

VECTORES VIRALES PARA EL SUMINISTRO DE NEOANTÍGENOS

REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

La presente solicitud reivindica el beneficio de las solicitudes provisionales estadounidenses n.º 62/425,996, presentada el 23 de noviembre de 2016; 62/435,266, presentada el 16 de diciembre de 2016; 62/503,196, presentada el 8 de mayo de 2017; y 62/523,212, presentada el 21 de junio de 2017; cada una de las cuales se incorpora en su totalidad a la presente mediante referencia.

LISTADO DE SECUENCIAS

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 el 27 de diciembre de 2017, se denomina 38804PCT_CRF_sequencelisting.txt y tiene un tamaño de 467.920 bytes.

ANTECEDENTES

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 T⁶ y que la terapia celular dirigida a neoantígenos puede producir regresión tumoral en determinadas circunstancias en pacientes seleccionados.⁷

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.

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 candidatos⁸. Sin embargo, estos métodos propuestos pueden fallar en modelar la totalidad del proceso de generación de epítomos, que contiene muchas etapas (p. ej., transporte de TAP, escisión proteasomal y/o reconocimiento de TCR) además de la expresión génica y unión a MHC⁹. Por consiguiente, es probable que los métodos existentes tengan un reducido valor predictivo positivo (PPV) bajo. (Figura 1A)

De hecho, los análisis de péptidos presentados por células tumorales realizados por

múltiple 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 inhibitoria del punto de control sobre la cantidad de mutaciones solamente.¹²

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)

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 genes¹³, y mutaciones que crean o eliminan los sitios de escisión de la proteasa.

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.

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 puede ser un importante obstáculo para el uso de virus recombinantes humanos para la administración de neoantígenos para tratamiento contra el cáncer.

COMPENDIO

En la presente se describe el vector de adenovirus de chimpancé que comprende un casete de neoantígeno, donde el casete de neoantígeno comprende: (1) múltiples secuencias de ácidos nucleicos que codifican el neoantígeno derivadas de un tumor presente en un sujeto, donde las múltiples secuencias comprenden: al menos dos secuencias de ácidos nucleicos que codifican el neoantígeno del MHC clase I específico del tumor y específico del sujeto cada una que comprende: a. una secuencia de ácidos nucleicos que codifican epítipo de MHC clase I con al menos una alteración que distingue a la secuencia peptídica codificada con respecto a la secuencia peptídica correspondiente codificada por la secuencia de ácidos nucleicos de tipo salvaje, b. opcionalmente una secuencia enlazadora 5' y c. opcionalmente una secuencia enlazadora 3'; (2) al menos una secuencia promotora enlazada de forma operativa a al menos una de las múltiples secuencias, (3) opcionalmente, al menos una secuencia de ácidos nucleicos que codifican el antígeno del MHC clase II; (4) opcionalmente, al menos una secuencia enlazadora GPGPG (SEQ ID NO:56); y (5) opcionalmente, al menos una secuencia de poliadenilación.

También se describe en la presente un vector de adenovirus de chimpancé que comprende: a. una secuencia de ChAdV68 modificada que comprende la secuencia de SEQ ID NO:1 con una eliminación de E1 (nt 577 a 3403) y una eliminación de E3 (nt 27,125-31,825); b. una secuencia promotora del CMV; c. una secuencia de nucleótidos de la señal de poliadenilación del SV40; y d. un casete de neoantígeno, donde el casete de neoantígeno comprende: (1) múltiples secuencias de ácidos nucleicos que codifican el neoantígeno derivadas de un tumor presente en un sujeto, donde las múltiples secuencias comprenden: al menos 20 secuencias de ácidos nucleicos que codifican el neoantígeno del MHC clase I específico del tumor y específico del sujeto enlazadas linealmente entre sí y cada una comprende: (A) una secuencia de ácidos nucleicos que codifican el epítipo del MHC clase I con al menos una alteración que distingue a la secuencia peptídica codificada con respecto a la secuencia peptídica correspondiente codificada por una secuencia de ácidos nucleicos de tipo salvaje, donde la secuencia de ácidos nucleicos que codifican el epítipo del MHC clase I codifica un epítipo del MHC clase I de 7-15 aminoácidos de longitud, (B) una secuencia enlazadora 5', donde la secuencia enlazadora 5' es una secuencia de ácidos nucleicos 5' natural del epítipo del MHC I y donde la secuencia enlazadora 5' codifica un péptido que es de al menos 5 aminoácidos de longitud, (C) una secuencia enlazadora 3', donde la secuencia enlazadora 3' es una secuencia de ácidos nucleicos 3' natural del epítipo del MHC I y donde la secuencia enlazadora 3' codifica un

péptido que es de al menos 5 aminoácidos de longitud y donde cada una de las secuencias de ácido nucleico que codifica neoantígenos de MHC de clase I codifica un polipéptido con 25 aminoácidos de longitud y donde cada extremo 3' de cada secuencia de ácidos nucleicos que codifican el neoantígeno del MHC clase I está enlazado al extremo 5' de la siguiente

5 secuencia de ácidos nucleicos que codifican el neoantígeno del MHC clase I a excepción de la secuencia de ácidos nucleicos que codifican el neoantígeno del MHC clase I final de las múltiples secuencias; y (2) al menos dos secuencias de ácidos nucleicos que codifican el antígeno del MHC clase II que comprenden: (A) una secuencia del MHC clase II de PADRE (SEQ ID NO:48), (B) una secuencia del MHC clase II de toxoide tetánico (SEQ ID

10 NO:46), (C) una secuencia de primer enlazador GP GPG (SEQ ID NO: 56) que enlaza la secuencia del MHC clase II de PADRE y la secuencia del MHC clase II de toxoide tetánico, (D) una segunda secuencia enlazadora GP GPG (SEQ ID NO: 56) que enlaza el extremo 5' de las al menos dos secuencias de ácidos nucleicos que codifican el antígeno del MHC clase II a las múltiples secuencias de ácidos nucleicos que codifican el neoantígeno, (E)

15 una tercera secuencia enlazadora GP GPG (SEQ ID NO: 56) que enlaza el extremo 3' de las al menos dos secuencias de ácidos nucleicos que codifican el antígeno del MHC clase II a la secuencia de nucleótidos de la señal de poliadenilación del SV40; y donde el casete de neoantígeno se inserta dentro de la eliminación E1 y la secuencia promotora del CMV se enlaza de forma operativa al casete de neoantígeno.

20 En algunos aspectos, el vector tiene una secuencia ordenada de cada elemento del vector que está descrito en la fórmula, de 5' a 3', que comprende:

$$P_a-(L5_b-N_c-L3_d)_X-(G5_e-U_f)_Y-G3_g-A_h$$

donde P comprende la o las secuencias promotoras enlazadas de forma operativa a al menos una secuencia de las múltiples secuencias, donde un vector de adenovirus de chimpancé, opcionalmente= 1, N comprende una de las secuencias de ácidos nucleicos que codifican el epítipo del MHC clase I con al menos una alteración que distingue a la secuencia peptídica codificada con respecto a la secuencia peptídica correspondiente codificada por la secuencia de ácidos nucleicos de tipo salvaje, donde c = 1, L5 comprende la secuencia enlazadora 5', donde b = 0 o 1, L3 comprende la secuencia enlazadora 3',

25 donde d = 0 o 1, G5 comprende una de la o las secuencias enlazadoras GP GPG (SEQ ID NO: 56), donde e = 0 o 1, G3 comprende una de la o las secuencias enlazadoras GP GPG (SEQ ID NO: 56), donde g = 0 o 1, U comprende una de la o las secuencias de ácidos nucleicos que codifican el antígeno del MHC clase II, donde f = 1, A comprende la o las secuencias de poliadenilación, donde h = 0 o 1, X = 2 a 400, donde para cada X el Nc

30

correspondiente es una secuencia de ácidos nucleicos que codifican el epítipo del MHC clase I C68distinto y $Y = 0-2$, donde para cada Y la correspondiente secuencia de ácidos nucleicos que codifican el antígeno del MHC clase II Uf. En un aspecto particular, $b = 1$, $d = 1$, $e = 1$, $g = 1$, $h = 1$, $X = 20$, $Y = 2$, P es una secuencia promotora del CMV, cada N codifica un epítipo MHC clase I con una longitud de 7-15 aminoácidos, $L5$ es a una secuencia de ácido nucleico 5' natural del epítipo del MHC I, y donde la secuencia enlazadora 5' codifica un péptido que es de al menos 5 aminoácidos de longitud, $L3$ es a una secuencia de ácidos nucleicos 3' natural del epítipo MHC I y donde la secuencia enlazadora 3' codifica un péptido que es de al menos 5 aminoácidos de longitud, U es cada una de una secuencia clase II de PADRE y una secuencia del MHC clase II de toxoide tetánico, el vector de adenovirus de chimpancé comprende una secuencia de ChAdV68 modificada que comprende la secuencia de SEQ ID NO:1 con una eliminación E1 (nt 577 a 3403) y una eliminación E3 (nt 27,125- 31,825) y el casete de neoantígeno se inserta en la eliminación E1, y cada una de las secuencias de ácidos nucleicos que codifican el neoantígeno del MHC clase I codifica un polipéptido que es de 25 aminoácidos de longitud.

En algunos aspectos, al menos 1, 2 u opcionalmente 3 secuencias de ácidos nucleicos que codifican el neoantígeno de las múltiples secuencias codifican secuencias de polipéptidos o partes de estas que son presentadas por MHC clase I en la superficie de la célula tumoral.

En algunos aspectos, cada una de las secuencias de ácidos nucleicos que codifican el antígeno de las múltiples secuencias están directamente enlazadas entre sí. En algunos aspectos, al menos una secuencia de ácidos nucleicos que codifican el antígeno de las múltiples secuencias está enlazada a una secuencia de ácidos nucleicos que codifican el antígeno distinta de las múltiples secuencias con 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 GP GPG (SEQ ID NO: 56).

En algunos aspectos, al menos una secuencia de las múltiples secuencias 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 las múltiples secuencias. 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.

En algunos aspectos, al menos una de las secuencias de ácidos nucleicos que codifican el neoantígeno de las múltiples secuencias 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 las secuencias de ácido nucleicos que codifican 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 parental de tipo salvaje correspondiente traducida. En algunos aspectos, al menos una de las secuencias de ácidos nucleicos que codifican 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.

En algunos aspectos, al menos una alteración comprende una mutación 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.

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 cerebro, linfoma de linfocitos B, leucemia mielógena aguda, leucemia mielógena crónica, leucemia linfocítica crónica, leucemia linfocítica de linfocitos T, cáncer pulmonar no microcítico y cáncer pulmonar microcítico.

5 En algunos aspectos, la expresión de cada secuencia de las múltiples secuencias es impulsada por el o los promotores.

En algunos aspectos, las múltiples secuencias comprenden al menos 2, 3, 4, 5, 6, 7, 8, 9 o 10 secuencia de ácidos nucleicos. En algunos aspectos, las múltiples secuencias comprenden al menos 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o hasta 400 secuencia de ácidos nucleicos. En algunos aspectos, las múltiples secuencias comprenden al menos 2-400
10 secuencia de ácidos nucleicos y donde al menos dos de las secuencias de ácidos nucleicos que codifican el neoantígeno de las múltiples secuencias codifican secuencias de polipéptidos o partes de estas que son presentadas por MHC I en la superficie de la célula tumoral. En algunos aspectos, las múltiples secuencias comprenden al menos 2-400
15 secuencia de ácidos nucleicos y donde, cuando se administran al sujeto y se traducen, al menos uno de los neoantígenos 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, las múltiples secuencias comprenden al menos 2-400 secuencias de ácidos nucleicos que codifican el neoantígeno
20 del MHC clase I y/o clase II, donde, 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 las al menos 2-400 secuencias de ácidos nucleicos que codifican
25 el neoantígeno del MHC clase I o clase II es impulsada por al menos un promotor.

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
30 aminoácidos.

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 alteración distingue a la secuencia peptídica codificada con respecto a la secuencia peptídica correspondiente codificada por la secuencia de ácidos nucleicos de tipo salvaje. 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.

En algunos aspectos, la o las secuencias promotoras son inducibles. En algunos aspectos, la o las secuencias promotoras son no inducibles. En algunos aspectos, la o las secuencias promotoras son una secuencia promotora del CMV, SV40, EF-1, RSV, PGK o EBV.

En algunos aspectos, el casete de neoantígeno comprende además al menos una secuencia de poliadenilación (poliA) enlazada de forma operativa a al menos una de las múltiples secuencias, donde opcionalmente la secuencia de poliA se encuentra en una ubicación 3' con respecto a la o las múltiples secuencias. En algunos aspectos, la secuencia de poliA comprende una secuencia de poliA del SV40. 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) 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 enlaza de manera operativa a al menos una de las múltiples secuencias. En algunos aspectos, el casete de neoantígeno comprende además un gen indicador, que incluye, de modo no taxativo, proteína fluorescente verde (GFP), una variante de GFP, fosfatasa alcalina secretada, luciferasa o una variante de luciferasa.

En algunos aspectos, el vector comprende 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.

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.

En algunos aspectos, el vector es un vector C68 de adenovirus de chimpancé. En algunos aspectos, el vector comprende la secuencia que se establece en las SEQ ID NO:1. En algunos aspectos, el vector 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 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.

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

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

En algunos aspectos, el vector de adenovirus comprende una o más eliminaciones entre el par de bases número 577 y 3403 o entre el par de bases 456 y 3014, y donde opcionalmente el vector comprende además una o más eliminaciones entre el par de bases 27 125 y 31 825 o entre el par de bases 27 816 y 31 333 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.

En algunos aspectos, las al menos dos secuencias de ácidos nucleicos que codifican el neoantígeno del MHC clase I se seleccionan al llevar a cabo las etapas de: 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; 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 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 las al menos dos secuencias de ácidos nucleicos que codifican el neoantígeno del MHC clase I.

En algunos aspectos, cada una de las secuencias de ácidos nucleicos que codifican el epítipo del MHC clase I se seleccionan al llevar a cabo las etapas de: 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; 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 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 las al menos dos secuencias de ácidos nucleicos que codifican el neoantígeno del MHC clase I.

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

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.

En algunos aspectos, seleccionar el conjunto de neoantígenos seleccionados
 5 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
 10 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
 15 (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
 20 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
 25 en paralelo.

En algunos aspectos, el casete de neoantígeno comprende secuencias de epítipo de unión formadas por secuencias adyacentes en el casete de neoantígeno. En algunos aspectos, la o cada secuencia de epítipo de unión tiene una afinidad mayor que 500 nM para MHC. En algunos aspectos, cada secuencia de epítipo de unión es no propia. En
 30 algunos aspectos, el casete de neoantígeno no codifica una secuencia de ácidos nucleicos de epítipo 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 epítipo no terapéutico aparecerá en un alelo del MHC del sujeto. En algunos aspectos, la secuencia de epítipo del MHC clase I o clase II no terapéutico predicho es una secuencia de epítipo 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 epítomos no terapéuticos en un modelo de presentación. En algunos aspectos, un orden de las múltiples secuencias de ácidos nucleicos que codifican el antígeno en el casete de neoantígeno se determina mediante una serie de etapas que comprenden: 1. generar un conjunto de secuencias de casete de neoantígeno correspondiente a diferentes órdenes de las múltiples secuencias de ácidos nucleicos que codifican el antígeno; 2. determinar, para cada secuencia de casete de neoantígeno candidato, una puntuación de presentación en función de la presentación de epítomos no terapéuticos en la secuencia de casete de neoantígeno candidato y 3. 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.

También se describe en la presente una composición farmacéutica que comprende un vector descrito en la presente (tal como un vector basado en ChAd descrito en la presente) y un portador farmacéuticamente aceptable. En algunos aspectos, la composición comprende adicionalmente un adyuvante. En algunos aspectos, la composición 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.

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.

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 CHO, HEK293 o variantes de esta, 911, HeLa, A549, LP-293, PER.C6 o AE1-2a.

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

También se describe en la presente un kit que comprende un vector descrito en la presente e instrucciones de uso.

5 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, el vector o la composición se administran por vía intramuscular (IM), intradérmica (ID) o subcutánea (SC). En algunos aspectos, el método comprende además administrar al sujeto un
10 modulador inmunitario, donde opcionalmente el modulador inmunitario se administra antes de, al mismo tiempo que, o después de la administración del vector o la composición farmacéutica. 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
15 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, donde la administración subcutánea se encuentra cerca del sitio de administración del vector o la composición o en proximidad estrecha con
20 respecto a uno o más ganglios linfáticos de drenaje del vector o de la composición.

En algunos aspectos, el método comprende además administrar al sujeto una segunda composición de vacuna. En algunos aspectos, la segunda composición de vacuna se administra antes de la administración del vector o la composición farmacéutica de cualquiera de los vectores o composiciones anteriores. En algunos aspectos, la segunda
25 composición de vacuna se administra después de la administración del vector o la composición farmacéutica de cualquiera de los vectores o composiciones anteriores. En algunos aspectos, la segunda composición de vacuna es la misma que el vector o la composición farmacéutica de cualquiera de los vectores o composiciones anteriores. En algunos aspectos, la segunda composición de vacuna es diferente del vector o la
30 composición farmacéutica de cualquiera de los vectores o composiciones anteriores. En algunos aspectos, la segunda composición de vacuna comprende un vector de ARN autorreplicante (srRNA) que codifica múltiples secuencias de ácidos nucleicos que codifican el neoantígeno. En algunos aspectos, las múltiples secuencias de ácidos nucleicos que codifican el neoantígeno codificadas por el vector de srRNA son iguales que las múltiples

secuencias de ácidos nucleicos que codifican el neoantígeno de cualquiera de las reivindicaciones de vectores anteriores.

También se describe en la presente un método de elaboración de un vector 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 la o las células hospedadoras.

En algunos aspectos, aislar comprende: lisar la célula hospedadora para obtener un lisado celular que comprenda el vector; y purificar el vector del lisado celular y opcionalmente también de medios usados para cultivar la célula hospedadora.

En algunos aspectos, la secuencia plasmídica se genera mediante el uso de una o más de las siguientes: recombinación de ADN o 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 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

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:

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

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

La FIG. 1C muestra el impacto del problema de especificidad de predicción de neoantígeno.

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

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

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

La FIG. 1G muestra cómo la adición de características aumenta el valor predictivo

positivo del modelo.

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

Las FIG. 2B y FIG. 2C ilustran un método para obtener información de presentación, de acuerdo con una modalidad. La FIG. 2B muestra la SEQ ID NO: 61. La FIG. 2C muestra las SEQ ID NOs: 61-66, respectivamente, en orden de aparición.

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.

La FIG. 4 ilustra un conjunto de datos de entrenamiento de ejemplo, de acuerdo con una modalidad. La figura describe las secuencias pépticas de las SEQ ID Nos: 68-71 y las secuencias flanqueantes C de las SEQ ID NOs: 72 y 154-155, respectivamente, en orden de aparición.

La FIG. 5 ilustra un modelo de red de ejemplo asociado con un alelo de MHC.

La FIG. 6 ilustra un modelo de red de ejemplo compartido por alelos de MHC.

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.

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.

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.

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.

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.

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.

La FIG. 13 ilustra los resultados de rendimiento de varios modelos de presentación de ejemplo.

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

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.

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

La FIG 16B ilustra la evaluación de secuencias enlazadoras en casetes cortos y muestra información de secuencia sobre los epítomos de linfocitos T integrados en los casetes cortos. La figura muestra las SEQ ID NOs: 128-129, 132, 131, 130, 49 y 156, respectivamente, en orden de aparición.

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 epítomos de linfocitos T humanos marcadores (epítomos 1 a 5) también dos epítomos 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 epítomos de linfocitos T en ambos lados (25mer).

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.

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

La FIG. 19B ilustra evaluación *in vivo* del impacto de la posición del epítomo en casetes largos 21-mer y muestra la información de secuencia en los epítomos de linfocitos

T usados. La figura describe las SEQ ID NOs: 128-129, 132, 131, 130, 157-159, 133, 160-171, respectivamente, en orden de aparición.

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 epítomos del MHC I contenidos en su secuencia natural 25-mer (enlazador = secuencias naturales flanqueantes), compuesto por 6 epítomos de primates no humanos (NHP), 5 epítomos humanos, 9 epítomos murinos, así como también 2 epítomos del MHC clase II universales.

La FIG. 20B ilustra el diseño de casete final para estudios preclínicos para permitir IND y muestra la información de secuencia para los epítomos de linfocitos T usados que se presentan en MHC clase I de origen primates no humanos (SEQ ID NOs: 172-177, respectivamente, en orden de aparición), ratón (SEQ ID NOs: 57-58 y 178-184, respectivamente, en orden de aparición) y humano (SEQ ID NOS 130-132 y 128-129, respectivamente, en orden de aparición), así como secuencias de 2 epítomos del MHC clase II universales PADRE y toxoide tetánico (SEQ ID NOS 49 and 47, respectivamente, en orden de aparición).

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).

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.

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.

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)

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.

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.

La FIG. 23 ilustra el esquema de producción de partícula viral.

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

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 µg por ratón, inyección intramuscular bilateral, encapsulado en MC3) en diversos puntos temporales.

La FIG. 26A ilustra respuestas de linfocitos T medidas 14 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 µg 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 bazo y ganglios linfáticos 14 días después de la inmunización. Se evaluaron las respuestas de linfocitos T específicos de SIINFEKL ("SIINFEKL" mostrada 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.

La FIG. 26B ilustra respuestas de linfocitos T medidas 14 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 µg 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 bazo y ganglios linfáticos 14 días después de la inmunización. Se evaluaron las respuestas de linfocitos T específicos de SIINFEKL ("SIINFEKL" mostrada 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.

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 bazo y ganglios linfáticos 14 días después de la inmunización con adenovirus.

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 bazo 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).

5 **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
10 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
15 respuestas de linfocitos T se midieron mediante tinción con pentámeros del MHC clase I . Los ratones se sacrificaron y se recolectaron los bazo y ganglios linfáticos 14 días después de la inmunización con adenovirus.

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
20 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-
25 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 bazo y ganglios linfáticos 14 días después
30 de la inmunización con adenovirus y 14 días después del refuerzo con srRNA (día 28 después de la sensibilización).

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 bazo y ganglios linfáticos 12 días después de la inmunización con adenovirus.

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 bazo 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).

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 epítipo del MHC clase I SIINFEKL (OVA) (SEQ ID NO: 57) se midieron en ratones hembra C57BL/6J y al epítipo 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.

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) por 10^6 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.

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.

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.

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.

La FIG. 34 ilustra respuestas inmunitarias celulares en macacos Rhesus indios después de una inmunización de sensibilización/refuerzo heterólogos. Se midió la producción de IFN-gamma específico de antígeno en PBMC a seis epítomos con restricción de mamu A01 diferentes para el grupo de sensibilización/refuerzo con ChAdV68.5WTnt.MAG25mer/srRNA de VEE-MAG25mer heterólogos (6 macacos Rhesus)

mediante el uso de ELISpot 7, 14, 21, 28 o 35 días después de la inmunización de sensibilización inicial y 7 días después de la primera inmunización de refuerzo. Los resultados se presentan como media de células formadoras de manchas (SFC) por 10^6 PBMC para cada epítipo en un formato de gráfico de barras apiladas.

5 **La FIG. 35** ilustra respuestas inmunitarias celulares en macacos Rhesus indios después de una inmunización de ChAdV con o sin anti-CTLA4. Se midió la producción de IFN-gamma específico de antígeno en PBMC a seis epítopos con restricción de mamu A01 diferentes después de la inmunización con ChAdV68.5WTnt.MAG25mer con o sin la adición de anti-CTLA4 administrado por vía intravenosa (IV) o localmente (SC) (6 macacos Rhesus
10 por grupo) mediante el uso de ELISpot 14 después de la inmunización inicial. Los resultados se presentan como media de células formadoras de manchas (SFC) por 10^6 PBMC para cada epítipo en un formato de gráfico de barras apiladas.

DESCRIPCIÓN DETALLADA

I. Definiciones

15 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.

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

 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 postraducciona específica de
25 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
30 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 al., 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.

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.

- 5 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.

- 10 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.

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

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

- 15 Como se usa en la presente, el término «ORF» significa marco de lectura abierto.

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.

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.

- 20 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.

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.

- 25 Como se usa en la presente, el término «indel» es una inserción o eliminación de uno o más un ácidos nucleicos.

- 30 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.

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).

Las 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. Natl. 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 al., infra).

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 al., 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.

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

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

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.

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.

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.

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.

5 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.

10 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.

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.

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.

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

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.

20 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.

25 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.

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.

30 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.

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.

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.

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).

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.

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.

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.

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).

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.

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.

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.

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.

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.

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.

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.

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.

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.

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.

El término «lípidos» 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.

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 esteroides (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.

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.

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.

Se entenderá que cualesquiera términos no definidos directamente en la presente
 5 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
 10 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
 15 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.

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
 20 esta referencia a todos los efectos.

