inclusive.

76. En determinadas modalidades, entrenar el conjunto de parámetros numéricos del modelo de presentación comprende:

codificar las secuencias peptídicas de entremaniento mediante el uso de un esquema de codificación one-hot con left-pad.

77. En determinadas modalidades, entrenar el conjunto de parámetros numéricos comprende además:

determinar valores para el conjunto de parámetros mediante el uso de un algoritmo de aprendizaje profundo.

78. En la presente, también se describe un método para generar un modelo para identificar uno o más neoantígenos con probabilidad de presentarse sobre una superficie celular tumoral de una célula tumoral, que comprende ejecutar las siguientes etapas:

recibir datos de espectrometría de masas que comprendan datos asociados a múltiples péptidos asilados eluidos de un complejo principal de histocompatibilidad (MHC) derivado de múltiples muestras de tumores frescas o congeladas;

obtener un conjunto de datos de entrenamiento mediante al menos la identificación de un conjunto de secuencias peptídicas de entrenamiento presentes en las muestras de tumores y uno o más alelos del MHC asociados a cada secuencia peptídica de entrenamiento;

obtener un conjunto de secuencias proteicas de entrenamiento en función de las secuencias peptídicas de entrenamiento; y

entrenar un conjunto de parámetros numéricos de un modelo de presentación mediante el uso de las

secuencias proteicas de entrenamiento y las secuencias peptídicas de entrenamiento, donde el modelo de presentación proporciona múltiples probabilidades numéricas de que uno o más alelos del MHC sobre la superficie celular tumoral presenten secuencias peptídicas de la célula tumoral.

79. En determinadas modalidades, el modelo de presentación representa una dependencia entre:

la presencia de un par de alelos de MHC de un tipo particular y un aminoácido particular en una posición particular de una secuencia peptídica; y

la probabilidad de la presentación sobre la superficie de la célula tumoral, por parte del alelo del MHC particular del par, de esa secuencia peptídica que comprende el aminoácido particular en la posición particular.