II. Métodos para indentificar neoantígenos

En la presente se describen métodos para identificar neoantígenos de un tumor de un sujeto que probablemente se presenten en la superficie celular y/o probablemente sean inmunógenos. A modo de ejemplo, uno de estos métodos puede comprender las
 25 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
 30 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.

5 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
10 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
15 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
20 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.

El modelo de presentación puede comprender un conjunto de rasgos que proviene
25 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.

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
30 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.

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.

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.

Una cantidad del conjunto de neoantígenos seleccionados puede ser 20.

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.

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.

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

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

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.

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.

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.

- 5 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.

- 10 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
15 MHC derivados de las múltiples muestras.

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

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

- 20 Las muestras también pueden incluir líneas celulares humanas obtenidas o derivadas de múltiples pacientes.

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

- 25 Las muestras también pueden incluir muestras de tejido frescas o congeladas obtenidas de múltiples pacientes.

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

- 30 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.

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.

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.

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.

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

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

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.

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.

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

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

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.

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

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.

Un método para tratar a un sujeto con un tumor, que comprende realizar los pasos de la reivindicación 1 y que además comprende obtener una vacuna tumoral que

comprende el conjunto de neoantígenos seleccionados y administrarle la vacuna tumoral al sujeto.

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.

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.

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.

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

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

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

10 La vacuna tumoral puede incluir un adyuvante.

La vacuna tumoral puede incluir un excipiente.

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.

15 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.

20 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).

25 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.

30 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.

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

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

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

5 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

10 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,

15 HLA-C, HLA-DQ, HLA-DR, HLA-DP) en otros sujetos distintos.

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

20 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»)

25 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

30 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

- 5 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
- 10 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
- 15 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,
- 20 CALR, CNX, ERP57, HLA-DM, HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOBHLA-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
- 25 maquinaria 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,
- 30 HLA-DMA, HLA-DMB, HLA-DO, HLA-DOA, HLA-DOBHLA-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.

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.

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.

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.

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 1 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 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

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.

5 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.

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

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

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

15 Las muestras también pueden incluir muestras de tumores frescas o congeladas obtenidas de múltiples pacientes.

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

20 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.

25 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.

30 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.

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

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.

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

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

10 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.

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.

15 El conjunto de datos de entrenamiento también puede incluir datos asociados con transcriptomas asociados con las muestras,

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

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

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.

25 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.

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

30 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.

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.

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.

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.

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).

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.

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.

5 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).

10 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., edición actual); Sambrook, et al., *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).

20 **III. Identificación de mutaciones específicas de tumores en neoantígenos**

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.

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:

30 (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.

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.

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).

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.

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.

5 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.º 4.656.127). De acuerdo con el método, un cebador complementario a la secuencia alélica
10 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
15 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
20 de grandes cantidades de datos externos de secuencias.

Un método basado en solución se puede usar para determinar la identidad de un nucleótido de un sitio polimórfico. Cohen, D. et al. (Patente francesa 2.650.840; solicitud PCT n.º WO91/02087). Al igual que en el método de Mundy de la patente estadounidense n.º 4.656.127, se emplea un cebador que es complementario a las secuencias alélicas
25 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.

Un método alternativo, conocido como Genetic Bit Analysis o GBA se describe por
30 Goelet, P. et al. (Solicitud PCT n.º 92/15712). El método de Goelet, P. et al. 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 al.

(Patente francesa 2.650.840; solicitud PCT n.º WO91/02087) el método de Goelet, P. et al. Puede ser un ensayo de fase heterogénea, en el cual el cebador o la molécula objetivo se inmoviliza en una fase sólida.

Se han descrito varios procedimientos de incorporación de nucleótidos guiados por
 5 cebadores para el ensayo de sitios polimórficos en el ADN (Komher, J. S. et al., Nucl Acids. Res. 17:7779-7784 (1989); Sokolov, B. P., Nucl. Acids Res. 18:3671 (1990); Syvanen, A.-C., et al., Genomics 8:684-692 (1990); Kuppuswamy, M. N. et al., Proc. Natl. Acad. Sci. (EUA) 88:1143-1147 (1991); Prezant, T. R. et ál., Hum. Mutat. 1:159-164 (1992); Ugozzoli, L. et al., GATA 9:107-112 (1992); Nyren, P. et al., Anal. Biochem.
 10 208:171-175 (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
 15 longitud de la ejecución (Syvanen, A.-C, et al., Amer. J. Hum. Genet. 52:46-59 (1993)).

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
 20 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
 25 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,
 30 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.

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: 5 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 10 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 15 (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.

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 20 avidina-biotina como se describe en, por ejemplo, la solicitud de patente estadounidense n.º 2006/0252077) puede unirse a cada fragmento que se capturará en una superficie recubierta con un segundo miembro respectivo de ese par de acoplamiento.

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 25 en la patente estadounidense n.º 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 30 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.

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.

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.

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.

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

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.

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.

- 5 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 1 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
10 motivos de secuencia que promueven el transporte de TAP

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

 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.

- 15 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.

- 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,
20 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,
25 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
30 moléculas amino, y cualquier intervalo derivado de estos. En modalidades específicas, las moléculas de péptidos neoantigénicos son iguales o menores a 50 aminoácidos.

 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, 15-24 residuos.

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 neoepítomos 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.

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.

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

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.

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.

- 5
 - 10
 - 15
- 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).

- 20
 - 25
 - 30
- 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 al., 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.

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 epítipo 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.

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.

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.

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.,

- 5 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 al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- 10
- 15

V. Composiciones de vacuna

- 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.
- 20

- 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
- 25
 - 30

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, 5 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.

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 10 pueden asociarse con diferentes moléculas de MHC, como una molécula de clase I de MHC diferente. En algunos aspectos, una composición de vacuna comprende una secuencia codificante de péptidos y/o polipéptidos que pueden asociarse con las moléculas de clase I de MHC que se producen más frecuentemente. Por lo tanto, las composiciones de vacuna pueden comprender diferentes fragmentos que pueden 15 asociarse con al menos 2 moléculas de clase I de MHC preferidas, al menos 3 preferidas o al menos 4 preferidas.

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

Una composición de vacuna también puede comprender un adyuvante y/o un 20 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.

Los adyuvantes son cualquier sustancia cuya mezcla en una composición de 25 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.

La capacidad de un adyuvante de aumentar una respuesta inmunitaria a un 30 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.

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, JuvImmune, 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 Ribí. Quil o Superfos de Ribí. 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 al., 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.º 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 al., J Immunother Emphasis Tumor Immunol. 1996 (6):414-418).

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.

Otros ejemplos de adyuvantes útiles incluyen, de modo no taxativo, CpG modificados químicamente (p. ej., CpR, Idera), Poly(I:C)(p. ej., polyi:Cl2U), 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).

Una composición de vacuna puede comprender más de un adyuvante diferente.

- 5 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.

- 10 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
15 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
20 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.

- 25 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
30 MHC correspondiente. Por lo tanto, en algunas modalidades, una composición de vacuna contiene además al menos una célula presentadora de antígenos.

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 al., 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 al., Immunization Delivered by Lentiviral Vectors for Cancer and Infectious Diseases, *Immunol Rev.* (2011) 239(1): 45-61, Sakuma et al., Lentiviral vectors: basic to translational, *Biochem J.* (2012) 443(3):603-18, Cooper et al., 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 al., 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 al., Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, *Nat Med.* (2016) 22 (4):433-8, Stronen et al., Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, *Science.* (2016) 352 (6291):1337-41, Lu et al., 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 al. (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

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.

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 al, 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 al., Nucl. Acids Res., 11(23):8287 (1983)]. Un experto en la técnica puede seleccionar otros promotores deseables o adecuados.

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 (poly-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 al., "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.

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.

- 5 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.

V.B. Puntos de control inmunitarios

- 10 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
15 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.

- 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,
20 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-
25 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),
30 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 al., 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

Los péptidos trancos, es decir los presentados por todos o la mayoría de los subclones tumorales, pueden priorizarse para incluirse en la vacuna.⁵³ Opcionalmente, si no hay péptidos trancos que se predice que se presenten y sean inmunógenos con alta probabilidad, o si la cantidad de péptidos trancos que se predice que se presentará y será inmunógena con alta probabilidad es lo suficientemente pequeña que se pueden incluir péptidos no trancos 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

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)

5

V.D. Alfavirus

V.D.1. Biología del alfavirus

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 Microbial 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).

Un ciclo de vida modelo de un alfavirus implica varias etapas diferentes (Strauss Microbial 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.

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

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.

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.

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*

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

5 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
10 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

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
15 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
20 casetes del neoantígeno.

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
25 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.

Una alternativa al suministro de genes mediado por partículas es el uso de nanomateriales para suministrar vectores de expresión (Riley 2017). Cabe destacar que
30 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.

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.

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.

- 5 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

- 10 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
- 15 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.

- 20 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
- 25 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.

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

- 30 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.

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.

Incluso otro aspecto proporciona un método para provocar una respuesta inmunitaria en un hospedador mamífero para tratar un cáncer. El método puede

- 5 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.

- 10 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.

- 15 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
- 20 que consiste en E1A, E1B, E2A, E2B, E3, E4, L1, L2, L3, L4 y L5 de la SEQ ID NO: 1.

- También se describe un 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
- 25 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
- 30 ID NO: 1. En algunos aspectos, el vector puede carecer del gen E1A y/o E1B.

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.

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.

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

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.

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.

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.

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.º de acceso de ATCC CCL 2], A549 [n.º de acceso de ATCC CCL 185], KB [CCL 17], Detroit [p. ej., Detroit 510, CCL 72] y WI-38 [CCL 75] 5 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.

Una línea celular que expresa E1 puede ser útil en la generación de vectores 10 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 15 genes E1 Ad también son útiles para generar Ad recombinantes de chimpancé.

V.E.3. Partículas virales recombinantes como vectores

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 20 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 25 envasado opcionales pueden suministrarle al vector viral de chimpancé cualquier producto necesario de genes adenovirales eliminados.

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 30 funcionales de expresión génica. Si se desea, puede retirarse toda la región del gen.

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

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.º 5.240.846].

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

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

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.

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.

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.

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.

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.

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.

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

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.

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.

Para C68, el virus «auxiliar» puede ser un fragmento formado mediante el recorte del extremo terminal C del genoma C68 con Sspl, 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.

Los virus auxiliares también pueden formarse en conjugados policatiónicos, como se describe en Wu et al., 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

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.

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.

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

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.

Los vectores recombinantes descritos anteriormente se administran a humanos de acuerdo con los métodos publicados para terapia génica. Se le puede administrar a un
 5 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
 10 conocidas por los expertos en la técnica y que se sepa que son portadores farmacéuticamente aceptables.

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
 15 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
 20 parenteral. Si se desea, las vías de administración se pueden combinar.

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
 25 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.

Los adenovirus recombinantes y con deficiencia de replicación pueden administrarse en una «cantidad farmacéuticamente eficaz», es decir, una cantidad de
 30 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.

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.

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

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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 estero, 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 al., *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.

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.

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 al., *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

5 incluir vectores virales, vectores de ARNm y vectores de ADN con o sin electroporación.

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

10 estadounidense n.º 5.279.833 Rose, la patente estadounidense n.º 5.279.833; 9106309WOAWO 91/06309; y en Felgner et al., *Proc. Natl. Acad. Sci. EUA* 84: 7413-7414 (1987).

También pueden incluir los neoantígenos en plataformas de vacuna basadas en vectores virales, como vaccinia, viruela aviar, alfavirus de autorreplicación, marabavirus,

15 adenovirus (remitirse, p. ej., a Tatsis et al., *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 al., *Immunization Delivered by Lentiviral Vectors for Cancer and*

20 *Infectious Diseases, Immunol Rev.* (2011) 239(1): 45-61, Sakuma et al., *Lentiviral vectors: basic to translational, Biochem J.* (2012) 443(3):603-18, Cooper et al., *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 al., *Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery, J. Virol.* (1998)

25 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

30 dirigidas al compartimiento subcelular (remitirse, p. ej., a Gros et al., *Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients, Nat Med.* (2016) 22 (4):433-8, Stronen et al., *Targeting of cancer neoantigens with donor-derived T cell receptor repertoires, Science.* (2016) 352 (6291):1337-41, Lu et al., *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 al. (*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.

Un medio de administración de ácidos nucleicos utiliza construcciones de minigenes que codifican uno o múltiples epítomos. Para crear una secuencia de ADN que codifique los epítomos de CTL selectos (minigenes) para expresión en células humanas, las secuencias de aminoácidos de los epítomos 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 epítomos 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, epítomos, una secuencia líder (señal) y una señal de retención del retículo endoplásmico. Adicionalmente, se puede mejorar la presentación de epítomos de CTL en el MHC incluyendo secuencias flanqueantes de origen natural o sintéticas (p. ej., poli-alanina) adyacentes a los epítomos de CTL. Las secuencias 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.

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.

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.

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.

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

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.

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 1×10^{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 μg de ARN, en particular 0,1 o 1 μg ; o es posible utilizar una o más inyecciones de srRNA a una dosis de vacunación elevada, seleccionada del intervalo de 1 a 100 μg de ARN, en particular 10 o 100 μg .

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 1×10^{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 μg de ARN, en particular 0,1 o 1 μg ; o es posible utilizar una o más inyecciones de srRNA a una dosis de vacunación elevada, seleccionada del intervalo de 1 a 100 μg de ARN, en particular 10 o 100 μg .

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.

- En determinados casos, se puede utilizar un anticuerpo anti-PD-L1 como
- 5 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.

- 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
- 10 eficacia, entre otros parámetros.

- 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
- 15 8 semanas).

- 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
- 20 MHC o ensayos de citotoxicidad. Es posible controlar las respuestas de linfocitos T a los epítomos 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 epítomos codificados en las vacunas a partir de las PBMC, midiendo la inducción de citocinas capturadas a nivel
 - 25 intracelular o extracelular, como IFN-gamma, utilizando un ensayo de citometría de flujo. Las respuestas específicas de linfocitos T CD4 o CD8 a los epítomos 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 epítomos/complejos MHC de clase I, mediante tinción multimérica de MHC. Las respuestas específicas de
 - 30 linfocitos T CD4 o CD8 a los epítomos 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 epítomos codificados

en las vacunas pueden evaluarse de forma funcional a través de un ensayo de liberación de cromó o ensayos colorimétricos alternativos de citotoxicidad.

VIII. Identificación de neoantígenos

VIII.A. Identificación de candidatos neoantigénicos

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

Las mejoras de los procesos que se establecen en la presente abordan dificultades 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 objetivo¹⁶ 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 QC¹⁷

- 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 estándar¹⁸

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 sondas¹⁹, 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

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 individuales²⁰ aunando los resultados de diferentes programas⁵

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 pureza²³.

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.

5 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:

10 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 base²⁷.

15 c. Eliminación de variantes derivadas de distorsiones recurrentes en la secuenciación, incluso en caso de no observarse en la forma normal correspondiente²⁷. Los ejemplos incluyen variantes que se detectan principalmente en una cadena.

20 d. Eliminación de variantes detectadas en un conjunto de controles sin relación²⁷

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 exomas²⁸. Posibles optimizaciones adicionales incluyen implementación de un ensayo específico de tipificación de HLA, como
25 secuenciación de ADN de lectura prolongada³⁰, o adaptación de un método de unión de fragmentos de ARN para mantener la continuidad³¹.

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
30 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 SpliceR³⁶ y MAMBA³⁷, con reintroducción de secuencias mutantes. La expresión génica se determina con una herramienta como Cufflinks³⁵ 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 ASE³⁸ o HTSeq³⁹. Las posibles etapas de filtración incluyen:

- 10 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).
- 15 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:
 - 20 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 uveal⁴⁰, el segundo una línea celular de melanoma uveal⁴¹ y el tercero pacientes con cáncer de mama⁴².
 - 25 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.
 - 30 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 GTEx⁴³ (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).

En las muestras con ARN poliadenilado, se evalúa la presencia de ARN viral y microbiano en los datos de secuenciación de ARN utilizando RNA CoMPASS⁴⁴ o un método similar, para identificar factores adicionales que puedan pronosticar la respuesta del paciente.

VIII.B. Aislamiento y detección de péptidos de HLA

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.

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)

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.

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.

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).

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

Utilizando el péptido YVYVADVAAK (SEQ ID NO: 59), 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

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.

- 5 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. Esto resulta útil en diversos contextos.
- 10 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
- 15 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.

- El sistema de identificación de presentación 160 determina las probabilidades de
- 20 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 presenten 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-B*07:02, HLA-
- 25 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
- 30 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.

IX.B. Información de presentación

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

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. 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.

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 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: 61), HROEIFSHDFJ (SEQ ID

NO: 62), FJIEJFOESS (SEQ ID NO: 63), NEIOREIREI (SEQ ID NO: 64), JFKSIFEMMSJDSSU (SEQ ID NO: 65), and KNFLENFIESOFI (SEQ ID NO: 66), 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

5 espectrometría de masas. A diferencia de las líneas celulares de alelos individuales, 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.

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.

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. 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: 61) y el alelo HLA-A*01:01. El MHC presenta pocos péptidos con una IC50 > 1000 nm y los menores valores de IC50 aumentan la probabilidad de presentación.

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 HLA-A*01:01.

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.

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.

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 postraduccion específicas en el péptido codificado por el neoantígeno. La presencia de motivos de cinasa afecta la probabilidad de modificación postraduccion, lo cual puede potenciar la unión del MHC o interferir con ella.

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

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 proteómica de espectrometría de masas u otros medios.

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.

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.

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 11.

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

También se puede realizar un modelo de cualquier información sobre ausencia de
5 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

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
10 proteica original. 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
15 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: 67) del péptido
presentado FJIEJFOESS (SEQ ID NO: 63) que se identifica a partir de la proteína original
20 del péptido.

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
25 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*
30 *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).

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

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.

La información de no interacción 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.

La información de no interacción 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.

La información de no interacción 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.

La información de no interacción 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.

La información de no interacción 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 al., Science 2015.

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.

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.

La información de no interacción 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.

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.

La información de no interacción 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).

La información de no interacción con alelos puede incluir también la probabilidad de que el péptido no sea detectado o se sobrerrepresente mediante espectrometría de masas debido a inclinaciones técnicas.

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).

5 La información de no interacción 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.

10 La información de no interacción 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.

15 La información de no interacción 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). Los niveles de expresión de TAP más altos aumentan la probabilidad de presentación de todos los péptidos.

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

20 i. Mutaciones conductoras en genes conductores del cáncer conocidos tales como

EGFR, KRAS, ALK, RET, ROS1, TP53, CDKN2A, CDKN2B, NTRK1, NTRK2, NTRK3.

25 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-DOBHLA-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
30 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.

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-DOBHLA-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).

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

La información de no interacción 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 epítomos; la nomenclatura de sufijos de alelos HLA se describe en <https://www.ebi.ac.uk/ipd/imgt/hla/nomenclature/suffixes.html>.

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

La información de no interacción con alelos puede incluir también el historial de tabaquismo.

La información de no interacción con alelos puede incluir también el historial de quemaduras de sol, la exposición solar o la exposición a otros mutágenos.

La información de no interacción 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.

La información de no interacción 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.

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
 5 provoca traducción del RNAm disminuido, lo cual disminuye la probabilidad de presentación.

IX.C. Sistema de identificación de presentación

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
 10 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
 15 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

20 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 $\mathbf{z}^i$ que incluyen al menos una secuencia peptídica presentada o no presentada $\mathbf{p}^i$, uno o más alelos MHC asociados $\mathbf{a}^i$ que se asocian con la
 25 secuencia peptídica $\mathbf{p}^i$, 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.

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
 30 el péptido $\mathbf{p}^i$ se presentó mediante el uno o más alelos MHC asociados $\mathbf{a}^i$. Sin embargo, se aprecia que en otras implementaciones, la variable dependiente 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 $\mathbf{z}^i$. Por ejemplo, en otra implementación, la variante dependiente y^i también puede ser un valor numérico que indica

la corriente de iones de espectrometría de masas identificada para la instancia de datos.

La secuencia peptídica p^i para la instancia de datos 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 9-30 para la clase II de MHC. En una implementación específica del sistema 160, todas las secuencia peptídicas p^i en un conjunto de datos de entrenamiento pueden tener la misma longitud, por ejemplo, 9. 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 a^i para la instancia de datos i indican qué alelos de MHC estaban presentes en asociación con la secuencia peptídica correspondiente p^i .

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

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

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.

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.

La figura 4 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: 68), FIEUHFWI (SEQ ID NO: 69) y FEWRHRJTRUJR (SEQ ID NO: 70). 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*01:01 y una secuencia peptídica QIEJOEIJ (SEQ ID NO: 71). La primera instancia de datos indica que la secuencia peptídica QCEIOWARE (SEQ ID NO: 68) 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: 72) y una medición de cuantificación de RNAm de 10² FPKM. La cuarta instancia de datos indica que la secuencia peptídica QIEJOEIJ (SEQ ID NO: 71) no fue presentada mediante uno de los alelos HLA-B*07:02, HLA-C*01:03 o HLA-A*01:01. Los datos de entrenamiento 170A incluyen también las predicciones de afinidad de unión y predicciones de estabilidad para cada uno de los alelos, así como también la secuencia flanqueante en C del péptido y la medición de cuantificación de RNAm para el péptido.

IX.C.2. Módulo de codificación

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

[illegible]

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.

El módulo de codificación 314 codifica también los uno o más alelos de MHC $\mathbf{a}^i$ para la instancia de datos i como un vector de fila de elementos m , en el que cada elemento $h=1, 2, \dots, m$ 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 $\mathbf{a}^i=[0 \ 0 \ 1 \ 1]$, en el 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 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 $\mathbf{p}_i$.

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 $\mathbf{x}^i$ se presentó por medio de uno de los alelos de MHC asociados $\mathbf{a}^i$, y un valor de 0 indica que el péptido $\mathbf{x}^i$ no fue presentado por ninguno de los alelos de MHC asociado $\mathbf{a}^i$. 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 $[-\infty, \infty]$ para valores actuales de iones entre $[0, \infty]$.

El módulo de codificación 314 puede representar un par de variables que interactúan con alelos $\mathbf{x}_h^i$ f para el péptido $\mathbf{p}_i$ 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 $\mathbf{x}_h^i$ como un vector de fila igual a $[\mathbf{p}^i]$, $[\mathbf{p}^i \ b_h^i]$, $[\mathbf{p}^i \ s_h^i]$, o $[\mathbf{p}^i \ b_h^i \ s_h^i]$, donde b_h^i es la predicción de afinidad de unión para el péptido $\mathbf{p}_i$ 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).

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 $\mathbf{x}_h^i$.

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 $\mathbf{x}_h^i$,

- 5 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 $\mathbf{x}_h^i$.

En un caso, el módulo de codificación 314 representa la longitud de péptidos como un vector $\mathbf{T}_k = [\mathbb{1}(L_k=8) \mathbb{1}(L_k=9) \mathbb{1}(L_k=10) \mathbb{1}(L_k=11) \mathbb{1}(L_k=12) \mathbb{1}(L_k=13) \mathbb{1}(L_k=14) \mathbb{1}(L_k=15)]$

- 10 donde $\mathbb{1}$ es la función de indicador, y L_k indica la longitud del péptido $\mathbf{p}_k$. El vector $\mathbf{T}_k$ puede incluirse en las variables que interactúan con alelos $\mathbf{x}_h^i$.

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 $\mathbf{x}_h^i$.

- 15 De manera similar, el módulo de codificación 314 puede representar las variables que no interactúan con alelos $\mathbf{w}^i$ 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, $\mathbf{w}^i$ puede ser un vector de fila igual a $[\mathbf{c}^i]$ o $[\mathbf{c}^i \mathbf{m}^i \mathbf{w}^i]$ en el que $\mathbf{w}^i$ 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 $\mathbf{p}^i$ y la medición de cuantificación del RNAm $\mathbf{m}^i$ 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).
- 20

- 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 $\mathbf{w}^i$.
- 25

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 $\mathbf{w}^i$.

- 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 $\mathbf{w}^i$.
- 30

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 .

5 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 .

10 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 .

15 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 .

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 .

20 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 .

25 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 .

30 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 .

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 .

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.

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 .

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.

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.

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 p^k en una muestra con muestra de tipo tumoral, no solo se incluye el nivel de expresión de transcripción o génica medido del gen o transcripción de origen del péptido p^k en las variables que no interactúan con alelos w^i , sino también la expresión de transcripción o génica media y/o mediana del gen o transcripción de origen del péptido p^k en melanomas según lo medido por TCGA.

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 w^i .

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 w^i . En otro caso, el módulo de codificación 314 representa las anotaciones de nivel de residuos de la proteína fuente para el péptido p^k al incluir una variable indicadora que es igual a 1 si el péptido p^k se superpone con un motivo de hélice y 0 de otro modo, o que es igual a 1 si el proteína p^k se encuentra completamente incluida dentro de un motivo de hélice en las variables que no interactúan con alelos w^i . En otro caso, una característica que representa la proporción de residuos en el péptido p^k que se encuentra incluida dentro de una anotación de motivo de hélice puede incluirse en las variables que no interactúan con alelos w^i .

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

El módulo de codificación 314 puede representar también un conjunto general de variables z^i para el péptido 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 x^i y las variables que no interactúan con alelos w^i están concatenadas unas detrás de otras. Por ejemplo, el módulo de codificación 314 puede representar z_h^i como un vector de fila igual a $[x_h^i w^i]$ o $[w_i x_h^i]$.

X. Módulo de entrenamiento

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

X.A. Visión general

5 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
10 independientes y variables dependientes en los datos de entrenamiento 170 de modo que se minimiza una función de pérdida. Específicamente, la función de pérdida $\ell(y_{i \in S}, u_{i \in S}; \theta)$ 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
15 una implementación particular a la que se hace referencia a lo largo del resto de la memoria descriptiva, la función de pérdida $(y_{i \in S}, u_{i \in S}; \theta)$ es la función de probabilidad log negativa dada por la ecuación (1a) de la siguiente manera:

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

Sin embargo, en la práctica, se puede utilizar otra función de pérdida. Por ejemplo,
20 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:

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

El modelo de presentación puede ser un modelo paramétrico en el que uno o más
25 parámetros θ 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 $(y_{i \in S}, u_{i \in S}; \theta)$ se determinan a través de los algoritmos de optimización numéricos basados en gradientes tales como algoritmos de gradientes, algoritmos de gradientes estocásticas y
30 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

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.

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

$$u_k^h = \Pr(\mathbf{p}^k \text{ presentados; alelo de MHC } h) = f(g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h)), \quad (2)$$

donde la secuencia peptídica $\mathbf{x}_h^k$ indica las variables que interactúan con alelos codificadas para el péptido $\mathbf{p}^k$ y el alelo de MHC correspondiente h , $f(\cdot)$ 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 $\mathbf{x}_h^k$ 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 .

El resultado de la función de dependencia $g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h)$ 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 $\mathbf{x}_h^k$, y en particular, de acuerdo con las posiciones de aminoácidos de la secuencia peptídica del péptido $\mathbf{p}^k$. 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 $\mathbf{p}^k$ y puede tener un valor bajo si la presentación no es probable. La función de transformación $f(\cdot)$ transforma el resultado, y más específicamente, transforma la puntuación de dependencia generada por $g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h)$ en este caso, en un valor adecuado para indicar la probabilidad de que el péptido $\mathbf{p}^k$ será presentado por un alelo de MHC.

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

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

Como ejemplo adicional, $f(\cdot)$ 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 $[0, 1]$, $f(\cdot)$ puede ser cualquier función tal como la función de identidad, la función exponencial, la función log y similares.

Por lo tanto, 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 de dependencia $g_h(\cdot)$ para el alelo de MHC h a la versión codificada de la secuencia peptídica 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 p^k se presentará por el alelo de MHC h .

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

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:

$$g_h(x_h^i; \theta_h) = x_h^i \cdot \theta_h. \quad (6)$$

que combina de forma lineal cada variable que interactúa con alelos en x_h^k con un parámetro correspondiente en el conjunto de parámetros θ_h determinados por el alelo de MHC asociado h .

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:

$$g_h(x_h^i; \theta_h) = NN_h(x_h^i; \theta_h). \quad (7)$$

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 θ_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 modelos 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.

5 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.

10 En un caso al que se hace referencia en todo lo que resta de la memoria descriptiva, cada alelo de MHC en $h=1, 2, \dots, m$ 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 .

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 $\theta_3(1), \theta_3(2), \dots, \theta_3(10)$. 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 $NN_3(x_3^k)$.

En otro caso, los alelos de MHC identificados $h=1, 2, \dots, m$ 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 θ_h puede corresponder a un conjunto de parámetros para el modelo de red único, y por lo tanto, el conjunto de parámetros θ_h se puede compartir por todos los alelos de MHC.

La figura 6A ilustra un modelo de red de ejemplo $NN_H(\cdot)$ compartido por los alelos de MHC $h=1, 2, \dots, m$. 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 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 $h=3$.

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 x_h^k y la secuencia de proteínas codificada d_h de un alelo de MHC h . En dicho caso, el conjunto

de parámetros θ_h puede corresponder de nuevo a un conjunto de parámetros para el modelo de red único, y por lo tanto, el conjunto de parámetros θ_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 $[x_h^k d_h]$ 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.

La figura 6B 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 $h=3$.

En incluso otro caso, la función de dependencia $g_h(\cdot)$ se puede expresar como:

$$g_h(x_h^k; \theta_h) = g'_h(x_h^k; \theta'_h) + \theta_h^0$$

donde $g'_h(x_h^k; \theta'_h)$ es la función afín con un conjunto de parámetros θ'_h , la función de red o similares, con un parámetro de inclinación θ_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 .

En otra implementación, el parámetro de inclinación θ_h^0 puede compartirse de acuerdo con la familia génica del alelo de MHC h . Es decir, el parámetro de inclinación θ_h^0 para el alelo de MHC h puede ser igual a $\theta_{gen(h)}^0$, donde el $gen(h)$ es la familia génica del alelo de MHC h . Por ejemplo, los alelos de MHC 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 θ_h^0 para cada uno de estos alelos de MHC.

Volviendo a la ecuación (2), como ejemplo, la probabilidad de que el péptido 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:

$$u_k^3 = f(x_3^k \cdot \theta_3),$$

donde x_3^k son las variables que interactúan con alelos identificadas para el alelo de MHC $h=3$, y θ_3 son el conjunto de parámetros determinados por el alelo de MHC $h=3$ a través de la minimización de función de pérdida.

Como ejemplo adicional, la probabilidad de que el péptido 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:

$$u_k^3 = f(NN_3(\mathbf{x}_3^k; \boldsymbol{\theta}_3)),$$

donde $\mathbf{x}_3^k$ 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 $h=3$.

La figura 7 ilustra la generación de una probabilidad de presentación para el péptido $\mathbf{p}^k$ 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 $\mathbf{x}_3^k$ para el alelo de MHC $h=3$ y genera el resultado $NN_3(\mathbf{x}_3^k)$. El resultado es mapeado por la función $f(\cdot)$ para generar la probabilidad de presentación estimada u_k .

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

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 $\mathbf{p}^k$ mediante:

$$u_k^h = \Pr(\mathbf{p}^k \text{ presentados}) = f(g_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + g_h(\mathbf{x}_h^i; \boldsymbol{\theta}_h)), \quad (8)$$

donde $\mathbf{w}^k$ indica las variables que no interactúan con alelos codificadas para el péptido $\mathbf{p}^k$, $g_w(\cdot)$ es una función para las variables que no interactúan con alelos $\mathbf{w}^k$ 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.

El resultado de la función de dependencia $g_w(\mathbf{w}^k; \boldsymbol{\theta}_w)$ representa una puntuación de dependencia para las variables que no interactúan con alelos que indica si el péptido $\mathbf{p}^k$ 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 $\mathbf{p}^k$ se asocia con una secuencia flanqueante de extremo C que es conocida por impactar positivamente la presentación del péptido $\mathbf{p}^k$ y puede tener un valor bajo si el péptido $\mathbf{p}^k$ se asocia con una secuencia flanqueante de extremo C que es conocida por impactar negativamente la presentación del péptido $\mathbf{p}^k$.

De acuerdo con la ecuación (8), la probabilidad por alelo de que una secuencia peptídica $\mathbf{p}^k$ 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 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 p^k será presentada por el alelo de MHC h .

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

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

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

alelos

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 w^k .

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

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

que combinan de forma lineal las variables que no interactúan con alelos en w^k con un parámetro correspondiente en el conjunto de parámetros θ_w .

La función de dependencia $g_w(\cdot)$ puede ser también una función de red dada por:

$$g_w(w^k; \theta_w) = NN_w(w^k; \theta_w).$$

representada por un modelo de red $NN_w(\cdot)$ que tiene un parámetro asociado en el conjunto de parámetros θ_w .

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

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

Donde $g'_w(w^k; \theta'_w)$ es la función afín, la función de red con el conjunto de parámetros que no interactúan con alelos θ'_w , o similares, m^k es la medición de cuantificación de RNAm

para el péptido $\mathbf{p}^k$, $h(\cdot)$ 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 cuantificación 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.

En incluso otro caso, la función de dependencia $\mathbf{g}_w(\cdot)$ para las variables que no interactúan con alelos puede estar dada por:

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

donde $g'_w(\mathbf{w}^k; \theta'_w)$ es la función afín, la función de red con el conjunto de parámetros que no interactúan con alelos θ'_w , o similares, $\mathbf{o}^k$ es el vector indicador descrito anteriormente que representa proteínas e isoformas en el proteoma humano para el péptido $\mathbf{p}^k$, y θ_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 $\mathbf{o}^k$ y el conjunto de parámetros θ_w^o son considerablemente altos, un término de regularización de parámetros, tal como $\lambda \cdot \|\theta_w^o\|$, 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.

Volviendo a la ecuación (8), como ejemplo, la probabilidad de que el péptido $\mathbf{p}^k$ 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:

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

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

Como ejemplo adicional, la probabilidad de que el péptido $\mathbf{p}^k$ 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:

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

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

con alelos.

La figura 8 ilustra la generación de una probabilidad de presentación para el péptido 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 x_3^k para el alelo de MHC $h=3$ y genera el resultado $NN_3(x_3^k)$. El modelo de red $NN_w(\cdot)$ recibe las variables que no interactúan con alelos w^k para el péptido p^k y genera el resultado $NN_w(w^k)$. 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

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

En una implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido 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 .

$$u_k = \Pr(p^k \text{ presentados; alelos } H) = \max(u_k^{h \in H}). \quad (12)$$

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

En una implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido p^k mediante:

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

donde los elementos a_h^k son 1 para los múltiples alelos de MHC H asociados con la secuencia peptídica p^k y x_h^k indica las variables que interactúan con alelos codificadas para el péptido p^k y los alelos de MHC correspondientes. Los valores para el conjunto de parámetros θ_h para cada alelo de MHC h se pueden determinar al minimizar la función de pérdida con respecto a θ_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 g_h puede ser en forma de cualquiera de las funciones de dependencia g_h introducidas anteriormente en las secciones X.B.1.

De acuerdo con la ecuación (13), la probabilidad de presentación de que una secuencia peptídica 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 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 p^k será presentada por el conjunto de alelos de MHC H .

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 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 p^k .

Como ejemplo, la probabilidad de que el péptido 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:

$$u_k = f(x_2^k \cdot \theta_2 + x_3^k \cdot \theta_3),$$

donde x_2^k, x_3^k son las variables que interactúan con alelos identificadas para el alelo de MHC $h=2$, $h=3$, y θ_2, θ_3 son el conjunto de parámetros determinados por el alelo de MHC $h=2$, $h=3$.

Como ejemplo adicional, la probabilidad de que el péptido 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:

$$u_k = f(NN_2(x_2^k; \theta_2) + NN_3(x_3^k; \theta_3)),$$

donde $NN_2(\cdot)$, $NN_3(\cdot)$ son los modelos de red identificados para alelos de MHC $h=2$,

$h=3$, y θ_2, θ_3 son el conjunto de parámetros determinado por los alelos de MHC $h=2, h=3$.

La Figura 9 ilustra la generación de una probabilidad de presentación para el péptido 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 x_2^k para el alelo de MHC $h=2$ y genera el producto $NN_2(x_2^k)$ y el modelo de red $NN_3(\cdot)$ recibe las variables de interacción con el alelo x_3^k para el alelo de MHC $h=3$ y genera el producto $NN_3(x_3^k)$. 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

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 p^k mediante:

$$u_k = \Pr(p^k \text{ presentados}) = f\left(g_w(w^k; \theta_w) + \sum_{h=1}^m a_h^k \cdot g_h(x_h^k; \theta_h)\right), \quad (14)$$

donde w^k denota las variables codificadas de no interacción con alelos para el péptido p^k . Específicamente, los valores para el conjunto de parámetros θ_h para cada alelo de MHC h y el conjunto de parámetros θ_w para las variables de no interacción con alelos se pueden determinar al minimizar la función de pérdida con respecto a θ_h y θ_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 g_w puede encontrarse en forma de cualquiera de las funciones de dependencia g_w introducidas anteriormente en las secciones X.B.3.

Por tanto, de acuerdo con la ecuación (14), la probabilidad de presentación de que una secuencia peptídica 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 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 p^k va a presentar mediante los alelos de MHC H .

En el modelo de presentación de la ecuación (14), la cantidad de alelos asociados para cada péptido 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 p^k .

A modo de ejemplo, la probabilidad de que el péptido 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:

$$u_k = f\left(w^k \cdot \theta_w + x_2^k \cdot \theta_2 + x_3^k \cdot \theta_3\right),$$

donde w^k son las variables identificadas de no interacción con alelos para el péptido p^k y θ_w son el conjunto de parámetros determinado para las variables de no interacción con alelos.

A modo de otro ejemplo, la probabilidad de que el péptido 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:

$$u_k = f\left(NN_w(w^k; \theta_w) + NN_2(x_2^k; \theta_2) + NN_3(x_3^k; \theta_3)\right)$$

donde w^k son las variables de interacción con alelos para el péptido p^k y θ_w son el conjunto de parámetros determinado para las variables de interacción con alelos.

La Figura 10 ilustra la generación de una probabilidad de presentación para el péptido 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 x_2^k para el alelo de MHC $h=2$ y genera el producto $NN_2(x_2^k)$. El modelo de red $NN_3(\cdot)$ recibe las variables de interacción con alelos x_3^k para el alelo de MHC $h=3$ y genera el producto $NN_3(x_3^k)$. El modelo de red $NN_w(\cdot)$ recibe las variables de no interacción con alelos w^k para el péptido p^k y genera el producto $NN_w(w^k)$. Los productos se combinan y se mapean con la función $f(\cdot)$ para generar la probabilidad de presentación estimada u_k .

De manera alternativa, el módulo de entrenamiento 316 puede incluir variables de no interacción con alelos w^k en la predicción al agregar las variables de no interacción con alelos w^k a las variables de interacción con alelos x_h^k en la ecuación (15). Por tanto, la probabilidad de presentación se puede proporcionar con:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = f\left(\sum_{h=1}^m a_h^k \cdot g_h([x_h^k w^k]; \theta_h)\right). \quad (15)$$

X.C.4. Ejemplo 3.1: Modelos con probabilidades implícitas por alelo

En otra implementación, el módulo de entrenamiento 316 modela la probabilidad de presentación estimada u_k para el péptido $\mathbf{p}^k$ con:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = r\left(s\left(\mathbf{v} = \left[a_1^k \cdot u_k^1(\theta) \dots a_m^k \cdot u_k^m(\theta)\right]\right)\right), \quad (16)$$

donde los elementos a_h^k son 1 para los múltiples alelos de MHC $h \in H$ asociados a la secuencia peptídica $\mathbf{p}^k$, u_k^{h} es una probabilidad de presentación implícita por alelo para el alelo de MHC h , el vector $\mathbf{v}$ es un vector en el cual el elemento v_h corresponde a $a_h^k \cdot u_k^{h}$, $s(\cdot)$ es una función que mapea los elementos de $\mathbf{v}$ 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 θ para las probabilidades implícitas por alelo se pueden determinar al minimizar la función de pérdida con respecto a θ , 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.

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_k^{h} donde cada una corresponde a la probabilidad de que el péptido $\mathbf{p}^k$ 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 $\mathbf{p}^k$ será presentado por un conjunto de alelos de MHC H por completo, sino que también puede proporcionar probabilidades individuales $u_k^{h \in H}$ que indiquen qué alelo de MHC h tiene más probabilidad de haber presentado el péptido $\mathbf{p}^k$. 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.

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:

$$r(z) = \min(\max(z, 0), 1),$$

- 5 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.

10 **X.C.5. Ejemplo 3.2: Modelo de suma de funciones**

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:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = r\left(\sum_{h=1}^m a_h^k \cdot u_k^h(\boldsymbol{\theta})\right). \quad (17)$$

- 15 En una implementación, la probabilidad de presentación implícita por alelo para el alelo de MHC h se genera mediante:

$$u_k^h = f(g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h)), \quad (18)$$

de modo que la probabilidad de presentación se estima con:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = r\left(\sum_{h=1}^m a_h^k \cdot f(g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h))\right). \quad (19)$$

- 20 De acuerdo con la ecuación (19), la probabilidad de presentación de que una secuencia peptídica $\mathbf{p}^k$ 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 $\mathbf{p}^k$ para cada uno de los alelos de MHC H para generar la puntuación de dependencia correspondiente para las variables de interacción con alelos. Cada
- 25 puntuación de dependencia se transforma primero con la función $f(\cdot)$ para generar las probabilidades de presentación implícita por alelo u_k^h . Las probabilidades por alelo u_k^h 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 $\mathbf{p}^k$ se presente mediante un conjunto de alelos de MHC H . La
- 30 función de dependencia g_h puede encontrarse en forma de cualquiera de las funciones de

dependencia g_h introducidas anteriormente en las secciones X.B.1.

A modo de ejemplo, la probabilidad de que el péptido 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:

$$u_k = r\left(f\left(x_2^k \cdot \theta_2\right) + f\left(x_3^k \cdot \theta_3\right)\right),$$

donde x_2^k , x_3^k son las variables de interacción con alelos para los alelos de MHC $h=2$, $h=3$ y θ_2 , θ_3 son el conjunto de parámetros determinado para los alelos de MHC $h=2$, $h=3$.

A modo de otro ejemplo, la probabilidad de que el péptido 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:

$$u_k = r\left(f\left(NN_2(x_2^k; \theta_2)\right) + f\left(NN_3(x_3^k; \theta_3)\right)\right),$$

donde $NN_2(\cdot)$, $NN_3(\cdot)$ son los modelos de red identificados para los alelos de MHC $h=2$, $h=3$ y θ_2 , θ_3 son el conjunto de parámetros determinado para los alelos de MHC $h=2$, $h=3$.

La Figura 11 ilustra la generación de una probabilidad de presentación para el péptido 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 x_2^k para el alelo de MHC $h=2$ y genera el producto $NN_2(x_2^k)$ y el modelo de red $NN_3(\cdot)$ recibe las variables de interacción con el alelo x_3^k para el alelo de MHC $h=3$ y genera el producto $NN_3(x_3^k)$. Cada producto se mapea con la función $f(\cdot)$ y se combina para generar la probabilidad de presentación estimada u_k .

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

En una implementación, la probabilidad de presentación implícita por alelo para el aleo de MHC h se genera mediante:

$$u_k^h = f\left(g_h\left(x_h^k; \theta_h\right) + g_w\left(w^k; \theta_w\right)\right), \quad (20)$$

de modo que la probabilidad de presentación se genere con:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = r\left(\sum_{h=1}^m a_h^k \cdot f\left(g_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + g_h(\mathbf{x}_h^k; \boldsymbol{\theta}_h)\right)\right), \quad (21)$$

para incorporar el impacto de las variables de no interacción con alelos en la presentación del péptido.

De acuerdo con la ecuación (21), la probabilidad de presentación de que una secuencia peptídica $\mathbf{p}^k$ 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 $\mathbf{p}^k$ 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 combinan 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 $\mathbf{p}^k$ se presente mediante los alelos de MHC H . La función de dependencia g_w puede encontrarse en forma de cualquiera de las funciones de dependencia g_w introducidas anteriormente en las secciones X.B.3.

A modo de ejemplo, la probabilidad de que el péptido $\mathbf{p}^k$ 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:

$$u_k = r\left(f\left(\mathbf{w}^k \cdot \boldsymbol{\theta}_w + \mathbf{x}_2^k \cdot \boldsymbol{\theta}_2\right) + f\left(\mathbf{w}^k \cdot \boldsymbol{\theta}_w + \mathbf{x}_3^k \cdot \boldsymbol{\theta}_3\right)\right),$$

donde $\mathbf{w}^k$ son las variables identificadas de no interacción con alelos para el péptido $\mathbf{p}^k$ y $\boldsymbol{\theta}_w$ son el conjunto de parámetros determinado para las variables de no interacción con alelos.

A modo de otro ejemplo, la probabilidad de que el péptido $\mathbf{p}^k$ 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:

$$u_k = r\left(f\left(NN_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + NN_2(\mathbf{x}_2^k; \boldsymbol{\theta}_2)\right) + f\left(NN_w(\mathbf{w}^k; \boldsymbol{\theta}_w) + NN_3(\mathbf{x}_3^k; \boldsymbol{\theta}_3)\right)\right)$$

donde $\mathbf{w}^k$ son las variables de interacción con alelos para el péptido $\mathbf{p}^k$ y $\boldsymbol{\theta}_w$ son el

conjunto de parámetros determinado para las variables de interacción con alelos.

La Figura 12 ilustra la generación de una probabilidad de presentación para el péptido $\mathbf{p}^k$ 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 $\mathbf{x}_2^k$ para el alelo de MHC $h=2$ y genera el producto $NN_2(\mathbf{x}_2^k)$. El modelo de red $NN_w(\cdot)$ recibe las variables de no interacción con alelos $\mathbf{w}^k$ para el péptido $\mathbf{p}^k$ y genera el producto $NN_w(\mathbf{w}^k)$. 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 $\mathbf{x}_3^k$ para el alelo de MHC $h=3$ y genera el producto $NN_3(\mathbf{x}_3^k)$, que se combina nuevamente con el producto $NN_w(\mathbf{w}^k)$ 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 .

En otra implementación, la probabilidad de presentación implícita por alelo para el alelo de MHC h se genera mediante:

$$u_k^{'h} = f\left(g_h([\mathbf{x}_h^k \mathbf{w}^k]; \boldsymbol{\theta}_h)\right). \quad (22)$$

de modo que la probabilidad de presentación se genere con:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = r \left(\sum_{h=1}^m a_h^k \cdot f\left(g_h([\mathbf{x}_h^k \mathbf{w}^k]; \boldsymbol{\theta}_h)\right) \right).$$

X.C.7. Ejemplo 4: Modelos de segundo orden

En una implementación, $s(\cdot)$ es una función de segundo orden y la probabilidad de presentación estimada u_k para el péptido $\mathbf{p}^k$ se proporciona con:

$$u_k = \Pr(\mathbf{p}^k \text{ presentados}) = \sum_{h=1}^m a_h^k \cdot u_k^{'h}(\boldsymbol{\theta}) - \sum_{h=1}^m \sum_{1j < h} a_h^k \cdot a_j^k \cdot u_k^{'h}(\boldsymbol{\theta}) \cdot u_k^{'j}(\boldsymbol{\theta}) \quad (23)$$

donde los elementos $u_k^{'h}$ 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.

En un aspecto, el modelo de ecuación (23) puede implicar que existe una posibilidad de que el péptido $\mathbf{p}^k$ 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.

De acuerdo con la ecuación (23), la probabilidad de presentación de que una secuencia peptídica 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 p^k de la suma para generar la probabilidad de presentación de que la secuencia peptídica p^k se presente mediante los alelos de MHC H .

A modo de ejemplo, la probabilidad de que el péptido 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:

$$u_k = f(x_2^k \cdot \theta_2) + f(x_3^k \cdot \theta_3) - f(x_2^k \cdot \theta_2) \cdot f(x_3^k \cdot \theta_3),$$

donde x_2^k, x_3^k son las variables de interacción con alelos para los alelos de HLA $h=2, h=3$ y θ_2, θ_3 son el conjunto de parámetros determinado para los alelos de HLA $h=2, h=3$.

A modo de otro ejemplo, la probabilidad de que el péptido 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:

$$u_k = f(NN_2(x_2^k; \theta_2)) + f(NN_3(x_3^k; \theta_3)) - f(NN_2(x_2^k; \theta_2)) \cdot f(NN_3(x_3^k; \theta_3)),$$

donde $NN_2(\cdot), NN_3(\cdot)$ son los modelos de red identificados para los alelos de HLA $h=2, h=3$ y θ_2, θ_3 son el conjunto de parámetros determinado para los alelos de HLA $h=2, h=3$.

XI.A Ejemplo 5: Módulo de predicción

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 p^k que tienen 8-15 aminoácidos. Por ejemplo, el módulo de predicción 320 puede procesar la secuencia determinada «IEFROEIFJEF» (SEQ ID NO: 73) en tres secuencias peptídicas con 9 aminoácidos «IEFROEIFJ» (SEQ ID NO: 74), «EFROEIFJE» (SEQ ID NO: 75) y «FROEIFJEF» (SEQ ID NO: 76). 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.

5 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
10 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
15 (donde N es en general la cantidad máxima de epítomos 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

20 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 $\mathbf{p}^k$, $k=1, 2, \dots, v$ para la inclusión en una vacuna de capacidad v , la secuencia de casete se proporciona
25 mediante concatenación de una serie de secuencias terapéuticas de epítomos $\mathbf{p}^k$, $k=1, 2, \dots, v$ que incluye, cada una, la secuencia de un péptido $\mathbf{p}^k$ correspondiente. En una modalidad, el módulo de diseño de casete 324 puede concatenar los epítomos directamente contiguos entre sí. Por ejemplo, un casete de vacuna $\mathbf{C}$ puede representarse como:

$$30 \quad \mathbf{C} = [\mathbf{p}^{t_1} \mathbf{p}^{t_2} \dots \mathbf{p}^{t_v}] \quad (24)$$

donde $\mathbf{p}^{t_i}$ denota el epítomo número i del casete. Por lo tanto, t_i corresponde a un índice $k=1, 2, \dots, v$ para el péptido seleccionado en la posición número i del casete. En otra modalidad, el módulo de diseño del casete 324 puede concatenar los epítomos con una o más secuencias enlazadoras opcionales entre los epítomos contiguos. Por ejemplo,

un casete para vacuna **C** se puede representar como:

$$C = [p^{t_1} l_{(t_1, t_2)} p^{t_2} l_{(t_2, t_3)} \cdots l_{(t_{v-1}, t_v)} p^{t_v}] \quad (25)$$

donde $l_{(ti, tj)}$ denota una secuencia enlazadora ubicada entre el epítipo número i p^{ti} y el epítipo número $j=i+1$ $p^{j=i+1}$ del casete. El módulo de diseño del casete 324 determina cuál de los epítopos seleccionados p^k , $k=1, 2, \dots, v$ 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.

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.

En una modalidad, los epítopos terapéuticos p^k pueden corresponder a los propios péptidos seleccionados p^k . En otra modalidad, los epítopos terapéuticos p^k también pueden incluir secuencias flanqueantes de extremo C y/o N, además de los péptidos seleccionados. Por ejemplo, un epítipo 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 epítipo 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 epítipo terapéutico p^k representa un epítipo de longitud fija. En otro ejemplo, el epítipo terapéutico p^k puede representar un epítipo de longitud variable, en el cual la longitud del epítipo 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 epítipo p^k .

En una modalidad, el módulo de diseño de casete 324 genera secuencias de casete teniendo en cuenta la presentación de epítipos de unión que abarquen la unión entre un par de epítipos terapéuticos en el casete. Los epítipos de unión son secuencias de epítipos no propios novedosos pero irrelevantes que surgen en el casete debido al proceso de concatenación de los epítipos terapéuticos y las secuencias enlazadoras en el casete. Las secuencias novedosas de epítipos de unión son diferentes de los epítipos terapéuticos propios del casete. Un epítipo de unión que abarca los epítipos p^{ti} y p^{tj} puede incluir cualquier secuencia de epítipos que se superponga tanto con p^{ti} como con p^{tj} que sea diferente a las secuencias de los propios epítipos terapéuticos p^{ti} y p^{tj} . Específicamente, cada unión entre el epítipo p^{ti} y un epítipo contiguo p^{tj} del casete con o sin una secuencia enlazadora opcional $I^{(ti,tj)}$ se puede asociar a $n_{(ti,tj)}$ epítipos de unión $e_n^{(ti,tj)}$, $n=1, 2, \dots, n_{(ti,tj)}$. Los epítipos de unión pueden ser secuencias que se superponen al menos parcialmente tanto con el epítipo p^{ti} como con p^{tj} o pueden ser secuencias que se superponen al menos parcialmente con secuencias enlazadoras colocadas entre los epítipos p^{ti} y p^{tj} . Los epítipos de unión pueden presentarse mediante los MHC clase I, MHC clase II o ambos.

La Figura 13 muestra dos secuencias de casete ilustrativas, el casete 1 (C_1) y el casete 2 (C_2). Cada casete tiene una capacidad para vacuna de $v=2$ e incluye los epítipos terapéuticos $p^{t1} = p^1 = \text{SINFEKL}$ (SEQ ID NO: 185) y $p^{t2} = p^2 = \text{LLLLLVVV}$ (SEQ ID NO: 77) y una secuencia enlazadora $I^{(t1,t2)} = \text{AAY}$ entre los dos epítipos. Específicamente, la secuencia de casete C_1 se proporciona mediante $[p^1 I^{(t1,t2)} p^2]$, mientras que la secuencia de casete C_2 se proporciona mediante $[p^2 I^{(t1,t2)} p^1]$. Los epítipos de unión ilustrativos $e_n^{(1,2)}$ del casete C_1 pueden ser secuencias tal como EKLAAYLLL (SEQ ID NO: 78), KLAAYLLLLL (SEQ ID NO: 79) y FEKLAAYL (SEQ ID NO: 80) que abarcan tanto el epítipo p^{t1} como p^{t2} en el casete y pueden ser secuencias tal como AAYLLLLL (SEQ ID NO: 81) y YLLLLLVVV (SEQ ID NO: 82) que abarcan la secuencia enlazadora y un único epítipo seleccionado en el casete. De manera similar, los epítipos de unión ilustrativos $e_m^{(2,1)}$ del casete C_2 pueden ser secuencias tal como VVVVAAYSIN (SEQ ID NO: 83), VVVVAAY (SEQ ID NO: 84) y AYSINFEK (SEQ ID NO: 85). Si bien ambos casetes implican el mismo conjunto de secuencias p^1 , $I^{(c1,c2)}$ y p^2 , el conjunto de epítipos de unión que se identifica es diferente según la secuencia

ordenada de los epítomos terapéuticos dentro del casete.

En una modalidad, el módulo de diseño de casete 324 genera una secuencia de casete que reduce la probabilidad de que los epítomos de unión se presenten en el paciente. Específicamente, cuando se inyecta el casete en el paciente, los epítomos 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 epítomos de unión no tienen un beneficio terapéutico y pueden disminuir la respuesta inmunitaria a los epítomos terapéuticos seleccionados en el casete mediante competencia de antígenos.⁷⁶

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 epítomos de unión asociados a la secuencia del casete se encuentra por debajo de un umbral numérico. La puntuación de presentación de epítomos de unión es una cantidad asociada a las probabilidades de presentación de los epítomos de unión en el casete y un valor mayor de puntuación de presentación de epítomos de unión indica una probabilidad mayor de que los epítomos de unión del casete se presenten mediante HLA clase I o HLA clase II, o ambos.

En una modalidad, el módulo de diseño de casete 324 puede determinar una secuencia de casete asociada a la puntuación de presentación de epítomos 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 $d(\mathbf{e}_n^{(ti,tj)}, n=1, 2, \dots, n_{(ti,tj)}) = d_{(ti,tj)}$, 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 epítomos de unión que abarquen entre el par de epítomos terapéuticos contiguos $\mathbf{p}^{ti}$ y $\mathbf{p}^{tj}$. La puntuación de presentación de epítomos 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:

$$puntuación = h(d_{(t_1, t_2)}, d_{(t_2, t_3)}, \dots, d_{(t_{v-1}, t_v)}) \quad (26)$$

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.

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 epítomos de unión para los casetes candidatos e identificar una secuencia de casete óptima asociada a una puntuación de epítomos 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 epítomos 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 epítomos 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 epítomos 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 epítomos que se predice que se van a presentar mediante cualquiera de los alelos de HLA clase I o HLA clase II del paciente.

El módulo de diseño de casete 324 puede verificar adicionalmente una o más secuencias de casetes candidatos para identificar si alguno de los epítomos de unión en las secuencias de casetes candidatos son epítomos 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 epítomos 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 epítomos de unión propios al configurar la medición de distancia $d_{(t_i, t_j)}$ en un valor muy grande (por ejemplo, 100) para pares de epítomos t_i, t_j donde la concatenación del epítomo t_i al extremo N del epítomo t_j produce la formación de un epítomo de unión propio.

Nuevamente con referencia al ejemplo en la Figura 13, 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 epítomos de unión posibles $\mathbf{e}_n^{(t1,t2)} = \mathbf{e}_n^{(1,2)}$ con longitudes, por ejemplo, de 8 a 15 aminoácidos para MHC clase I o 9-30 aminoácidos para MHC clase II. Dado que no se encuentran presentes otras uniones en el casete $\mathbf{C}_1$, la puntuación de presentación de epítomos 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 epítomos de unión posibles $\mathbf{e}_n^{(t1,t2)} = \mathbf{e}_n^{(2,1)}$ 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 epítomos 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 epítomos de unión es menor que la de la secuencia de casete de $\mathbf{C}_1$.

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 epítomos 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 $v=20$ epítomos, el módulo de diseño de casete 324 tiene que repetir $\sim 10^{18}$ casetes candidatos posibles para determinar el casete con la menor puntuación de presentación de epítomos 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 epítomos 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.

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 epítomos de unión menor que un umbral predeterminado como

5 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 epítomos 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 $v=20$ epítomos seleccionados y seleccionar el casete

10 candidato con una puntuación mínima de presentación de epítomos 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 epítomos 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

15 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 epítomos de unión, lo que hace que no valga la pena desde una perspectiva de adjudicación de recursos.

20 En otra modalidad, el módulo de diseño de casete 324 determina una configuración de casete mejorada mediante la formulación de la secuencia de epítomos 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

25 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

30 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.

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 p^k . La distancia desde un nodo correspondiente al epítopo p^k a otro nodo

correspondiente al epítopo p^m se proporciona mediante la medición de distancia de epítomos de unión $d_{(k,m)}$, mientras que la distancia desde el nodo correspondiente al epítopo p^m al nodo correspondiente al epítopo 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 epítomos del casete. La solución del TSP asimétrico indica una secuencia de epítomos terapéuticos que corresponden al orden en el cual debería concatenarse los epítomos en un casete para minimizar la puntuación de presentación de epítomos de unión en las uniones del casete. Específicamente, dado un conjunto de epítomos terapéuticos $k=1, 2, \dots, v$, el módulo de diseño de casete 324 determina las mediciones de distancia $d_{(k,m)}$, $k, m = 1, 2, \dots, v$ para cada par ordenado posible de epítomos terapéuticos en el casete. En otras palabras, para un par determinado k, m de epítomos, se determina tanto la medición de distancia $d_{(k,m)}$ para la concatenación del epítopo terapéutico p^m luego del epítopo p^k como la medición de distancia $d_{(m,k)}$ para la concatenación del epítopo terapéutico p^k luego del epítopo p^m , ya que estas mediciones de distancia pueden ser diferentes entre sí.

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 $(v + 1) \times (v + 1)$ $\mathbf{P}$ determinada por lo siguiente:

$$\mathbf{P} = \begin{bmatrix} 0 & \mathbf{0}^{1 \times v} \\ \mathbf{0}^{v \times 1} & \mathbf{D} \end{bmatrix} \quad (26).$$

La matriz $v \times v$ $\mathbf{D}$ es una matriz de distancia asimétrica, donde cada elemento $D(k, m)$, $k=1, 2, \dots, v$; $m=1, 2, \dots, v$ corresponde a la medición de distancia para una unión del epítopo p^k al epítopo p^m . Las filas $k = 2, \dots, v$ de $\mathbf{P}$ corresponden a nodos de los epítomos 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 epítomos terapéuticos para vacuna v y sea $S \subset E$ un subconjunto de epítomos. Para cualquiera de tal subconjunto S , sea $out(S)$ la cantidad de uniones de un epítomo con otro $x_{km} = 1$ donde k es un epítomo en S y m es un epítomo en $E \setminus S$. Dada una matriz de vía $\mathbf{P}$ conocida, el módulo de diseño de casete 324 encuentra

5 una matriz de vía $\mathbf{X}$ que resuelve el siguiente problema de programación lineal de enteros:

$$\min_x \sum_{k=1}^{v+1} \sum_{m=1}^{v+1} P_{km} \cdot x_{km} \quad (27)$$

en el cual P_{km} denota el elemento $P(k,m)$ de la matriz de vía $\mathbf{P}$, sujeto a las siguientes limitaciones:

$$\begin{aligned} \sum_{k=1}^{v+1} x_{km} &= 1, \quad m = 1, 2, \dots, v+1 \\ \sum_{m=1}^{v+1} x_{km} &= 1, \quad k = 1, 2, \dots, v+1 \\ x_{kk} &= 0, \quad k = 1, 2, \dots, v+1 \end{aligned}$$

$$out(S) \geq 1, \quad S \subset E, 2 \leq |S| \leq |V|/2$$

15 Las primeras dos limitaciones garantizan que cada epítomo 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.

Las soluciones para x_{km} , $k, m = 1, 2, \dots, v+1$ en el problema de
 20 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 epítomos terapéuticos para el casete que disminuyen la puntuación de presentación de los epítomos 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 epítomo terapéutico $\mathbf{p}^m$ debería
 25 concatenarse luego del epítomo terapéutico $\mathbf{p}^k$ en la secuencia de casete mejorada. Una solución de $x_{km} = 0$ indica que no existe tal vía o, en otras palabras, que el epítomo terapéutico $\mathbf{p}^m$ no debería concatenarse luego del epítomo terapéutico $\mathbf{p}^k$ en la secuencia de casete mejorada. En conjunto, los valores de x_{km} en el problema de
 30 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 $fantasma \rightarrow 1 \rightarrow 3 \rightarrow 2 \rightarrow fantasma$ de nodos y nodos fantasma.

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 epítomos terapéuticos en el casete. La secuencia refinada indica el orden en el cual deberían concatenarse los epítomos 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 $1 \rightarrow 3 \rightarrow 2$. La secuencia refinada indica un modo posible de concatenar epítomos en el casete, a saber $p^1 \rightarrow p^3 \rightarrow p^2$.

En una modalidad, cuando los epítomos terapéuticos p^k son epítomos de longitud variable, el módulo de diseño de casete 324 determina las mediciones de distancia candidatas correspondientes a las diferentes longitudes de los epítomos terapéuticos p^k y p^m e identifica la medición de distancia $d_{(k,m)}$ como la menor medición de distancia candidata. Por ejemplo, cada uno de los epítomos $p^k=[n^k \ p^k \ c^k]$ y $p^m=[n^m \ p^m \ c^m]$ 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 epítomos p^k y p^m se asocia a 16 conjuntos diferentes de epítomos de unión según 4 valores de longitud posibles de n^k y los 4 valores de longitud posibles de 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 epítomos 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.

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 $v \times (v-1)$, que cada una corresponde a un par de epítomos terapéuticos en la vacuna. Una secuencia de casete determinada mediante este enfoque puede producir una secuencia con una presentación de epítomos 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 epítomos de unión para secuencias de casete generadas mediante muestreo aleatorio en oposición a TSP asimétrico

Se generaron dos secuencias de casete que incluían $v=20$ de epítomos

5 terapéuticos mediante muestreo aleatorio de 1.000.000 permutaciones (secuencia de casete C_1) y mediante resolución del problema de programación lineal de enteros en la ecuación (27) (secuencia de casete 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, x_h^i es la secuencia de péptidos p^i , $g_h(\cdot)$ 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 p^i , la antigenicidad de la proteína del péptido p^i y la muestra de identificación de origen del péptido p^i y $g_w(\cdot)$ de la secuencia flanqueante y TPM logarítmicas son respectivamente funciones de redes neuronales. Cada una de las funciones de redes

15 neuronales para $g_h(\cdot)$ 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

25 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 epítomos 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

30

aleatorio.

Específicamente, los $v=20$ epítomos se proporcionaron mediante:

p^{*1} = YNYSYWISIFAHTMWYNIWHVQWVK (SEQ ID NO: 86)

p^{*2} = IEALPYVFLQDQFELRLLKGEQGNN (SEQ ID NO: 87)

5 p^{*3} = DSEETNTNYLHYCHFHWTTWAQQTTV (SEQ ID NO: 88)

p^{*4} = GMLSQYELKDCSLGFSWNDPAKYLR (SEQ ID NO: 89)

p^{*5} = VRIDKFLMYVWYSAPFSAYPLYQDA (SEQ ID NO: 90)

p^{*6} = CVHIYNNYPRMLGIPFSVMVSGFAM (SEQ ID NO: 91)

p^{*7} = FTFKGNIWIEMAGQFERTWNYPLSL (SEQ ID NO: 92)

10 p^{*8} = ANDDTPDFRKCYIEDHSFRFSQTMN (SEQ ID NO: 93)

p^{*9} = AAQYIACMVNRQMTIVYHLTRWGMK (SEQ ID NO: 94)

p^{*10} = KYLKEFTQLLTFVDCYMWITFCGPD (SEQ ID NO: 95)

p^{*11} = AMHYRTDIHGYWIEYRQVDNQMWNT (SEQ ID NO: 96)

p^{*12} = THVNEHQLEAVYRFHQVHCRFPYEN (SEQ ID NO: 97)

15 p^{*13} = QTFSECLFFHCLKVWNNVKYAKSLK (SEQ ID NO: 98)

p^{*14} = SFSSWHYKESHIALLMSPKKNHNNT (SEQ ID NO: 99) p^{*15} =
ILDGIMSRWEKVCTRQTRYSCQCA (SEQ ID NO: 100)

p^{*16} = YRAAQMSKWPKNKYFDFPEFMAYMPI (SEQ ID NO: 101)

p^{*17} = PRPGMPCQHHNTHGLNDRQAFDDFV (SEQ ID NO: 102)

20 p^{*18} = HNIISDETEVWEQAPHITWVYMWCR (SEQ ID NO: 103)

p^{*19} = AYSWPVVPKMWIPYRALCANHPPGT (SEQ ID NO: 104)

p^{*20} = HVMPHVAMNICNWEFLYRISHIGR (SEQ ID NO: 105).

En el primer ejemplo, se generaron en forma aleatoria 1.000.000 secuencias de
25 casete candidatas diferentes con los 20 epítomos 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:

C_1 =

30 THVNEHQLEAVYRFHQVHCRFPYENAMHYQMWNTYRAAQMSKWPKNKYFDFPEFMAYM
PICVHIYNNYPRMLGIPFSVMVSGFAMAYSWPVVPKMWIPYRALCANHPPGTANDDTPD
FRKCYIEDHSFRFSQTMNIEALPYVFLQDQFELRLLKGEQGNNDSSEETNTNYLHYCHFHW
TTWAQQTTVILDGIMSRWEKVCTRQTRYSCQCAFTFKGNIWIEMAGQFERTWNYPLSLS
FSSWHYKESHIALLMSPKKNHNNTQTFSECLFFHCLKVWNNVKYAKSLKHVMPHVAMNI

CNWYEFlyRISHIGRHNIISDETEVWEQAPHITWVYMWCRVRIDKFLMYVWYSAPFSAYP
 LYQDAKYLKEFTQLLTFVDCYMWITFCGPDAAQYIACMVNRQMTIVYHLTRWGMKYNYS
 YWISIFAHTMWYNIWHVQWNGMLSQYELKDCSLGFSWNDPAKYLRPRPGMPCQHHNT
 HGLNDRQAFDDFV (SEQ ID NO: 106)

5

con una puntuación de presentación de 6,1 de la cantidad esperada de epítomos
 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
 epítomos de unión presentados se puede reducir significativamente mediante la
 identificación de una secuencia de casete entre los casetes muestreados en forma
 aleatoria.

10

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 epítomos 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:

15

$C_2 =$

IEALPYVFLQDQFELRLLKGEQGNNILDGIMSRWEKVCTRQTRYSCQCAHVMPHVAMNI
 CNWYEFlyRISHIGRTHVNEHQLEAVYRFHQVHCRFPYENFTFKGNIWIEMAGQFERTW
 NYPLSLAMHYQMWNTSFSSWHYKESHIALLMSPKKNHNNTVRIDKFLMYVWYSAPFSAY
 PLYQDAQTFSECLFFHCLKVWNNVKYAKSLKYRAAQMSKWPKNKYFDFPEFMAYMPIAYS
 WPVVPWKWIPYRALCANHPPGTCVHIYNNYPRMLGIPFSVMVSGFAMHNIISDETEVWE
 QAPHITWVYMWCRAAQYIACMVNRQMTIVYHLTRWGMKYNYSYWISIFAHTMWYNIWH
 VQWNGMLSQYELKDCSLGFSWNDPAKYLRKYLKEFTQLLTFVDCYMWITFCGPDAND
 DTPDFRKYIEDHSFRFSQTMNDSEETNTNYLHYCHFHWTTWAQQTTPRPGMPCQHHN
 THGLNDRQAFDDFV (SEQ ID NO: 107)

25

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

30

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.

5 Los resultados muestran que el problema de programación de enteros posiblemente puede proporcionar una secuencia de casete con una cantidad menor de epítomos de unión presentados que una identificada a partir de muestreo aleatorio, posiblemente con menos recursos informáticos.

10

XI.B.3. Comparación de la presentación de epítomos de unión para la selección de secuencias de casete generadas mediante MHCflurry y el modelo de presentación

En el presente ejemplo, se generaron secuencias de casete que incluyen $v=20$ epítomos terapéuticos se seleccionaron según la secuenciación de exoma
 15 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
 20 puntuación de presentación, según la cantidad de epítomos 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 epítomos terapéuticos de entre 98
 25 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 epítomos terapéuticos se pueden seleccionar según otros criterios; tal como aquellos basados en la estabilidad, o combinaciones de criterios tales como puntuación de presentación, afinidad y otros.
 30 Además, se aprecia que los criterios usados para priorizar epítomos 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.

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.

Específicamente en este ejemplo, los $v=20$ epítomos terapéuticos fueron

5 SSTPYLYYGTSSVSQFPMVPGGDR (SEQ ID NO: 108)
 EMAGKIDLLRDSYIFQLFWREAAEP (SEQ ID NO: 109)
 ALKQRTWQALAHKYNSQPSVSLRDF (SEQ ID NO: 110)
 VSSHSSQATKDSAVGLKYSASTPVR (SEQ ID NO: 111)
 KEAIDAWAPYLPEYIDHVISPGVTS (SEQ ID NO: 112)
 SPVITAPPSSPVFDTSDIRKEPMNI (SEQ ID NO: 113)
 10 PAEVAEQYSEKLVYMPHTFFIGDHA (SEQ ID NO: 114)
 MADLDKLNHSIIQRLLEVRGS (SEQ ID NO: 115)
 AAAYNEKSGRITLLSLLFQKVFAQI (SEQ ID NO: 116)
 KIEEVRDAMENEIRTQLRRQAAAHT (SEQ ID NO: 117)
 DRGHYVLCDFGSTTNKFQNPQTEGV (SEQ ID NO: 118)
 15 QVDNRKAEAEAAIKRLSYISQKVSD (SEQ ID NO: 119)
 CLSDAGVRKMTAAVRVMKRGLENLT (SEQ ID NO: 120)
 LPPRSLPSPDPSQVPASPQSQSSSQ (SEQ ID NO: 121)
 ELVLEDLQDGDVKMGGSFGRGAFSNS (SEQ ID NO: 122)
 VTMDGVREEDLASFSLRKRWESEPH (SEQ ID NO: 123)
 20 IVGVMFFERAFDEGADAIYDHINEG (SEQ ID NO: 124)
 TVTPTPTPTGTQSPTPTPITTTTTV (SEQ ID NO: 125)
 QEEMPPRPCGGHTSSSLPKSHLEPS (SEQ ID NO: 126)
 PNIQAVLLPKKTDSHHKAKGK (SEQ ID NO: 127)

25 Los resultados de este ejemplo en la tabla a continuación comparan la cantidad de epítomos 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
 30 luego de 1 millón de muestras aleatorias. Para el tercer método, la cantidad mediana de epítomos de unión se halló en 1 millón de muestras aleatorias.

Umbral (nM)	ATSP # Unión	Muestreo aleatorio	Mediana
-------------	--------------	--------------------	---------

	Epítomos de unión	# Epítomos de unión con unión	# Epítomos de unión con unión
50	0	0	3
100	0	0	7
150	0	1	12
500	15	26	55
1000	68	91	131

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 epítomos 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 epítomos 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.

En otro ejemplo, se generaron los mismos casetes con el mismo tipo de HLA y 20 epítomos 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 epítomos 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 # Epítomos de unión	Muestreo aleatorio # Epítomos de unión	Mediana # Epítomos 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

Los ejemplos que anteceden identificaron que los criterios para la

- 5 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 epítomos 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 epítomos que cumple o supera un umbral
- 10 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 epítomos de unión que supera un umbral de probabilidad de presentación.
- 15 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 epítomos que supera/es escaso respecto a algún umbral, se puede determinar en su lugar qué proporción de epítomos de unión supera o es escasa respecto al umbral, por ejemplo si el X % superior
- 20 de epítomos de unión tiene una probabilidad de presentación que supera cierto umbral Y, o si X % de epítomos 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 epítomos de unión individuales o estadística derivada de las acumulaciones de algunos o todos los epítomos de unión. Aquí, X puede ser en general
- 25 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.

Como tal, en determinados aspectos, los epítomos de unión con mayores probabilidades de presentación se pueden eliminar; los epítomos de unión con pocas probabilidades de presentación pueden conservarse; los epítomos de unión que se unen potentemente, es decir, los epítomos de unión con una afinidad de unión menor que 1000nM o 500nM o algún otro umbral, pueden eliminarse; y/o los epítomos con unión frágil, es decir, epítomos de unión con una afinidad de unión que supera los 1000nM o 500nM o algún otro umbral, pueden conservarse.

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 epítomos 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

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.

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 p^j 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

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.

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.

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.

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.

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 epítomos de linfocitos T en oposición a los modelos de vanguardia

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 epítomos de linfocitos T. Los datos de epítomos 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 epítomos 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 epítomos 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 inmunogénica en el sistema inmunitario.

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.

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 epítomos 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.

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.

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 epítomos que fueron reconocidos por los linfocitos T.

XII.C. Comparación de diferentes rendimientos del modelo de presentación en los datos de espectrometría de masas

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.

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.

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 %.

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

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.

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 $g_n(\cdot)$, 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.

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 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.

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

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.

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 B7 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-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.

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

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.

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 posición 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

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 compartan a través de los alelos de MHC para mejorar el rendimiento del modelo de presentación.

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.

- 5 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
- 10 la potencia estadística de las variables de no interacción con alelos a través de los alelos de MHC mediante el moldeado 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

- 15 La Figura 13H ilustra la dependencia entre la fracción de péptidos presentados para genes en función de la cuantificación de ARNm para los datos de espectrometría de masas en células tumorales. Los resultados mostraron que existe una dependencia fuerte entre la expresión de ARNm y la presentación de péptidos.

- 20 Específicamente, el eje horizontal en la Figura 13G 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 13G indica la fracción de epítomos 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.
- 25 Como se muestra en la Figura 13G, 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.

- 30 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

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.

«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 $g_w(\cdot)$.

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.

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.

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.

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

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.

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 «Probabilidad de datos de prueba reales» 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 «Gritstone, con ARN» indica mediciones generadas a partir del modelo de presentación «Gritstones, con ARN». Como se muestra en la Figura 13J, las mediciones generadas por el modelo «Gritstone, con ARN» sigue estrechamente aquellas mostradas en la «Probabilidad de datos de prueba real» y representaron correctamente diferentes grados de presentación de péptidos para las longitudes 8, 9, 10 y 11.

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

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 :

$$u_k = \expit\left(\text{relu}\left(x_h^k \cdot W_h^1 + b_h^1\right) \cdot W_h^2 + b_h^2\right),$$

donde $\text{relu}(\cdot)$ es la función de unidad lineal rectificada (RELU) y W_h^1 , b_h^1 , W_h^2 , y b_h^2 son el conjunto de parámetros θ determinado para el modelo. Las variables de interacción con alelos x_h^k consiste en secuencias de péptidos. Las dimensiones de W_h^1 son (231 x 256), las dimensiones de b_h^1 (1 x 256), las dimensiones de W_h^2 son (256 x 1), y b_h^2 es un escalar. Para fines demostrativos, los valores para b_h^1 , b_h^2 , W_h^1 , y 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.

XIII. Computadora de ejemplo

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.

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
 5 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
 10 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.

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,
 15 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.

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
 20 de los componentes descritos anteriormente, tales como los adaptadores gráficos 1412 y pantallas 1418.

XIV. Ejemplo de vector de administración del neoantígeno

A continuación figuran ejemplos de modalidades específicas para llevar a cabo la presente invención. Los ejemplos se proporcionan solamente con fines ilustrativos y no
 30 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.

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 al., *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

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 epítomos como un producto génico único donde los epítomos 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 epítomos 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 epítomos; (3) si la posición relativa de los epítomos dentro del casete tienen un efecto en las respuestas de linfocitos T; (4) si la cantidad de epítomos dentro de un casete influencia la magnitud o calidad de las respuestas del linfocito T a epítomos individuales; (5) si la adición de las secuencias de direccionamiento celular mejora las respuestas de linfocitos T.

Se desarrollaron dos lecturas para evaluar la presentación de antígenos y respuestas de linfocitos T específicas para los epítomos 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 al., 2002; Nagai et al., 2012); y (2) un ensayo in vivo que utilizó ratones transgénicos HLA-A2 (Vitiello et al., 1991) para evaluar la inmunogenicidad luego de la vacunación de epítomos derivados del casete de origen

humano por sus respuestas de linfocitos T específicos del epítipo correspondientes (Cornet et al., 2006; Depla et al., 2008; Ishioka et al., 1999).

XIV.B. Evaluación del diseño del casete del neoantígeno

XIV.B.1. Métodos y materiales

Diseño y clonación del TCR y el casete

Los TCR seleccionados reconocen péptidos NLVPMVATV (SEQ ID NO: 128) (PDB# 5D2N), CLGGLLTMV (SEQ ID NO: 129) (PDB#3REV), GILGFVFTL (SEQ ID NO: 130) (PDB#1OGA), LLFGYPVYV (SEQ ID NO: 131) (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 epítipo in vitro

Se adquirieron péptidos de ProlImmune 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 %.

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.

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 puomicina, 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 epítomos in vitro

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: 128), CLGGLTMV (SEQ ID NO: 129), GLCTLVAML (SEQ ID NO: 132), LLFGYPVYV (SEQ ID NO: 131), GILGFVFTL (SEQ ID NO: 130) y dos péptidos irrelevantes (WLSLLVPFV (SEQ ID NO: 133), FLLTRICT (SEQ ID NO: 134) se cargaron en linfocitos T2. Brevemente, los linfocitos T2 se contaron y diluyeron hasta 1x10⁶ células/mL con IMDM más FBS_{hi} 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 FBS_{hi} al 20 % diluidas hasta 5x10⁵ 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 5x10⁵ células/mL en RPMI 1640 más FBS_{hi} al 10 % y se agregaron 100 µL a las células T2. Se incubaron placas durante la noche 37 °C, 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.

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 5x10⁵ células/100 µl en una

placa para cultivo de tejido Costar de 96 pocillos. Las placas se incubaron durante aproximadamente 2 horas a 37 °C. Se diluyeron casetes adenovirales con MEM más FBS_{hi} 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 °C. Se contaron y se diluyeron clones de TCR Jurkat-Lucia hasta 5x10⁵ células/mL en RPMI más FBS_{hi} 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 °C, 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

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 al., 1991). Los ratones utilizados para estos estudios fueron progeñe 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)

Los ratones HLA-A2 Tg se inmunizaron con partículas 1x10¹⁰ a 1x10⁶ 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

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 estreptomycin (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

El análisis ELISPOT se llevó a cabo de acuerdo con los lineamientos generales de armonización ELISPOT (Janetzki et al., 2015) con el kit ELISpotPLUS de IFNγ de ratón (MABTECH). Se incubaron 1x10⁵ esplenocitos con 10 µM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNγ 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 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.

Tinción de citosina intracelular (ICS) ex vivo y análisis de citometría de flujo

Los linfocitos aislados recientemente a una densidad de $2-5 \times 10^6$ células/mL se incubaron con 10 μ M de los péptidos indicados durante 2 horas. Luego de dos horas, se agregó brefeldin A a una concentración de 5 μ g/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-IFN γ 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 IFN γ ⁺ de las células CD8⁺ como la cantidad de células IFN γ ⁺ total/ 1×10^6 de células vivas en respuesta a cada estimulante de péptidos.

XIV.B.2. Evaluación in vitro de diseños del casete de neoantígenos

Como ejemplo de la evaluación del diseño del casete de neoantígenos, se desarrolló un ensayo basado en células in vitro para evaluar si los epítomos 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.

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 al., 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 al., 2006; Yachi et al., 2006).

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
Epítipo	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 epítipo 3 no generado aún

En otro ejemplo, una serie de casetes cortos, todos los epítipos se incorporaron en la misma posición (Figura 16A) y solamente los enlazadores que separan los epítipos 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 epítipo pobre (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 epítipo de linfocitos T solamente (sin enlazador) = 9AA, enlazador natural un solo lado = 17AA, enlazador natural ambos lados = 25AA, enlazadores no naturales = AAY, RR, DPP.

Epítipo	Diseños del casete corto					
	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 ± 0,4	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 epítipo 3 no generado aún

En otro ejemplo, se construyó una serie de casetes cortos que, aparte de epítomos 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 significativamente 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 epítomos marcadores restringidos HLA-A*0201 se liberan de forma eficaz de los casetes modelo y las secuencias de direccionamiento no mejoraron significativamente el reconocimiento y activación de linfocitos T.

	Diseños del casete corto									
Epítipo	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 epítipo 3 no generado aún

XIV.B.3. Evaluación in vivo de diseños del casete de neoantígenos

A modo de ejemplo adicional de la evaluación del diseño del casete de neoantígenos, los casetes de vacuna se diseñaron para contener 5 epítomos 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 epítomos 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, $\alpha 1$ y $\alpha 2$ ligados al dominio $\alpha 3$, transmembrana y citoplasmático de H2-Kb (Vitiello et al., 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 $\alpha 3$ en el MHC.

Para los casetes cortos, todos los epítomos marcadores generaron una respuesta de linfocitos T potente, como se determina mediante ELISPOT de IFN-gamma, que fue aproximadamente 10-50 veces más potente de lo que se describe comúnmente (Cornet et al., 2006; Depla et al., 2008; Ishioka et al., 1999). De todos los enlazadores evaluados, el concatémoro de secuencias 25mer, cada uno con un epítomo 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 epítomos restringidos de MHC clase I en el casete.

	Diseños del casete corto					
Epítomo	9AA	17AA	25AA	AAV	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

En otro ejemplo, se construyó una serie de casetes de vacuna y se incorporó en vectores adenovirales que, junto con los 5 epítomos marcadores originales, contuvieron 16 epítomos 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 epítomos 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 epítomos no tuvo un impacto en la magnitud de la respuesta inmunitaria al conjunto original de epítomos y (b) la posición de un epítomo en un casete no influenció la respuesta consiguiente de linfocitos T a este (Tabla 6).

Tabla 6: Evaluación in vivo del impacto de la posición del epítomo 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.

Epítomo	Diseños del casete largo		
	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+/-	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

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 óptima cuando se empleó un enfoque de «tira de cuentas» que codifica
 alrededor de 20 epítomos en el contexto de un vector a base de adenovirus. Los epítomos
 se ensamblaron mejor mediante la concatenación de secuencias 25mer, cada una integra
 un epítomo de linfocito T CD8 mínimo (por ejemplo, 9 residuos de aminoácidos) que fueron
 5 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 epítomo dado en el contexto natural de ese epítomo dentro de su proteína de
 fuente. Por ejemplo, el epítomo de MHC I HCMV pp65 NLVPMVATV (SEQ ID NO: 128) está
 10 flanqueado en su extremo 5' por la secuencia 5' WQAGILAR (SEQ ID NO: 135 en su 3' por
 la secuencia 3' natural QGQNLKYQ (SEQ ID NO: 136), por lo que genera el péptido 25mer
 WQAGILARNLVPMVATVQGQNLKYQ (SEQ ID NO: 137) encontrado dentro de la proteína
 fuente HCMV pp65. La secuencia natural también puede hacer referencia a una secuencia
 15 naturales. Cada secuencia 25mer se conecta directamente a la siguiente secuencia 25mer.
 En los casos donde el epítomo 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 epítomo de linfocito
 T CD8 de 10 aminoácidos puede estar flanqueada por una secuencia de 8 aminoácidos y
 20 una de 7 aminoácidos. El concatémoro fue seguido por dos epítomos 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 al.,
 1994; Panina-Bordignon et al., 1989). Los epítomos de clase II se enlazaron al final del
 epítomos de clase I mediante un enlazador de aminoácidos GPGPG (SEQ ID NO:56). Los
 25 dos epítomos 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
 epítomos demostraron tener un impacto sustancial en el reconocimiento o la respuesta de
 linfocitos T. Las secuencias de direccionamiento tampoco parecieron afectar
 30 sustancialmente la inmunogenicidad de los antígenos derivados del casete.

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
 epítomos caracterizados ampliamente de linfocitos T conocidos por ser inmunogénicos en
 primates no humanos (NHP), ratones y seres humanos. Los 20 epítomos, todos integrados

en sus secuencias 25mer naturales, son seguidos de dos epítomos 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

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).

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.

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 complete "ChAdV68.5WTnt" (SEQ ID NO:1);
secuencia AC_000011.1 con nucleótidos de ATCC VR-594 correspondientes
sustituídos en cinco posiciones.

CCATCTTCAATAATATACCTCAAACCTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTGGGGAG
GAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGACGTGGTT
GCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTTGAACACGGAA
ATACTCAATTTTCCGCGCTCTCTGACAGGAAATGAGGTGTTTTCTGGGCGGATGCAAGTGAAAACGGGCCA
TTTTCGCGCGAAAACCTGAATGAGGAAGTGAAAATCTGAGTAATTTTCGCGTTTATGGCAGGGAGGAGTATTT
GCCGAGGGCCGAGTAGACTTTTGACCGATTACGTGGGGGTTTTCGATTACCGTGTTTTTACCTAAATTTCCG
CGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTGACGTGATCGCCAGGGTATTTAAACCTGCGCTCT
CCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACAC
TTTGAAAGATGAGGCACCTGAGAGACCTGCCCCGATGAGAAAATCATCATCGCTTCCGGGAACGAGATTCTG
GAACCTGGTGGTAAATGCCATGATGGGCGACGACCTCCGGAGCCCCCACCCTTTGAGACACCTTCGCT
GCACGATTTGTATGATCTGGAGGTGGATGTGCCCGAGGACGATCCCAATGAGGAGGCGGTAAATGATTTTT
TTAGCGATGCCGCGCTGCTAGCTGCCGAGGAGGCTTCGAGCTCTAGCTCAGACAGCGACTCTTCACTGCAT
ACCCCTAGACCCGGCAGAGGTGAGAAAAAGATCCCCGAGCTTAAAGGGGAAGAGATGGACTTGCGCTGCTA
TGAGGAATGCTTGCCCCGAGCGATGATGAGGACGAGCAGGCGATCCAGAACGCGAGCGAGCCAGGGAGTGC
AAGCCGCCAGCGAGAGCTTTGCGCTGGACTGCCCGCTCTGCCCGGACACGGCTGTAAGTCTTGTGAATTT
CATCGCATGAATACTGGAGATAAAGCTGTGTTGTGTGCACTTTGCTATATGAGAGCTTACAACCATTGTGT
TTACAGTAAGTGTGATTAAGTTGAACTTTAGAGGGAGGCAGAGAGCAGGGTGACTGGGCGATGACTGGTTTT
ATTTATGTATATATGTTCTTTATATAGGTCCCGTCTCTGACGCAGATGATGAGACCCCTACTACAAAGTCC
ACTTCGTCACCCCCAGAAATTGGCACATCTCCACCTGAGAATATTGTTAGACCAGTTCCTGTTAGAGCCAC
TGGGAGGAGAGCAGCTGTGGAATGTTTGGATGACTTGCTACAGGGTGGGGTTGAACCTTTGGACTTGTGTA
CCCGGAAACGCCCCAGGCACTAAGTGCCACACATGTGTGTTTACTTGAGGTGATGTCAGTATTTATAGGGT
GTGGAGTGCAATAAAAAATGTGTTGACTTTAAGTGCGTGGTTTTATGACTCAGGGGTGGGGACTGTGAGTAT
ATAAGCAGGTGCAGACCTGTGTGGTTAGCTCAGAGCGGCATGGAGATTTGGACGGTCTTGAAGACTTTCA
CAAGACTAGACAGCTGCTAGAGAACGCCTCGAACGGAGTCTCTTACCTGTGGAGATTCTGCTTCGGTGGCG
ACCTAGCTAGGCTAGTCTACAGGGCCAAACAGGATTATAGTGAACAATTTGAGGTTATTTTGAGAGAGTGT
TCTGGTCTTTTTGACGCTCTTAACTTGGGCCATCAGTCTCACTTTAACCAGAGGATTTTCGAGAGCCCTTGA
TTTTACTACTCCTGGCAGAACCACTGCAGCAGTAGCCCTTTTTTGCTTTTATTCTTGACAAATGGAGTCAAG
AAACCCATTTTCAGCAGGGATTACCAGCTGGATTTCTTAGCAGTAGCTTTGTGGAGAACATGGAAGTGCCAG
CGCTTGAATGCAATCTCCGGCTACTTGCCGGTACAGCCGCTAGACACTCTGAGGATCCTGAATCTCCAGGA
GAGTCCCAGGGCAGCCAAACGTGCGCAGCAGCAGCAGCAGGAGGAGGATCAAGAAGAGAACCCGAGAGCCG
GCCTGGACCCTCCGGCGGAGGAGGAGGAGTAGCTGACCTGTTTTCTGAACTGCGCCGGGTGCTGACTAGGT
CTTCGAGTGGTCCGGAGAGGGGGATTAAGCGGGAGAGGCATGATGAGACTAATCACAGAATGAACTGACT
GTGGGTCTGATGAGTCGCAAGCGCCAGAAACAGTGTGGTGGCATGAGGTGCAGTCGACTGGCACAGATGA
GGTGTGCGGTGATGCATGAGAGGTTTTCTCTAGAACAAGTCAAGACTTGTTGGTTAGAGCCTGAGGATGATT
GGGAGGTAGCCATCAGGAATTATGCCAAGCTGGCTCTGAGGCCAGACAAGAAGTACAAGATTACTAAGCTG
ATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTGAAGTGGAGATCTGTCTCCAGGAAAGGGT
GGCTTTTCAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGGGCATGGATGGGGTTACCTTTATGAACA
TGAGGTTTCAGGGGAGATGGGTATAATGGCACGGTCTTTATGGCCAATACCAAGCTGACAGTCCATGGCTGC
TCCTTCTTTGGGTTTAATAACACCTGCATCGAGGCCTGGGGTCAGGTCGGTGTGAGGGGCTGCAGTTTTTC
AGCCAACCTGGATGGGGGTGCTGGGCAGGACCAAGAGTATGCTGTCCGTGAAGAAATGCTTGTGTTGAGAGGT
GCCACCTGGGGGTGATGAGCGAGGGGCGAAGCCAGAATCCGCCACTGCGCCTCTACCGAGACGGGCTGCTTT
GTGCTGTGCAAGGGCAATGCTAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCA
GATGCTGACCTGCGCCGGCGGGAACAGCCATATGCTGGCCACCGTACATGTGGCTTCCCATGCTCGCAAGC
CCTGGCCCCGAGTTCGAGCACAATGTCATGACCAGGTGCAATATGCATCTGGGGTCCCGCCGAGGCATGTTT
ATGCCCTACCAAGTGAACCTGAATTATGTGAAGGTGCTGCTGGAGCCCGATGCCATGTCCAGAGTGAGCCT
GACGGGGGTGTTTGACATGAATGTGGAGGTGTGGAAGATTCTGAGATATGATGAATCCAAGACAGGTCGC
GAGCCTGCGAGTCCGGAGGGAAGCATGCCAGGTTCCAGTCCCGTGTGTGTGGATGTGACGGAGGACCTGCGA
CCCGATCATTTGGTGTGTCCTGACACGGGACGGAGTTCGGTTCCAGCGGGGAAGAATCTGACTAGAGTGA
GTAGTGTCTGGGGCGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTTCTGTGTGTT
GCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGGCGTCTCCCC
TCCTGGGCGGGAGTGCCTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCGCGAACTC
TTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGAGCTGCTGCATCTG
CCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACCTCGAGTTCC
ACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCAGCTCGAGGCCTTGACCCA

GCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGA
 AATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAAACACAGAGTCTGAATCTTT
 ATTTGATTTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCC
 AGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCT
 CCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGT
 GTTGACACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTG
 TTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTT
 ACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGTCAGGACCACCAGCACGGTGTATCCGGTGCACCTGGGGA
 ATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCCGCCAGGTTTTCC
 ATGCACTCATCCATGATGATGGCGATGGGCCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTCCGA
 CACATCATAGTTGTGGTCTCGGTGAGGTGCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGG
 ACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCTCACAGATCTGCATCTCCAGGCTTTG
 AGCTCGGAGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGAGATGAG
 CTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGACGCGGTGGGGCCGTAGATGACCCCGATGA
 CCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCTCCCGAGGAGGGGGGCCACCTCGTTTCATC
 ATCTCGCGCACGTGCATGTTCTCGCGCACCGATTCCGCCAGGAGGCGCTCTCCCCCAGGGATAGGAGCTC
 CTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTGCGCCATGGGCATTTTGAGAGGGGTTTGTGCA
 AGAGTTCCAGGCGGTCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTC
 GCGGGTTGGGACGGCTGCGGGAGTAGGGCACAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCTTTC
 CAGGGTTCGACGCTCCGCGTCAGGGTGGTCTCCGTACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGC
 GAGGGTGCCTTCAGGCTCATCCGGCTGGTTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTGGCCAGGT
 AGCAATTGACCATGAGTTCTGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAA
 GTCTGCCCCGAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGG
 GGCGTAGGCGTCCGCGCCGAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTGGGGTGGT
 CGGGGTCAAAAACAGTTTTCCCGCGGTTCTTTTTGATGCGTTTTCTTACCTTTGGTCTCCATGAGCTCGTGT
 CCCCCTGGGTGACAAAGAGGCTGTCCGTGTCCCGTAGACCGACTTTATGGGCCGGTCTCGAGCGGTGT
 GCCGCGGTCTCTCTCGTAGAGGAACCCCGCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAAGG
 AGGCCACGTGGGACGGGTAGCGGTGCTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATG
 TCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGGCCGG
 GGGGGTATAAAAGGGTCCGGGTCCCTGCTCGTCTCTACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCT
 GTTGGGGTAGGTATTCCCTCTCGAAGCGGGCATGACCTCGGCACTCAGGTGTGTCAGTTTCTAGAAACAGAG
 GAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCCTCGTCCATCTGGTTCAGAAAAGAC
 GATCTTTTTGTTGTGCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGC
 GCATGGTCTGGTTTTTTTTCTTGTGCGCGCGCTCCTTGGCGCGCATGTTGAGCTGCACGTACTCGCGCGCC
 ACGCACTTCCATTGCGGGAAGACGGTGGTGCAGCTCGTGGGCACGATTCTGACCTGCCAGCCCCGATTATG
 CAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGC
 CTTTGGCGGAGCAGAAGGGGGGAGGGGGTCCAGCATGACCTCGTGGGGGGGTGGGCATCGATGGTGAAG
 ATGCCGGGCAGGAGGTGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTC
 GCGCACGGCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAAGCGGGAGG
 CGTACATGCCGAGATGTCTGATAGCTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGC
 CCCCCGCGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGGCCAGGTT
 GGTGCGACTGGGCTTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGG
 GCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCT
 TGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATGAT
 GTCATACTTGAGCTGTCCCTTTTGTTCACAGCTCGCGGTTGAGAAGGAACCTCTCGCGGTCTTCCAGT
 ACTCTTCGAGGGGGAACCCGTCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTG
 TAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGGCAGGGAGGTGTGCGTGAG
 GCGGAAAGTGTCCCTGACCATGACCTTGAGAACTGGTGTGTAAGTCGATATCGTCGACGCCCCCTGTCT
 CCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGG
 ATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTGAT
 GACCTGGGCGGCGAGCACGATCTCGTGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAATC
 GCGGACGGCCCTTGACGTGGGGCAGTTTTCTTGAGCTCCTCGTAGGTGAGCTCGTGGGGTCTGAGCCCG
 TGCTGCTCGAGCGCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCCAG
 GGCGGTTTGCAGACGGTCCCGGTACTGACGGAAGTCTGCCCGACGGCCATTTTTTCGGGGGTGACGCAGT
 AGAAGGTGCGGGGGTCCCGGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACG
 AGCCGGTCTGTCGCCGAGAGTTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCA
 GGTGTAGGTTTTCCACATCGTAGGTGAGGAAGAGCCTTTCCGTGCGAGGATGCGAGCCGATGGGGAAGAACT

GGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAA
 CACTCGTGCTTGTGTTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAG
 CTGTACCTGAGTTCTTTTACGAGGAATTTTCACTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTGTA
 CTACGTCTGTTGGTGGTTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGCCCGCGCGGGAGG
 CAGGTCCAGACCTCGGCGCGAGCGGGTTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGT
 CCTGAGACGCTGCGGAGTCAGGTCAGTGGGCAGCGCGCGCGCGGTTGACTTGCAGGAGTTTTTCCAGGG
 CGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTGATGGCTTGCAGGGTC
 CCGTGCCCTGGGGTGTGACCACCGTCCCCCGTTTTCTTCTTGGGCGGCTGGGGCGACGGGGCGGTGCCTC
 TTCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCGCGGGCGGCGAGGGGCGGCTCGGGGCGCGGAGGCAGGGG
 CGGCAGGGGCACGTGCGCGCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGA
 CGACGCGACGGTTGACGTCTTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTG
 AAAGAGAGTTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCTGCGGCAGGATCTCTTGCACGTGCGC
 CGAGTTGTCTCTGGTAGGCGATCTCGGTATCATGAACGTGCTCGATCTCTCTCTTGAAGGTCTCCGCGGCGG
 CGCGCTCCACGGTGGCGCGAGGTTCGTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTTCATGCCCGC
 TCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCGGGCGCGCATGACCACCTGGGCGAGGTT
 GAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGCGA
 TGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACGTGCGCCAGCGCCTCC
 AAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGAGACGGTCAA
 CTCCTCTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCGGGAGTT
 CCTCCACTTCTCTTCTTCTCTCTCCACTAACATCTCTTCTACTTCTCTCCTCAGGCGGCAGTGGTGGCGGG
 GGAGGGGGCCTGCGTCGCGGCGGCGCACGGGCAGACGGTCGATGAAGCGCTCGATGGTCTCGCCGCGCCG
 GCGTCGCATGGTCTCGGTGACGGCGCGCCGTCCTCGCGGGGCGCAGCGTGAAGACGCCGCCGCGCATCT
 CCAGGTGGCCGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCGTA
 GGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCGAG
 CCAGTCGCAGTCGCAAGGTAGGCTGAGCACGGTTTTCTTCTGGCGGGTTCATGTTGGTTGGGAGCGGGGCGGG
 CGATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACCAGGTCT
 TTGGGCGCGGCTTGTCTGGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCTGACACCTGGCCAGGTG
 CTTGTAGTAGTCTGTCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCCGCGCGGCGCGTGCATGCGCGTGA
 GCGCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTTCGGCGACGACGCGCTCGGCGAGGATGGCTTGTCTGG
 ATCTGGGTGAGGTTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGTTGTTGATGGTGA
 GAGCAGTTGGCCATGACGGACAGTTGACGGTCTGGTGGCCCGGACGACGAGCTCGTGGTACTTGAGGAC
 GCGAGTAGGCGCGCGTGTGCAAGATGTAGTTCGTTGAGGTGCGCACCAAGGTACTGGTAGCCGATGAGGAAG
 TGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCTTCGAGCAT
 GGTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGCGGTTGGTGGAGGCGCGCGGGA
 ACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTTCATGGTGGGCACGGTCTGGCCCGTG
 AGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTGAGCGGCTCGACTCCGTGGCCTG
 GAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTTCAATCTCGAATCAGGCTGGAGCCGAGCT
 AACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCTGTTT
 TGCAACTTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTGCC
 GTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGTTTCGAGGCCGGCCGATTCCGCGGC
 TAACGAGGGCGTGGCTGCCCCGTCGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAG
 CCCCTCTTTTGTGTTTTGTTTTGTTTTGCCCCGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCTCC
 ACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGACCG
 CCGCGGCCCGCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGGAAGAGGGCGAGGGGCTG
 GCGCGCCTGGGGGCGTCTGTCGCCGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCCTA
 CGTGCCCCAAGCAGAACCTGTTTCAGAGACAGGAGCGGCGAGGAGCCGAGGAGATGCGCGCGGCCCCGTTCC
 ACGCGGGGCGGGAGCTGCGGCGCGGCTGGACCGAAAGAGGGTGTGAGGGACGAGGATTTTCAGGGCGGAC
 GAGCTGACGGGATCAGCCCCGCGCGCGCACGTGCGCCGCGGCAACCTGGTCACGGCGTACGAGCAGAC
 CGTGAAGAGGAGAGCAACTTCCAAAAATCCTTCAACAACACGTGCGCACCTGATGCGCACCTGACGCGAGGAGG
 TGACCTTGGGCTGATGCACCTGTGGGACCTGTGAGGACATCGTGCAGAACCCACCAGCAAGCCGCTG
 ACGGCGCAGCTGTTCTGTTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTTCAGGGAGGCGCTGCTGAATAT
 CACCGAGCCCGAGGGCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCGCG
 GGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGG
 AAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGAC
 CCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCAGTGAGCG
 CCAGCAGGCGGCGCGAGCTGAGCGACAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCGGG
 ACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCTTGGAGGC

GCGGCGAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGAT
 GGCGCGACCGTATTTTTTGTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCTGC
 AGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACG
 ACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGT
 GCCCTCGCGCTCCAACCCACGCACGAGAAGGTCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCA
 TCCGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAAC
 GTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCAGCGCGAGCGGTTCCACCG
 CGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCTCAGCACCCAGCCCCGCAACGTGCCCCGGG
 GCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTG
 TACCAGTCCGGGCGGACTACTTCTTCCAGACAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCCAGGC
 TTTCAAGAACTTGAGGGCCTGTGGGGCGTGCAGGCCCGGTCGGGGACCGCGCGACGGTGTGAGCCTGC
 TGACGCCGAACCTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTACGGACAGCGGCAGCATCAACCGCAAC
 TCGTACCTGGGCTACCTGATTAACCTGTACCCGCGAGGCATCGGCCAGGCGCACGTGGACGAGCAGACCTA
 CCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCGAGGACGACCCGGGCAACCTGGAAGCCACCTGAACT
 TTTTGCTGACCAACCGGTGCGAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTG
 CGTTACGTGCGAGCAGAGCGTGGGCCTGTTCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACAT
 GACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTCATCAATAAACTGATGGACTACT
 TGCATCGGGCGGCCCGCATGAACTCTGACTATTTACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCG
 CCGGGGTTCTACACGGGCGAGTACGACATGCCCGACCCCAATGACGGGTTCTGTGGGACGATGTGGACAG
 CAGCGTGTCTCCCCCGACCGGGTGCTAACGAGCGCCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCC
 CGTCTCGGCGCTGTCCGGCCGCGAGGGTGTGCGCGGGCGGTGCCCGAGGCCGCCAGTCTTTCCCGAGC
 TTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCGAGATCACGCGCCCGCGCTTGCTGGGCGA
 AGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAA
 GCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTGCAG
 GGGGCCACGAGCCGGGGCAGCGCCGCCCCGTAAACGCCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGA
 CGATGAGGACTCCGCCGACGACAGCAGCGTGTGGACTTGGGTGGGAGTGGTAACCCGTTTCTGCTCACCTGC
 GCCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACCAG
 CGTGCGTTTCTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCTCTCT
 CCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCGCTGGAGGCTCCTTA
 CGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCACCCCTTGT
 ACAGTACCACCCGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGAC
 CACAGAACTTCTGACCACCGTGGTGCAGAACAATGACTTCACCCCGAGGAGCCAGCAGCCAGACCAT
 CAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGTGA
 ACGAGTTTATGTACAGCAACAAGTTCAAGGCGCGGGTGTGGTCTCCCGCAAGACCCCCAATGGGGTGACA
 GTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGG
 CAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGC
 GGCAGAACGGGGTGTGGAGAGCGACATCGGCGTGAAGTTTCGACACTAGGAACCTCAGGCTGGGCTGGGAC
 CCCGTGACCGAGCTGGTTCATGCCCCGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCC
 CGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTTCGAAGAGGCAGCCCTTCC
 AGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGCAACATCCCCGCGCTCCTGGATGTGACGCC
 TATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTGAG
 GGGCGATAATTTTGAAGCGCCGCGAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCA
 TTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTAC
 CGCAGCTGGTACCTAGCCTACAACATATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACCAC
 CTCGGACGTACCTGCGGCGTGGAGCAAGTCTACTGGTGCCTGCCCCGACATGATGCAAGACCCGGTACCT
 TCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCCGTCTACTCCAAGAGC
 TTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTACCTCGCTTACGCACGTCTTCAACCG
 CTTCCTCCGAGAACAGATCCTCGTCCGCCCCCGCGCCACCATACCACCGTCAGTGAAAACGTTCTCTG
 CTTCTACAGATCACGGGACCTGCGCTGCGCAGTATCCGGGGAGTCCAGCGCGTACCGTTACTGAC
 GCCAGACCCGACCTGCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCTCCTCTCGAGCCG
 CACCTTCTAAATGTCCATTCTCATCTCGCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCAGCAAGA
 TGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCAGCGTGCAGCGGGCACTTCCGCGCTCCCTGG
 GGCGCCCTCAAGGGCCGCGTGCAGTGCAGCACACCGTGCAGACGTGATCGACAGGTGGTGGCCGACGC
 GCGCAACTACACCCCGCGCGCGCGCCGTCTCCACCGTGGACGCGGTTCATCGACAGCGTGGTGGCAGACG
 CGCGCCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCATCGCCCGGCGGCACCGGAGCACCCCGCCATG
 CGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACG
 CGCGGCTTACGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGCCA

GCATGTCCCGCCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTG
CGCACCCCGCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATG
TCCAAGCGCAAATTCAAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAA
GGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAGGAAGAAGAAAGTGATGTGGACG
GATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCGCGCGCGCTGCAGTGGCGCGGGCGGAAGGTGCAACCG
GTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCCGCGGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTA
CGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCGGAGCGCCTGGGCGAGTTTGTCTACGGCA
AGCGCAGCCGTTCCGCAACGAAGGAAGAGGCGGTGTCCATCCCCTGGACCACGGCAACCCACGCCGAGC
CTCAAGCCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCGCGGGGGTTCAAGCGCGAGGGCGA
GGATCTGTACCCCAACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGA
AGGTGGACCCGGACGTGCAGCCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCTGGGCGTG
CAGACCGTGGACATCAAGATTCCACGGAGCCCCATGGAACGCAGACCGAGCCCATGATCAAGCCCAAGCAC
CAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGCGCAAGTACG
GCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGC
ACGCGCTTCTACCGCGGTACATACCAGCAGCCGCCGCCAAGACCACCACTCGCCGCCGCCGTGCGCGCAC
CGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCGCGCACCTCTGACCCTGC
CGCGCGCGCGCTACCACCCGAGCATCGCCATTTAACTTTTCGCTTGTCTTGAGATCAATGGCCCTCACAT
GCCGCTTTCGCTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGTGGCGGGGAACGGG
ATGCGTCGCCACCACCACCGGCGGCGGCGGCCATCAGCAAGCGGTTGGGGGGAGGCTTCTGCCCGCGCT
GATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGC
GCCACTGAGACACACTTGGAAACATCTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCTGTGATGTG
TTTTCGTAGACAGATGGAAGACATCAATTTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTTATGG
GCACCTGGAGCGACATCGGCACCAGCCAATGAACGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGG
CTTAAGAATTTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACACAGGGCAGGCGCT
GAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTTCGATGGGCTCGCCTCGGGCATCAACGGGG
TGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCCGCCGGC
TCCGTGGAGATGCCGCGAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCC
CGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCGCGTACGAGGAGGCGGTGAACTGGGTG
TGCCCCACACGCGGCCCATCGCGCCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCCGCGACCTGGAC
TTGCCCTCCTCCCCAGCCTTCCCCGCCCTCTACAGTGGCTAAGCCCCCTGCCGCCGTGGCCGTGGCCCGCGC
TGACCCCGGGGCGACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAG
TGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTAT
GTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGG
CCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGT
CCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCAC
GGTGGCGCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACC
GCGAGGACAACACCTACTCGTACAAAGTGCCTACACGCTGGCCGTGGGCGACAACCGCGTGTGACATG
GCCAGCACCTACTTTGACATCCGCGGCGTGTGGATCGGGGCCCTAGCTTCAAACCTACTCCGGCACCGC
CTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTGAGTGGACATATAAAGCCGATGGTGAACTG
CCACAGAAAAAACCTATACATATGGAATGCACCCGTGCAGGGCATTAAACATCACAAAAGATGGTATTCAA
CTTGGAACCTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGGGTGA
TGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAA
TGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTGAGGCAAAATGTGAAAACAGGA
ACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCT
AGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTGGAAGTCCAGATACCCATATTGTATACAAAG
CAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTGAGCAAGCCATGCCCAACAGACCTAACTACATT
GGTTTCAGAGACAACCTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCA
GGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACGAAACACCGAGCTGTCTTACCAGCTCTTGCTTG
ACTCTCTGGGTGAGCAAGCCGGTATTTAGTATGTGGAATCAGGCGGTGGACAGTATGATCTCTGCTGTG
CGCATTATTGAAAATCATGGTGTGGAGGATGAACCTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAG
AACAGATACTTATCAGGGAATTAAGGCTAATGGAACCTGATCAAACCACATGGACCAAAGATGACAGTGTCA
ATGATGCTAATGAGATAGGCAAGGGTAATCCATTGCCATGGAATCAACATCCAAGCCAACCTGTGGAGG
AACTTCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCCT
GCCCCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACA
TCAACATCGGGGCGCGCTGGTTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACACCGCAATGCG
GGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAA
ATTTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGTCTTACACCTACGAGTGGAACCTCCGCAAGGACG

TCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCTCCATCTCCTTACCAGC
 ATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAA
 CGACACCAACGACCAGTCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACG
 CCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCGCCTTCCGCGGCTGGTCTTCACGCGTCTC
 AAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTA
 CCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTGAGCT
 GGCCCCGGAACGACCGGCTCCTGACGCCCAACGAGTTCGAAATCAAGCGCACCGTTCGACGGCGAGGGCTAC
 AACGTGGCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCACTACAACATCGGCTA
 CCAGGGCTTCTACGTGCCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGA
 GCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTACCCCTGGCCTACCAGCACAAAC
 TCGGGCTTTCGTGCGCTACCTCGCGCCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCC
 GCTCATCGGCAAGAGCGCGGTACCCAGCGTACCCAGAAAAAGTTCTCTGCGACAGGGTTCATGTGGCGCA
 TCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAAC
 TCCGCCACGCGCTAGACATGAATTTTGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTT
 CGAAGTCTTCGACGTCTGCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCC
 CTTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGC
 AGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCCTGGGCACCTTCGATAAGCGCTTC
 CCGGGATTTCATGGCCCCGCAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGA
 GCACTGGCTGGCCTTCGCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGG
 ACGAGCGCTCAAGCAGATCTACAGTTCGAGTACGAGGGCCTGCTGCGCCGACGCGCCCTGGCCACCGAG
 GACCGCTGCGTACCCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCGCTGCGGGCTCTT
 CTGCTGCATGTTCTGTCACGCTTTCGTGCACTGGCCCCACCGCCCCATGGACAAGAACCCACCATGAAC
 TGCTGACGGGGGTGCCAACCGCATGCTCCAGTGCAGGCGAGGTGGAACCCACCCTGCGCCGCAACCAGGAG
 GCGCTCTACCGCTTCTCAACTCCCACTCCGCTACTTTTCGCTCCCAACCGCGCGCGCATCGAGAAGGCCAC
 CGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTT
 TCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTCTGCGGGTCTCGGCATGGCCC
 GCGGGCAGGGACACGTTGCGGAACTGGTACTTGGCCAGCCACTTGAACCTCGGGGATCAGCAGTTTGGGCAG
 CGGGGTGTGCGGGAAGGAGTCCGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCAGCAGGTGCGGCGCGG
 AGATCTTGAAATCGCAGTTGGGACCCGCGTTCGTGCGCGCGGGAGTTGCGGTACACGGGGTTCAGCACTGG
 AACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTTCGCGTTCGGTTCATGCTCTCCACGTCAGGTC
 CTCGGCGTTGGCCATCCCGAAGGGGGTTCATCTTGCAGTCTGCGCTTCCATGTTGGGACGACACCCGGCT
 TGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTTCGGCGTTTCATCCCCGGGTACATG
 GCCTTCATGAAAGCCTCCAATTGCCTGAACGCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCGCA
 GGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTTCGTGCACGACGAGCGCGCGTTCGTTGTTGGCCA
 GCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTTCGGGGTCTCCTTCAGCGCGCGC
 TGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCA
 CCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCCTCCAGTTCTTGT
 GGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTTCAGGGTCTTGTG
 CTAGTGAAGGTTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTC
 GCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTTCAGGTTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAG
 TCATGATTTCCATACCTTCTCCAGGCCGAGACGATGGGCAGGCTCATAGGGTCTTTCACCATCATCTTA
 GCGCTAGCAGCCGCGGCCAGGGGGTTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCTTCTCGGT
 GATCCGCACCGGGGGGTAGCTGAAGCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTTCGTCTCTCGCTGT
 CCTGGCTGACGTCCTGCAGGACCACATGCTTGGTCTTTCGGGGTTTCTTCTTGGGCGGCAGCGGGCGGGA
 GATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCCTACTATCTCTTCTTCTTGGTCCGA
 GGCCACGCGCGGTAGGTATGTCTCTTCGGGGGAGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGCACTTG
 GCGGATGGCTGGCAGAGCCCCCTTCGCGGTTTCGGGGGTGCGCTCCCGCGCGCGCTCTGACTGACTTCTTCCG
 CGGCCGCCATTTGTGTTCTCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACTCGCCATC
 TGCCCCACCTCGCAGAGAAGCAGCAGCAGAGCATGAAGCTTAACCGCCCCGCGCCAGCCCGCCCA
 CCTCCGACGCGCGCTCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGTATGTGACG
 CCCGCGGAGCACGAGGAGGAGTGGCAGTGCCTTTTTACAAGAAGAGATACACCAAGAAGAGCCAGAGCA
 GGAAGCAGAGAATGAGCAGAGTACGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGG
 AGGACGCGCTCATCAAGCATCTGGCCCGGAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAG
 GTGCCCCCTCAGCGTGGAGGAGCTCAGCCGCGCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCACCCCAA
 GCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCTCAACTTCTACCCGGTCTTCGCGGTGCCCGAGG
 CCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCGCGCCAAACCGCACCCGC
 GCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCGCTACCTGATATCGCTCCTTGAAGAGGTTCCCAA

GATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATG
 AGCACCACAGCGCCCTGGTTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTTCGAG
 CTGACCCATTTTCGCTACCCGGCTCTGAACCTGCCCCCAAAGTCATGAGCGCGGTTCATGGACCAGGTGCT
 CATCAAGCGCGCTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCA
 GCGACGAGCAGCTGGCCCGGTGGCTGGGTCTAATGCTAGTCCCCAGAGTTTGAAGAGCGGCGCAAACCTC
 ATGATGGCCGTGGTCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGAGACCCT
 GCGCAAGGTTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCA
 ACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACC GCCTGGGGCAGAACGTGCTG
 CACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACAC
 CTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCC
 TGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACCTC
 ATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCATGTTGCA
 AAACCTTTCGCTCTTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGACT
 TCGTGCCGCTGACCTTCCGCGAGTGCCCCCGCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAAC
 TACCTGGCCTACCACTCGGACGTGATCGAGGACGTGAGCGGCGAGGGCCTGCTCGAGTGCCACTGCGCCTG
 CAACCTCTGCACGCCGACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCA
 CCTTCGAGTTGCAAGGGCCAGCGAAGGCGAGGGTTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGG
 CTGTGGACCTCGGCCTACTTGC GCAAGTTCTGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGA
 GGACCAATCCCATCCGCCCAAGGCCGAGCTGTGCGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAAT
 TGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAG
 ACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGC
 CGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAAGCAGAGGAGGAGGAGATGGAGGAAGAC
 TGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGA
 GGAGGTGGAAGAAGCAGCCGCCCGCCAGACCGTCTGCTCCTCGGCGGGGGAGAAAGCAAGCAGCACGGATACCA
 TCTCCGCTCCGGGTTCGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCC
 ACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCTTGGCGGGGGCACA AAAACGCCATCGTCTC
 CTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACT
 TTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCCTACTACTTCCAAGAAGAGGCAGCAGCA
 GCAGAAAAAGACCAGCAGAAAAACCAGCAGCTAGAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCG
 CGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCACCCCTCTATGCCATCTTCCAG
 CAGAGTCGGGGGCAGGAGCAGGAAGTGAAGTCAAGAACCCTTCTCTGCGCTCGCTCACCCGAGTTGTCT
 GTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCG
 CGCTCACTCTTAAAGAGTAGCCCCGCGCCCGCCAGTTCGAGAAAAAGGCGGGAATTACGTACCTGTGCCC
 TTGCGCCTAGCCGCTCCACCCATCATCATGAGCAAGAGATTCCACGCCTTACATGTGGAGCTACCAGC
 CCCAGATGGGCCTGGCCGCCGGTGCCGCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCC
 GCGATGATCTCACGGGTGAATGACATCCGCGCCACCAGAAACCAGATACTCCTAGAACAGTCAGCGCTCAC
 CGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGC
 CCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCG
 GGCGGCGCCACCCTGTGTGCTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGCAC
 ACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACCTCGCCGAT
 CGGGGAGATCTTCCCTTACGCCTCGTCAGGCCGCTGACTTTGGAGAGTTCGTCTCGCAGCCCCGCTCG
 GGTGGCATCGGCACTCTCCAGTTCTGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTC
 CCCCCGCCACTACCCGGACGAGTTTCATCCCGAATTTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATT
 GAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTTCGACACCTGGACCACTGCCGCCGCTTCCGCTGC
 TTCGCTCGGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCCCGAGGAGCACCCCTCAGGGCCCGGCCACGG
 AGTGCGGATCGTCGTGCAAGGGGGCCTCGACTCCCACCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGG
 TCGAGCGGAGCAAGGACAGACCCCTTCTGACTCTGTACTGCATCTGCAACCACCCCGGCTGCATGAAAGT
 CTTTGTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGACTTCCGTGTGTTTC
 CTGAATCCATCAACAGTCTTTGTTCTTACCAGGAACGAGACCGAGCTCCAGCTCCAGTGAAGCCCTCAC
 AAGAAGTACCTACCTGGCTGTTCCAGGGCTCCCGGATCGCCGTTGTCAACCACTGCGACAACACGGAGT
 CCTGCTGAGCGGCCCTGCCAACCTTACTTTTTTCCACCCGCGAGAAGCAAGCTCCAGCTCTTCCAACCCCTTCC
 TCCCCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACCACAGCG
 TCGCTCCCCGCTACTAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACCTTTCTGAATCTAATAC
 TACCACCCACACCGGAGGTGAGCTCCGAGGTCAACCAACCTCTGGGATTTACTACGGCCCTGGGAGGTGG
 TTGGGTAAATAGCGCTAGGCCTAGTTGCGGGTGGGCTTTTGGTTCTCTGCTACCTATACCTCCCTTGCTGT
 TCGTACTTAGTGGTGCTGTGTTGCTGGTTTAAAGAAATGGGGAAGATCACCTAGTGAGCTGCGGTGCGCTG
 GTGGCGGTGTTGCTTTTCGATTGTGGGACTGGGCGGTGCGGCTGTAGTGAAGGAGAAGGCCGATCCCTGCTT

GCATTTCAATCCCAACAAATGCCAGCTGAGTTTTTCAGCCCGATGGCAATCGGTGCGCGGTACTGATCAAGT
GCGGATGGGAATGCGGAGACGTGAGAATCGAGTACAATAACAAGACTCGGAACAATACTCTCGCGTCCGTG
TGGCAGCCCCGGGGACCCCGAGTGGTACACCGTCTCTGTCCCCGGTGCTGACGGCTCCCCGCGCACCGTGAA
TAATACTTTTCATTTTTTGCACATGTGCGACACGGTCATGTGGATGAGCAAGCAGTACGATATGTGGCCCC
CCACGAAGGAGAACATCGTGGTCTTCTCCATCGCTTACAGCCTGTGCACGGCGCTAATCACCGCTATCGTG
TGCCTGAGCATTACATGCTCATCGCTATTGCCCCAGAAATAATGCCGAAAAAGAAAAACAGCCATAACG
TTTTTTTTTACACCTTTTTTCAGACCATGGCCTCTGTAAATTTTTTGCTTTTTATTTGCCAGTCTCATTGCCG
TCATTCATGGAATGAGTAATGAGAAAATTACTATTTTACACTGGCACTAATCACACATTGAAAGGTCCAGAA
AAAGCCACAGAAGTTTCATGGTATTGTTATTTTAATGAATCAGATGTATCTACTGAACTCTGTGGAAACAA
TAACAAAAAAAATGAGAGCATTACTCTCATCAAGTTTCAATGTGGATCTGACTTAACCCTAATTAACATCA
CTAGAGACTATGTAGGTATGTATTATGGAACACAGCAGGCATTTTCGGACATGGAATTTTATCAAGTTTCT
GTGTCTGAACCCACACGCCTAGAATGACCACAACCAAAAACTACACCTGTTACCCTATGCAGCTCAC
TACCAATAACATTTTTTGGCATGCGTCAAATGGTCAACAATAGCACTCAACCCACCCACCGAGGAGGAAA
TTCCCAAATCCATGATTGGCATTATTGTTGCTGTAGTGGTGTGCATGTTGATCATCGCCTTGTGCATGGTG
TACTATGCCTTCTGCTACAGAAAGCACAGACTGAACGACAAGCTGGAACACTTACTAAGTGTTGAATTTTA
ATTTTTTTAGAACCATGAAGATCCTAGGCCTTTTTAATTTTTTCTATCATTACCTCTGCTCTATGCAATTCTG
ACAATGAGGACGTTACTGTGCTTGTGCGATCAAATTATACACTGAAAGGTCCAGCGAAGGGTATGCTTTTCG
TGGTATTGCTATTTTTGGATCTGACACTACAGAACTGAATTATGCAATCTTAAGAATGGCAAAATTCAAAA
TTCTAAAATTAACAATTATATATGCAATGGTACTGATCTGATACTCCTCAATATCACGAAATCATATGCTG
GCAGTTACACCTGCCCTGGAGATGATGCTGACAGTATGATTTTTTACAAAGTAAGTGTGTTGATCCCACT
ACTCCACCTCCACCCACCACAACACTACTCACACCACACACACAGATCAAACCGCAGCAGAGGAGGCAGCAAA
GTTAGCCTTGCAGGTCCAAGACAGTTTCAATTTGTTGGCATTACCCCTACACCTGATCAGCGGTGTCCGGGGC
TGCTAGTCAGCGGCATTGTGCGGTGTGCTTTTCGGGATTAGCAGTCATAATCATCTGCATGTTCAATTTTTGCT
TGCTGCTATAGAAGGCTTTACCGACAAAAATCAGACCCACTGCTGAACCTCTATGTTTAATTTTTTCCAGA
GTCATGAAGGCAGTTAGCGCTCTAGTTTTTTGTTCTTTGATTGGCATTGTTTTTTGCAATCCTATTTCCTAA
AGTTAGCTTTTATTAAGATGTGAATGTTACTGAGGGGGGCAATGTGACACTGGTAGGTGTAGAGGGTGCTG
AAAACACCACCTGGACAAAATACCACCTCAATGGGTGGAAAGATATTTGCAATTGGAGTGTATTAGTTTAT
ACATGTGAGGGAGTTAATCTTACCATTGTCAATGCCACCTCAGCTCAAAATGGTAGAATTCAAGGACAAAG
TGTCAGTGTATCTAATGGGTATTTTACCCAACATACTTTTATCTATGACGTTAAAGTCATACCACTGCCTA
CGCCTAGCCCCACTAGCACTACCACACAGACAACCCACACTACACAGACAACCACATACAGTACATTAAAT
CAGCCTACCACCACTACAGCAGCAGAGGTTGCCAGTCTGCTGCGGTCCGAGTGGCATTTTTGATGTGGGC
CCCATCTAGCAGTCCCAGTCTAGTACCAATGAGCAGACTACTGAATTTTTGTCCACTGTGAGAGCCACA
CCACAGCTACCTCCAGTGCCTTCTCTAGCACCGCCAATCTCTCCTCGCTTTTCTCTACACCAATCAGTCCC
GCTACTACTCCTAGCCCCGCTCCTCTTCCCACTCCCCTGAAGCAAACAGACGGCGGCATGCAATGGCAGAT
CACCTGCTCATTGTGATCGGGTTGGTCATCCTGGCCGTGTTGCTCTACTACATCTTCTGCCGCCGATTC
CCAACGCGCACCGCAAGCCGGTCTACAAGCCCATCATTGTGCGGCAGCCGGAGCCGCTTCAGGTGGAAGGG
GGTCTAAGGAATCTTCTCTTCTCTTTTACAGTATGGTGATTGAACATGATTTCCTAGACAATTCTTGATCA
CTATTCTTATCTGCCTCCTCCAAGTCTGTGCCACCCTCGCTCTGGTGGCCAACGCCAGTCCAGACTGTATT
GGGCCCTTTCGCTCCTACGTGCTCTTTGCCTTACCACCTGCATCTGCTGCTGTAGCATAGTCTGCCTGCT
TATCACCTTCTTCCAGTTTCAATGACTGGATCTTTGTGCGCATCGCCTACCTGCGCCACCACCCCACTACC
GCGACCAGCGAGTGGCGCGGCTGCTCAGGCTCCTCTGATAAGCATGCGGGCTCTGCTACTTCTCGCGCTTC
TGCTGTTAGTGCTCCCCCGTCCCGTCGACCCCGGTCCCCCACCCAGTCCCCCGAGGAGGTCCGCAAATGC
AAATTCCAAGAACCCTGGAAATTCCTCAAATGCTACCGCCAAAAATCAGACATGCATCCCAGCTGGATCAT
GATCATTGGGATCGTGAACATTCTGGCCTGCACCCTCATCTCCTTTGTGATTTACCCCTGCTTTGACTTTG
GTTGGAACCTCGCCAGAGGCGCTCTATCTCCCGCCTGAACCTGACACACCACCACAGCAACCTCAGGCACAC
GCACTACCACCACTACAGCCTAGGCCACAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACC
CATGCTCCCCGCTATTAGTTACTTCAATCTAACCGCGGAGATGACTGACCCACTGGCCAACAACAACGTC
AACGACCTTCTCCTGGACATGGACGGCCGCGCTCGGAGAGCGACTCGCCCACTTCGATTTCGCGAGCA
GCAGGAGAGAGCCGCTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGGAGCATCTTCTGCC
TGGTGAAACAGGCCAAGATCTCTACGAGGTCACTCAAACGACCATCGCCTCTCCTACGAGCTCCTGCAG
CAGCGCCAGAAGTTACCTGCCTGGTGGAGTCAACCCCATCGTCATCACCCAGCAGTCTGGCGATACCAA
GGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCACACTCTGATCAAGACCCTCTGCGGCCTCC
GCGACCTCCTCCCCATGAACATAACCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTAC
AGAAATAAAAAATAATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAAACAAAAA
ATAAAGAATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCCTTCACTCC
CCTCTTCCAGCTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACTTCTCCACACGCTGAAGGGGATGTCA
AATTCCTCCTGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAGCGCGTCCGGGTGGATGAT

GACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGT
 CTCTTCAGATGGATTCCAAGAGAAGCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCA
 AGAACGGGGAAATCACCCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCCTCGGGAAAACCTCATCTCCAAC
 ACGGCCACCAAGGCCGCCGCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCACCCCTT
 TTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAATATACTGAGAAACAAGCATTTCTAA
 ACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCT
 CCACTTACATTTGATACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGA
 TGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTTTAAAATTTGAAGATGGAGCCATAGCAACCAACATTG
 GAAATGGGTTAGAGTTTGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAA
 CTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCAC
 TTTGTGGACAACACCTGATCCATCACCAAACCTGTCAAATACTCGCAGAAAATGATGCAAACTAACACTTT
 GCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTGAGTCTTAGTTGTAGGAAGTGGAAACCTAAAC
 CCCATTACTGGCACCCTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAACCGGTGTTCTTTTAACAGA
 ACATTCTACACTAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATATACCAATG
 CTGTAGGATTTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTA
 GGGCAAGTATACATGAATGGAGATGTTTCAAAACCTATGCTTCTCACTATAACCCCTCAATGGTACTGATGA
 CAGCAACAGTACATATTCAATGTCATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTG
 GGGCTAACTCTTATACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCCT
 TCCCACCCCACTCTGTGGAACAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAA
 CAGTTTTTACAGGATTCGAGCAGTTATTTTTCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCC
 CCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTT
 CAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAG
 CTCAACAGCTGAGGATTGTCTCGGTGGTCGGGATCACGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGG
 GAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCGCAGCAGTCGCTGCCGCCGCCG
 CTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCACGGCCCTCAGCA
 TCAGTCGTCTGGTGCGGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTGCTGTCAGTACGTGCAACAC
 AGAACCACCAGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGGAAGGATGCT
 ACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCACGT
 ACATGATCTCCTTGGGCATGTGGCGGTTACACACCTCCCGGTACCACATCACCTCTGGTTGAACATGCAG
 CCCCCGATGATCTGCGGAACACAGGGCCAGCACCGCCCCCGCCCGCCATGCAGCGAAGAGACCCCGGGT
 CCGGCAATGGCAATGGAGAGCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGG
 CACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCAG
 GGCACGGGGAACCTCTTGAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTG
 CATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCT
 CACAGCGTGTTAAGGGGGCGGGCGGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTGCGGACCGTGTC
 ATGATGCAGTTGCTTTTCGGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATC
 GCCGGCGGGTCTCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAACACAGCCACTCTCTCAGACCGTGC
 AGCAGATCTAGGGCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGT
 GGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTGGGTTTCGGTGACGGCGGGGAGGGAAGAACAG
 GAAGAACCATGATTAACTTTAAATCCAACGGTCTCGGAGTACTTCAAATGAAGATCGCGGAGATGGCAC
 CTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCAC
 GGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTA
 ATTCTCAATCATCATGTTACACTCCTGCACCATCCCCAGATAATTTTCATTTTTCCAGCCTTGAATGATT
 CGAACTAGTTCcTGAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCAT
 TCTTAAGCACACCCTCATAATTCCAAGATATTCTGCTCCTGGTTACCTGCAGCAGATTGACAAGCGGAAT
 ATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTCTC
 CGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAACCAGGATCCT
 CCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATA
 ACTGGACAGAAAATCGCCAGGCAATTTTAAAGAAAATCAACAAAAGAAAAATCCTCAGGTGGACGTTTA
 GAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGTTAGCTGATGTAGA
 AAAACAAAAATGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAG
 GCAGGCCACGGGTCTCCGGCGCGACCCCTCGTAAAAATTGTGCTATGATTGAAAACCATCACAGAGAGAC
 GTTCCCGGTGGCCGGCGTGAATGATTGACAAGATGAATACACCCCGGAACATTGGCGTCCGCGAGTGAA
 AAAAGCGCCCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGG
 ATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCG
 ATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAG
 AGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCAGATCTACACTGA

CGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAACCGGTGACACTCAAAAAAATACGCGCACTTCCTCAAACGCCCCAAACTGCCGTCATTTCCGGGTTCCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAAAACGTCACCCGCCCCGCCCCCTAACGGTCGCCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG

C68 de ATCC VR-594 (SEQ ID NO:10); secuenciado de forma independiente;
C68 de longitud completa

CCATCTTCAATAATATACCTCAAACCTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTGGGGAGGAAGGGCGGTGATTGGTTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGACGTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTGAACACGGAAATACTCAATTTTCCGCGCTCTCTGACAGGAAATGAGGTGTTTTCTGGGCGGATGCAAGTGAAAACGGGCCATTTTCGCGCGAAAACCTGAATGAGGAAGTGAAAATCTGAGTAATTTTCGCGTTTATGGCAGGGAGGAGTATTTGCCGAGGGCCGAGTAGACTTTGACCGATTACGTGGGGGTTTTCGATTACCGTGTTTTTACCTAAATTTCCGCGTAGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTGAGTGTGATCGCCAGGGTATTTAAACCTGCGCTCTCCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCGAGTCAGATCTACACTTTGAAAGATGAGGCACCTGAGAGACCTGCCCCGATGAGAAAATCATCATCGCTTCCGGGAACGAGATTCTGGAACCTGGTGGAATGCCATGATGGGCGACGACCCTCCGGAGCCCCCACCCTTTGAGACACCTTCGCTGCACGATTTGTATGATCTGGAGGTGGATGTGCCCGAGGACGATCCCAATGAGGAGGCGGTAAATGATTTTTTTAGCGATGCCGCGCTGCTAGCTGCCGAGGAGGCTTCGAGCTCTAGCTCAGACAGCGACTCTTCACTGCATACCCCTAGACCCGGCAGAGGTGAGAAAAAGATCCCCGAGCTTAAAGGGGAAGAGATGGACTTGCCTGCTATGAGGAATGCTTGCCCCCGAGCGATGATGAGGACGAGCAGGCGATCCAGAACGAGCGAGCCAGGGAGTGC AAGCCGCCAGCGAGAGCTTTGCGCTGGACTGCCCGCTCTGCCCGGACACGGCTGTAAGTCTTGTGAATTT CATCGCATGAATACTGGAGATAAAGCTGTGTTGTGTGCACTTTGCTATATGAGAGCTTACAACCATTTGTGT TTACAGTAAGTGTGATTAAGTTGAACCTTTAGAGGGAGGCAGAGAGCAGGGTGACTGGGCGATGACTGGTTTT ATTTATGTATATATGTTCTTTATATAGGTCCCGTCTCTGACGCAGATGATGAGACCCCCACTACAAAGTCC ACTTCGTACCCCCAGAAATTGGCACATCTCCACCTGAGAATATTGTTAGACCAGTTCTGTAGAGCCAC TGGGAGGAGAGCAGCTGTGGAATGTTTGGATGACTTGCTACAGGGTGGGGTTGAACCTTTGGACTTGTGTA CCCGGAACGCCCCAGGCACTAAGTGCCACACATGTGTGTTTACTTGAGGTGATGTCAGTATTTATAGGGT GTGGAGTGCAATAAAAAATGTGTTGACTTTAAGTGCGTGGTTTTATGACTCAGGGGTGGGGACTGTGAGTAT ATAAGCAGGTGCAGACCTGTGTGGTTAGCTCAGAGCGGCATGGAGATTTGGACGGTCTTGGAAGACTTTCA CAAGACTAGACAGCTGCTAGAGAACGCCTCGAACGGAGTCTCTTACCTGTGGAGATTCTGCTTCGGTGGCG ACCTAGCTAGGCTAGTCTACAGGGCCAAACAGGATTATAGTGAACAATTTGAGGTTATTTTGAAGAGTGT TCTGGTCTTTTTGACGCTCTTAACCTTGGGCCATCAGTCTCACTTTAACAGAGGATTTCGAGAGCCCTTGA TTTTACTACTCCTGGCAGAACCCTGCAGCAGTAGCCTTTTTTGCTTTTATTCTTGACAAATGGAGTCAAG AAACCCATTTTACGAGGGATTACCAGCTGGATTTCTTAGCAGTAGCTTTGTGGAGAACATGGAAGTGCCAG CGCCTGAATGCAATCTCCGGCTACTTGCCGGTACAGCCGCTAGACACTCTGAGGATCCTGAATCTCCAGGA GAGTCCCAGGGCACGCCAACGTCGCCAGCAGCAGCAGCAGGAGGAGGATCAAGAAGAGAACCCGAGAGCCG GCCTGGACCCTCCGGCGGAGGAGGAGGAGTAGCTGACCTGTTTTCTGAACTGCGCCGGGTGCTGACTAGGT CTTGAGTGGTGCGGAGAGGGGGATTAAAGCGGAGAGGCATGATGAGACTAATCACAGAACTGAACTGACT GTGGGTCTGATGAGTCGCAAGCGCCAGAAACAGTGTGGTGGCATGAGGTGCAGTCGACTGGCACAGATGA GGTGTGCGGTGATGCATGAGAGTTTTCTCTAGAACAAGTCAAGACTTGTTGGTTAGAGCCTGAGGATGATT GGGAGGTAGCCATCAGGAATTATGCCAAGCTGGCTCTGAGGCCAGACAAGAAGTACAAGATTACTAAGCTG ATAAATATCAGAAATGCCTGCTACATCTCAGGGAATGGGGCTGAAGTGGAGATCTGTCTCCAGGAAAGGGT GGCTTTTACAGATGCTGCATGATGAATATGTACCCGGGAGTGGTGGGCATGGATGGGGTTACCTTTTATGAACA TGAGGTTTACGGGGAGATGGGTATAATGGCACGGTCTTTATGGCCAATACCAAGCTGACAGTCCATGGCTGC TCCTTCTTTGGGTTTAATAACACCTGCATCGAGGCCTGGGGTCAGGTGGTGTGAGGGGCTGCAGTTTTTTC AGCCAACTGGATGGGGGTGCTGGGCAGGACCAAGAGTATGCTGTCCGTGAAGAAATGCTTGTGTTGAGAGGT GCCACCTGGGGGTGATGAGCGAGGGCGAAGCCAGAATCCGCCACTGCGCCTCTACCGAGACGGGCTGCTTT GTGCTGTGCAAGGGCAATGCTAAGATCAAGCATAATATGATCTGTGGAGCCTCGGACGAGCGCGGCTACCA GATGCTGACCTGCGCGCGGGGAACAGCCATATGCTGGCCACCGTACATGTGGCTTCCCATTGCTCGCAAGC CCTGGCCCGAGTTCGAGCACAATGTTCATGACCAGGTGCAATATGCATCTGGGGTCCCGCCGAGGCTGCTT ATGCCCTACCAGTGCAACCTGAATTATGTGAAGGTGCTGCTGGAGCCGATGCCATGTCCAGAGTGAGCCT GACGGGGGTGTTTGACATGAATGTGGAGGTGTGGAAGATTCTGAGATATGATGAATCCAAGACCAGGTGCC GAGCCTGCGAGTGCGGAGGGAAGCATGCCAGGTTCCAGCCCGTGTGTGTGGATGTGACGGAGGACCTGCGA CCCGATCATTTGGTGTGTCCTGACCCGGGACGGAGTTCCGGTTCCAGCGGGGAAGAATCTGACTAGAGTGA GTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTT GCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTTCAGCCCTTATCTGACGGGGCGTCTCCCC

TCCTGGGCGGGAGTGCCTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCCGGAACCTC
 TTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGAGCTGCTGCATCTG
 CCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTTCC
 ACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCAGCTCGAGGCCTTGACCCA
 GCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGA
 AATCCAAATAAAAAATGAATCAATAAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTT
 ATTTGATTTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTTCC
 AGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCT
 CCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGT
 GTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTG
 TTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTT
 ACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGTCAGGACCACCAGCACGGTGTATCCGGTGCATCTGGGGA
 ATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCAGGTTTCC
 ATGCACTCATCCATGATGATGGCGATGGGCGCGTGGGCGGCGCTGGGCAAAGACGTTTCGGGGTTCGGA
 CACATCATAGTTGTGGTCTGGGTGAGGTGCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGG
 ACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTG
 AGCTCGGAGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTTCGGGGCGGGGAGATGAG
 CTGGGCGGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGA
 CCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCTCCCGGAGGAGGGGGGCCACCTCGTTCATC
 ATCTCGCGCACGTGCATGTTCTCGCGCACCAAGTTCCGCCAGGAGGCGCTCTCCCCCAGGGATAGGAGCTC
 CTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTGCGCCATGGGCATTTTGGAGAGGGTTTGTGCA
 AGAGTTCCAGGCGGTCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTC
 GCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCCTTC
 CAGGGTGCAGCGTCCGCGTCAGGGTGGTCTCCGTACCGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGC
 GAGGGTGCCTTCAGGCTCATCCGGCTGGTTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCCGCCAGGT
 AGCAATTGACCATGAGTTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAA
 GTCTGCCCCGAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGG
 GCGGTAGGCGTCCGCGCCGAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTCCGGCTGGT
 CCGGGTCAAAAACCAAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGT
 CCCCCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCCCTCGAGCGGTGT
 GCCCGGTCCTCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCAAGG
 AGGCCACGTGGGACGGGTAGCGGTGCTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAACACATG
 TCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAAGTGTAGGCCACGTGACCGGGGGTCCCGGCCGG
 GGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCT
 GTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTGTGAGTTTCTAGAAAACGAG
 GAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCCTCGTCCATCTGGTCAGAAAAGAC
 GATCTTTTTTGTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGC
 GCATGGTCTGGTTTTTTTTCTTGTGCGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCC
 ACGCACTTCCATTCCGGGAAGACGGTGGTCAGCTCGTGGGCACGATTCTGACCTGCCAGCCCCGATTATG
 CAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGC
 CTTTGCAGCAGCAGAAGGGGGGAGGGGGTCCAGCATGACCTCGTGGGGGGGTGCGCATCGATGGTGAAG
 ATGCCGGGCAGGAGGTGCGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTC
 GCGCACGGCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGG
 CGTACATGCCGAGATGTGCTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGC
 CCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGGCCAGGTT
 GGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGG
 GCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGTCT
 TGCACACTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGAGTAGTCGAGGGTCTCTCTGGATGAT
 GTCATACTTGAAGTGTCCCTTTTGTTCACAGCTCGCGGTTGAGAAGGAACCTTCGCGGCTCTCCAGT
 ACTCTTCGAGGGGGAACCCGTCCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCTTG
 TAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCAGAGGAGGTGTGCGTGAG
 GGCGAAAGTGTCCCTGACCATGACCTTGAGGAACTGGTGCTTGAAGTCGATATCGTCGAGCCCCCTGCT
 CCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGAGG
 ATCTTGCCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGAT
 GACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCACGATGTAGAGTTCCACGAATC
 GCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAAGTCTCTCGTAGGTGAGCTCGTGGGGTTCGTGAGCCCG
 TGCTGCTCGAGCGCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGTAAGTCCAGAGATCCACGGCCAG

GCGGGTTTGCAGACGGTCCCGGTACTGACGGAAGTCTGCCCCGACGGCCATTTTTTCGGGGGTGACGCAGT
 AGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGACG
 AGCCGGTTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATCCA
 GGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAACT
 GGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCGAA
 CACTCGTGCTTGTGTTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACGAG
 CTGTACCTGAGTTCCTTTGACGAGGAATTTAGTGGGAAAGTGGAGTCTGTGGCGCCTGCATCTCGTGCTGTA
 CTACGTGCTGGTGGTTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTTCATGCTGACGAGCCCGCGCGGGAGG
 CAGGTCCAGACCTCGGCGCGAGCGGGTTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGGGT
 CCTGAGACGCTGCGGAGTCAGGTCACTGGGACGCGCGCGCGCGGTTGACTTGACAGGAGTTTTTCCAGGG
 CGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTGATGGCTTGACAGGGT
 CCGTGCCCTTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGCGGTGCCTC
 TTCCATGGTTAGAAAGCGCGCGAGGACGCGCGCGGGCGGCGAGGGCGGCTCGGGGCCCGGAGGCAGGGG
 CGGCAGGGGCACGTGCGCGCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGCGA
 CGACGCGACGGTTGACGTCTGGATCTGACGCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACCTG
 AAAGAGAGTTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCTGCCGACGATCTCTTGACGTGCGC
 CGAGTTGTCTGGTAGGCGATCTCGGTCACTGACTGCTGATCTCCTCCTCTTGAAGGTCTCCGCGGCCGG
 CGCGCTCCACGGTGGCGCGAGGTGCTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTTCATGCCCGCC
 TCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCGGGCGCGCATGACCACCTGGGCGAGGTT
 GAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGCGA
 TGTGCTCGGTGACGAAGAAATACATGATCCAGCGCGGAGCGGCATCTCGCTGACGTGCGCCAGCGCCTCC
 AAACGTTCCATGGCCTCGTAAAGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGAGACGGTCAA
 CTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCGGGAGTT
 CCTCCACTTCTCTTCTTCTCCTCCTCCACTAACATCTCTTCTACTTCTCCTCAGGCGGCAGTGGTGGCGGG
 GGAGGGGGCCTGCGTTCGCCGGCGGCGCACGGGACAGCGGTGATGAAGCGCTCGATGGTCTCGCCGCGCCG
 GCGTTCGATGGTCTCGGTGACGGCGCGCCCGTCTCGCGGGGCGCAGCGTGAAGACGCCCGCGCGCATCT
 CCAGGTGGCCGGGGGGTCCCCGTGGGACAGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCGTA
 GGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCGAG
 CCAGTCGAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCGGG
 CGATGCTGCTGGTGTGAAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACAGGTCT
 TTGGGCCCGGCTTGTGATGCGCAGACGGTTCGGCCATGCCCCAGGCGTGGTCTGACACCTGGCCAGGTC
 CTTGTAGTAGTCTGCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCGCTGCATGCGCGTGA
 GCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTGCGCGACGACGCGCTCGGCGAGGATGGCTTGCTGG
 ATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTGATGGTGTA
 GGAGCAGTTGGCCATGACGGACAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAGGC
 GCGAGTAGGCGCGCTGTGCAAGATGTAGTCTGTCAGGTGCGCACAGGTAAGTGGTAGCCGATGAGGAAG
 TGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCTCGAGCAT
 GGTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGGGA
 ACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTTCATGGTGGGCACGGTCTGGCCCGTG
 AGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAACGAAAGCGGTGAGCGGCTCGACTCCGTGGCCTG
 GAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTTCAATCTCGAATCAGGCTGGAGCCGAGCT
 AACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCTGTTT
 TGCAACTTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGGCCGACCGCGATGGCTCGCTGCC
 GTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTTCGAGGCCGGCGGATTCCGCGGC
 TAACGAGGGCGTGGCTGCCCCGTGTTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCGAG
 CCCCTCTTTTGTGTTTTGTTTTGTTTGGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCTCC
 ACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCGCCCCAGCAGCAACTTCCAGCCACGACCG
 CCGCGGCGCGCTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGAAGAGGGCGAGGGCTG
 CGCGCCTTGGGGGCTGCTGCGCGGAGCGGACCCGCGCTGCGAGATGAAAAGGGACGCTCGCGAGCCCTA
 CGTGCCCAAGCAGAACCTGTTGAGAGACAGGAGCGGCGAGGAGCCGAGGAGATGCGCGCGGCCCGGTTCC
 ACGCGGGGCGGGAGCTGCGGCGCGGCTGGACCGAAAGAGGGTGTGAGGGACGAGGATTTTCGAGGCGGAC
 GAGCTGACGGGGATCAGCCCCGCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAGAC
 CGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCACCTGATCGCGCGCGAGGAGG
 TGACCCTGGGCTGATGCACCTGTGGGACCTGCTGGAGGCCATCGTGAGAACCCACCAGCAAGCCGCTG
 ACGGCGCAGCTGTTCTGTTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTTCAGGGAGGCGCTGCTGAATAT
 CACCGAGCCCGAGGGCGGCTGGCTCCTGGACCTGGTGAACATTCTGACAGAGCATCGTGGTGCAGGAGCGCG
 GGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTGGGCAAGTACTACGCTAGG

AAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGAC
 CCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCG
 CCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCGGG
 ACCGAGGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCTTGAGGGC
 GGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGAT
 GGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCTGC
 AGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACG
 ACCCGCAACCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGT
 GCCCTCGCGCTCCAACCCACGCACGAGAAGGTCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCA
 TCCGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCAAC
 GTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCAGCGCGAGCGGTTCCACCG
 CGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCTCAGCACCCAGCCGCCAACGTGCCCCGGG
 GCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTG
 TACCAGTCCGGGCGGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGACAGACCGTGAACCTGAGCCAGGC
 TTTCAAGAACTTGAGGGCCTGTGGGGCGTGAGGCCCGGTCGGGGACCGCGCGACGGTGTGAGCCTGC
 TGACGCCGAACCTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTACGACAGCGGCAGCATCAACCGCAAC
 TCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTA
 CCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAACT
 TTTTGCTGACCAACCGGTGCGAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTG
 CGTTACGTGCAGCAGAGCGTGGGCCTGTTCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACAT
 GACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTTCATCAATAAACTGATGGACTACT
 TGCATCGGGCGGCCGCCATGAACTCTGACTATTTACCAACGCCATCCTGAATCCCCACTGGCTCCCGCG
 CCGGGTTCTACACGGGCGAGTACGACATGCCCCACCCCAATGACGGGTTCTGTGGGACGATGTGGACAG
 CAGCGTGTCTCCCCCGACCGGGTGCTAACGAGCGCCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCC
 CGTCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCTTTCCCGAGC
 TTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGA
 AGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAA
 GCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTCGCAG
 GGGGCCACGAGCCGGGCGAGCGCCGCCCGTAAACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGGGA
 CGATGAGGACTCCGCCGACGACAGCAGCGTGTGGACTTGGGTGGGAGTGGTAACCCGTTTCGCTCACCTGC
 CCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTACCAAGGCCATGGCGACCAG
 CGTGCGTTTCGTTTCTCTCTGTTGTTGTTGTTATCTAGTATGATGAGGCGTGCATACCCGAGGGTCTCTCT
 CCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGCTCCTTA
 CGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTTCGTTACTCGGAGCTGGCACCCCTTGT
 ACGATAACCACCCGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGAC
 CACAGCAACTTCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCACCCAGACCAT
 CAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGTGA
 ACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACA
 GTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGG
 CAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGC
 GGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACCTTCAGGCTGGGCTGGGAC
 CCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCC
 CGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTTCGAAGAGGCAGCCCTTCC
 AGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTGACGCC
 TATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTGAG
 GGGCGATAATTTTGCAAGCGCCGCGAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCA
 TTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTAC
 CCGACCTGGTACCTAGCCTACAACATATGGCAGCCCCGAGAAGGGCGTGCGCTCCTGGACGCTGTCTACCA
 CTCGAGCTCACCTGCGGCGTGAGCAAGTCTACTGGTTCGCTGCCCCGACATGATGCAAGAACCCGCTCACCT
 TCCGCTCCACGCGTCAAGTTAGCAACTACCCGCTGTTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGAGC
 TTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTACCTCGCTTACGCACGTCTTCAACCG
 CTTCCCCGAGAACCAGATCCTCGTCCGCCCCGCGCGCCACCATTACCACCGTCAGTGAACACGTTCTCTG
 CTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGAC
 GCCAGACGCCGACCTGCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCTCTCGAGCCG
 CACCTTCTAAATGTCCATTCTCATCTCGCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCAGCAAGA
 TGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCAGCGTGCAGGGGCACTTCCGCGCTCCCTGG
 GGCGCCCTCAAGGGCCGCGTGCGGTGCGGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGC

GCGCAACTACACCCCGCCGCGCGCCGCTCTCCACCGTGACGCCGTCATCGACAGCGTGGTGGCcGACG
 CGCGCCGGTACGCCCCGCCAAGAGCCGGCGGCGGCATCGCCCGCGGCACCGGAGCACCCCGCCATG
 CGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACG
 CGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGGCGGCAGCGGCCATCGCCA
 GCATGTCCCCCGCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGGCCGTG
 CGCACCCGCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGGAGGATG
 TCCAAGCGCAAATTCAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAA
 GGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACG
 GATTGGTGGAGTTTGTGCGCGAGTTTCGCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCG
 GTGCTGAGACCCGGCACCCCGTGGTCTTCACGCCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTA
 CGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGTCTACGGCA
 AGCGCAGCCGTTCCGACCCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCACGCCGAGC
 CTCAAGCCCGTGACCTTGACGAGGTGCTGCCGACCCGCGGCCCGCGCCGCGCGGGGGTTCAAGCGCGAGGGCGA
 GGATCTGTACCCACCATGACGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGA
 AGGTGGACCCGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCGCTGGGCGTG
 CAGACCGTGGACATCAAGATTCCACGAGGCCCATGGAACGCAGACCGAGCCCATGATCAAGCCAGCAC
 CAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACG
 GCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGC
 ACGCGCTTCTACCGCGGTTCATACCAGCAGCCGCCGCGCAAGACCACCACTCGCCGCCGCGCTCGCCGCAC
 CGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGC
 CGCGCGCGCGCTACCACCCGAGCATCGCCATTTAACTTTTCGCTGCTTTGCGATCAATGGCCCTCACAT
 GCCGCTTTCGCGTTCCCATACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGG
 ATGCGTCGCCACCACCACCGGCGGCGGCCATCAGCAAGCGGTTGGGGGGAGGCTTCTGCCCCGCGCT
 GATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGC
 GCCACTGAGACACACTTGGAAACATCTTGTAAATAAACCaATGGACTCTGACGCTCCTGGTCTGTGATGTG
 TTTTTCGTAGACAGATGGAAGACATCAATTTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTTCATGG
 GCACCTGGAGCGACATCGGCACCAGCCAATGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGG
 CTTAAGAATTTTCGGGTCCACGCTTAAACCTATGGCAGCAAGGCGTGGAACAGCACAGGGCAGGCGCT
 GAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTTCGATGGGCTCGCCTCGGGCATCAACGGGG
 TGGTGGACATGGCCAAACAGGCCGTGCAGCGGCAGATCAACAGCCGCTGGACCCGGTGCCGCCCGCGGC
 TCCGTGGAGATGCCGACGGTGGAGGAGAGCTGCCCTCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCC
 CGATGCGGAGGAGACGCTGCTGACGCACACGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGGTC
 TGCCCCACCACGCGGCCCATCGCGCCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCCTGGAC
 TTGCCTCCTCCCCAGCCTTCCCGCCCCCTCTACAGTGGCTAAGCCCCCTGCCGCCGGTGGCCGTGGCCCCGCGC
 GCGACCCGGGGGACCGCCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAG
 TGCAGAGTGTGAAGCGCCGCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTAT
 GTATTATGTCGCCGCCGCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGG
 CCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGT
 CCGGGTCTGGTGCAGTTTGGCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCAC
 GGTGGCGCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACC
 GCGAGGACAACACCTACTCGTACAAAGTGCCTACACGCTGGCCGTGGGCGACAACCGCGTGTGACATG
 GCCAGCACCTACTTTGACATCCGCGGCGTGTGGATCGGGGCCCTAGCTTCAAACCTACTCCGGCACCGC
 CTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTGAGTGGACATATAAAGCCGATGGTGAAACTG
 CCACAGAAAAACCTATACATATGGAATGCACCCGTGCAGGGCATTAAACATCACAAAAGATGGTATTCAA
 CTTGGAAGTACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGGGTGA
 TGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAA
 TGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTGAGGCAAATGTGAAAACAGGA
 ACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGCCT
 AGTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTGGAACCTCCAGATACCCATATTGTATACAAAG
 CAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTGAGCAAGCCATGCCAACAGACCTAATCATATT
 GGTTTTAGAGACAACCTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCA
 GGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCTACCAGCTCTTGCTTG
 ACTCTCTGGGTGACAGAACCCGGTATTTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTG
 CGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAG
 AACAGATACTTATCAGGGAATTAAGGCTAATGGAAGTATGATCAAACCATGGACCAAAGATGACAGTGTCA
 ATGATGCTAATGAGATAGGCAAGGGTAATCCATTTCGCATGGAAATCAACATCCAAGCCAACCTGTGGAGG
 AACTTCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGCCAATGTTACCTT

GCCACCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACA
 TCAACATCGGGGCGCGCTGGTTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCG
 GGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAA
 ATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGGTCTACACCTACGAGTGGAACCTCCGCAAGGACG
 TCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCTCCATCTCCTTCACCAGC
 ATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAA
 CGACACCAACGACCAGTCTTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAACG
 CCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCGGCCTTCCGCGGCTGGTCTTCACGCGTCTC
 AAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTA
 CCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTACGCT
 GGCCCCGCAACGACCGGCTCCTGACGCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTAC
 AACGTGGCCCAGTGCAACATGACCAAGGACTGGTTCCTGGTCCAGATGCTGGCCCCACTACAACATCGGCTA
 CCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTCCGCAACTCCAGCCCCATGA
 GCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTACCCTGGCCTACCAGCACAACAAC
 TCGGGCTTCGTGCGCTACCTCGCGCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACTACCCCTACCC
 GCTCATCGGCAAGAGCGCCGTACCAGCGTCACCCAGAAAAAGTTCTCTGCGACAGGGTCATGTGGCGCA
 TCCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAAC
 TCCGCCCCACGCGCTAGACATGAATTTGGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTCTT
 CGAAGTCTTCGACGTGTCGAGTGACACCAGCCCCACCGCGCGTCATCGAGGCCGTCTACCTGCGCACCC
 CTTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGAGC
 AGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCTTGGGCACCTTCGATAAGCGCTTC
 CCGGGATTTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGA
 GCACTGGCTGGCCTTCGCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGG
 ACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGACGCGCCCTGGCCACCGAG
 GACCGCTGCGTCACCCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTT
 CTGCTGCATGTTCTGTCACGCCTTCGTGCACTGGCCCCGACCGCCCCATGGACAAGAACCCACCATGAACT
 TGCTGACGGGGGTGCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGGAG
 GCGCTCTACCGCTTCCCTCAACTCCCACTCCGCCTACTTTTCGCTCCACCGCGCGCGCATCGAGAAGGCCAC
 CGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCACTT
 TCATGTTACACATCGATCTGAGATGATTTATTTAGAAATCGAAAGGGTCTGCCGGGTCTCGGCATGGCCC
 CGGGGTGTCGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCAGCAGGTTCGGGCGCGG
 AGATCTTGAAATCGCAGTTGGGACCCGCGTTCGTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGG
 AACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTTCGCGTTCGGTGATGCTCTCCACGTGAGGTC
 CTCGGCGTTGGCCATCCCGAAGGGGGTTCATCTTGAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCT
 TGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTTCGGCGTTCATCCCCGGGTACATG
 GCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCGCA
 GGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTTCGTGCACGCAGCAGCGCGCGTCTGTTGTTGCCA
 GCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCCGTTCGGGGTTCTCCTTCAGCGCGCGC
 TGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCA
 CCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCCTCCAGTTCTTGT
 GGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTTCAGGGTCTTGTG
 CTAGTGAAGGTTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTC
 GCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTTCAGGTTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAG
 TCATGATTTCCATACCCTTCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTCTTTCACCATCATCTTA
 GCGCTAGCAGCCGCGGCCAGGGGGTTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCTTCTCGGT
 GATCCGCACCGGGGGGTAGCTGAAGCCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCTCTCGTGT
 CCTGGCTGACGTCTTGCAGGACCACATGCTTGGTCTTGCGGGGTTTCTTCTTGGGCGGCAGCGGGCGGA
 GATGTTGGAGATGGCGAGGGGGAGCGGAGTTCGCTCACCCTACTATCTCTTCTTCTTCTTGGTCCGA
 GGCCACGCGCGGTAGGTATGTCTCTTTCGGGGGAGAGGCGAGGCGAGGGCTCTCGCCGCCGCGACTTG
 GCGGATGGCTGGCAGAGCCCCCTTCCGCGTTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCTCCG
 CGGCCGGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACTCGCCATC
 TGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCGGCCAGCCCCGCCA
 CCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACG
 CCCGCGGAGCACGAGGAGGAGCTGGCAGTGCCTTTTACAAGAAGAGATACACCAAGAAGAGCCAGAGCA
 GGAAGCAGAGAATGAGCAGAGTACGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGG
 AGGACGCGCTCATCAAGCATCTGGCCCCGAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAG

GTGCCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCAA
 GCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCCGAGG
 CCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCCGACCCGC
 GCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAA
 GATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGGAACGCTCTGCAAGGAGAAGGAGGAGAGCATG
 AGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAG
 CTGACCCATTTTCGCTACCCGGCTCTGAACCTGCCCCCAAAAGTCATGAGCGCGGTTCATGGACCAGGTGCT
 CATCAAGCGCGCTGCCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCA
 GCGACGAGCAGCTGGCCCCGGTGGCTGGGTCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGGCGAAACTC
 ATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCGGACGCGGAGACCCT
 GCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTCCA
 ACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCCTGGGGCAGAACGTCGCTG
 CACACCACCTGCGCGGGGAGGCCGCGGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACAC
 CTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCC
 TGCAGAAGAACCTCAAGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCTCGGACCTGGCCGACCTC
 ATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCCGACTTTATGAGCCAAAGCATGTTGCA
 AAACCTTCGCTCTTTTCATCCTCGAACGCTCCGGAATCCTGCCCCGACCTGCTCCGCGCTGCCCTCGGACT
 TCGTGCCGCTGACCTTCCGCGAGTGCCCCCGCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAAC
 TACCTGGCCTACCACTCGGACGTGATCGAGGACGTGAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTG
 CAACCTCTGCACGCCGACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCA
 CCTTCGAGTTGCAAGGGCCAGCGAAGGCGAGGGTTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGGGG
 CTGTGGACCTCGGCCTACTTGCACAAGTTCGTGCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGA
 GGACCAATCCCATCCGCCCAAGGCCGAGCTGTGCGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCAAT
 TGCAAGCCATCCAGAAATCCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAG
 ACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGGAGCTGC
 CGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGAC
 TGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGA
 GGAGGTGGAAGAAGCAGCCGCCCGCCAGACCGTCTGCTCCTCGGCGGGGGAGAAAGCAAGCAGCACGGATACCA
 TCTCCGCTCCGGGTCCGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCC
 ACCACCCAGACCGGTGAAGAAGGAGCGGCAGGATACAAGTCTTGGCGGGGGCACAAAACGCCATCGTCTC
 CTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTACCCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAAC
 TTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCCTACTACTTCCAAGAAGAGGCAGCAGCA
 GCAGAAAAAGACCAGCAGAAAAACCAGCAGCTAGAAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCG
 CGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCACCCCTCTATGCCATCTTCCAG
 CAGAGTCGGGGGAGGAGCAGGAACCTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCT
 GTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCG
 CGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCAGTCGCAGAAAAAGGCGGGAATTACGTACCTGTGCCC
 TTCGCCCTAGCCGCTCCACCCATCATCATGAGCAAAGAGATTCCACGCCTTACATGTGGAGCTACCAGC
 CCCAGATGGGCCTGGCCGCCGGTGCCGCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCCC
 GCGATGATCTCACGGGTGAATGACATCCGCGCCACCAGAACAGATACTCCTAGAACAGTCAGCGCTCAC
 CGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCAGGAAATTCCCCAGC
 CCACGACCGTACTACTTCCGCGAGACGCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCG
 GGCGGCGCCACCCTGTGTGCTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGTACCAGGAGGAGGAC
 ACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACCTCGCCGGAT
 CGGGGAGATCTTCCTTACGCCTCGTCAGGCCGTCTGACTTTGGAGAGTTCGTCTCGCAGCCCCGCTCG
 GGTGGCATCGGCACTCTCCAGTTCGTGGAGGAGTTCACCTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTC
 CCCCCGCCACTACCCGGACGAGTTCATCCCGAATTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATT
 GAATGTCCCATGTTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGGACCACTGCCGCCGCTTCCGCTGC
 TTCGCTCGGGATCTCGCCGAGTTTGCTACTTTGAGCTGCCCGAGGAGCACCTCAGGCGCCGGCCACGG
 AGTGCAGATCGTCGTCGAAGGGGGCCTCGACTCCACCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGG
 TCGAGCGCGAGCAAGGACAGACCTTCTGACTCTGTACTGCATCTGCAACCACCCCGGCTGCATGAAAGT
 CTTTGTGTGTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGTTC
 CTGAATCCATCAACCAGTCTTTGTTCTTACCAGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGCCCCAC
 AAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCGACAACGACGGAGT
 CCTGCTGAGCGGCCCTGCCAACCTTACTTTTTTCCACCCGAGAAAGCAAGCTCCAGCTCTTCCAACCTTCC
 TCCCCGGGACCTATCAGTGCGTCTCGGGACCTGCCATCACACCTTCCACCTGATCCCGAATACCACAGCG
 TCGCTCCCCGCTACTAACAACCAACTAACCTCCACCAACGCCACCGTCGCGACCTTTCTGAATCTAATAC

TACCACCCACACCGGAGGTGAGCTCCGAGGTCAACCAACCTCTGGGATTTACTACGGCCCCCTGGGAGGTGG
 TTGGGTAAATAGCGCTAGGCCTAGTTGCGGGTGGGCTTTTGGTTCTCTGCTACCTATACCTCCCTTGCTGT
 TCGTACTTAGTGGTGCTGTGTTGCTGGTTTAAGAAATGGGGAAGATCACCTAGTGAGCTGCGGTGCGCTG
 GTGGCGGTGTTGCTTTTCGATTGTGGGACTGGGCGGTGCGGCTGTAGTGAAGGAGAAGGCCGATCCCTGCTT
 GCATTTCAATCCCAACAAATGCCAGCTGAGTTTTTCAGCCCGATGGCAATCGGTGCGCGGTACTGATCAAGT
 GCGGATGGGAATGCGGAGACGTGAGAATCGAGTACAATAACAAGACTCGGAACAATACTCTCGCGTCCGTG
 TGGCAGCCCCGGGGACCCCGAGTGGTACACCGTCTCTGTCCCCGGTGCTGACGGCTCCCCGCGCACCGTGAA
 TAATACTTTTCATTTTTGCGCACATGTGCGACACGGTCATGTGGATGAGCAAGCAGTACGATATGTGGCCCC
 CCACGAAGGAGAACATCGTGGTCTTCTCCATCGCTTACAGCCTGTGCACGGCGCTAATCACCGCTATCGTG
 TGCCTGAGCATTACATGCTCATCGCTATTCGCCCCAGAAATAATGCCGAAAAAGAAAAACAGCCATAACG
 TTTTTTTTTCACACCTTTTTTCAGACCATGGCCTCTGTAAATTTTTTGCTTTTATTTGCCAGTCTCATTGCCG
 TCATTCATGGAATGAGTAATGAGAAAATTACTATTTTACACTGGCACTAATCACACATTGAAAGGTCCAGAA
 AAGCCACAGAAGTTTCATGGTATTGTTATTTTAAATGAATCAGATGTATCTACTGAACCTCTGTGGAAACAA
 TAACAAAAAAAATGAGAGCATTACTCTCATCAAGTTTCAATGTGGATCTGACTTAACCCTAATTAACATCA
 CTAGAGACTATGTAGGTATGTATTATGGAACACAGCAGGCATTTTCGGACATGGAATTTTATCAAGTTTCT
 GTGTCTGAACCCACCACGCCTAGAATGACCACAACCACAAAACTACACCTGTTACCCTATGCAGCTCAC
 TACCAATAACATTTTTGCCATGCGTCAAATGGTCAACAATAGCACTCAACCCACCCACCCAGTGAGGAAA
 TTCCCAAATCCATGATTGGCATTATTGTTGCTGTAGTGGTGTGCATGTTGATCATCGCCTTGTGCATGGTG
 TACTATGCCTTCTGCTACAGAAAGCACAGACTGAACGACAAGCTGGAACACTTACTAAGTGTGAAATTTTA
 ATTTTTTTAGAACCATGAAGATCCTAGGCCTTTTAAATTTTTTCTATCATTACCTCTGCTCTATGCAATTCTG
 ACAATGAGGACGTTACTGTCTGTTGTGCGGATCAAATTATACACTGAAAGGTCCAGCGAAGGGTATGCTTTTCG
 TGGTATTGCTATTTTTGGATCTGACACTACAGAACTGAATTATGCAATCTTAAGAATGGCAAAATTCAAAA
 TTCTAAAATTAACAATTATATATGCAATGGTACTGATCTGATACTCCTCAATATCACGAAATCATATGCTG
 GCAGTTACACCTGCCCTGGAGATGATGCTGACAGTATGATTTTTTACAAAGTAACTGTTGTTGATCCCACT
 ACTCCACCTCCACCCACCACAACCTACTCACACCACACACACAGATCAAACCGCAGCAGAGGAGGCAGCAAA
 GTTAGCCTTGCAGGTCCAAGACAGTTTCAATTTGTTGGCATTACCCCTACACCTGATCAGCGGTGTCCGGGGC
 TGCTAGTCAGCGGCATTGTGCGGTGTGCTTTCCGGGATTAGCAGTCATAATCATCTGCATGTTTCAATTTTTGCT
 TGCTGCTATAGAAGGCTTTACCGACAAAAATCAGACCCACTGCTGAACCTCTATGTTTAAATTTTTTCCAGA
 GTCATGAAGGCAGTTAGCGCTCTAGTTTTTTGTTCTTTGATTGGCATTGTTTTTTGCAATCCTATTCTCTAA
 AGTTAGCTTTTATTAAAGATGTGAATGTTACTGAGGGGGGCAATGTGACACTGGTAGGTGTAGAGGGTGTG
 AAAACACCACCTGGACAAAATACCACCTCAATGGGTGGAAGATATTTGCAATTGGAGTGTATTAGTTTAT
 ACATGTGAGGGAGTTAATCTTACCATTGTCAATGCCACCTCAGCTCAAAATGGGTAGAATTCAAGGACAAAG
 TGTCAGTGTATCTAATGGGTATTTTTACCCAACATACTTTTATCTATGACGTTAAAGTCATACCACTGCCTA
 CGCCTAGCCACCTAGCACTACCACACAGACAACCCACACTACACAGACAACCACATACAGTACATTAAAT
 CAGCCTACCACCACTACAGCAGCAGAGGTTGCCAGCTCGTCTGGGGTCCGAGTGGCATTTTTTGATGTGGC
 CCCATCTAGCAGTCCCACTGCTAGTACCAATGAGCAGACTACTGAATTTTTGTCCACTGTGAGAGCCACA
 CCACAGCTACCTCCAGTGCCTTCTCTAGCACCGCCAATCTCTCCTCGCTTTCTCTACACCAATCAGTCCC
 GCTACTACTCCTAGCCCCGCTCCTCTTCCCACTCCCTGAAGCAAACAGACGGCGGCATGCAATGGCAGAT
 CACCCTGCTCATTGTGATCGGGTTGGTCATCCTGGCCGTGTTGCTCTACTACATCTTCTGCCGCCGCATTC
 CCAACGCGCACCGCAAGCCGGTCTACAAGCCCATCATTGTGCGGCAGCCGGAGCCGCTTCAGGTGGAAGGG
 GGTCTAAGGAATCTTCTCTTCTCTTTTACAGTATGGTGATTGAACTATGATTCCCTAGACAATTCTTGATCA
 CTATTCTTATCTGCCTCCTCCAAGTCTGTGCCACCCTCGCTCTGGTGGCCAACGCCAGTCCAGACTGTATT
 GGGCCCTTGCCTCCTACGTGCTCTTTGCCTTACCACCTGCATCTGCTGCTGTAGCATAGTCTGCCTGCT
 TATCACCTTCTTCCAGTTTCAATGACTGGATCTTTGTGCGCATCGCCTACCTGCGCCACCACCCCACTACC
 GCGACCAGCGAGTGGCGCGGTGCTCAGGCTCCTCTGATAAGCATGCGGGCTCTGCTACTTCTCGCGCTTC
 TGCTGTTAGTGCTCCCCCGTCCCGTCGACCCCGGTCCCCACCCAGTCCCCGAGGAGGTCCGCAAATGC
 AAATTCGAAGAACCTGGAAATTCCTCAAATGCTACCGCCAAAAATCAGACATGCATCCAGCTGGATCAT
 GATCAATTGGGATCGTGAACATTCTGGCCTGCACCCCTCATCTCCTTTGTGATTTACCCCTGCTTTGACTTTG
 GTTGAACCTCGCAGAGGCGCTCTATCTCCGCTGACCTGACACACCACCACAGCAACCTCAGGACACAC
 GCACTACCACCACTACAGCCTAGGCCACAATACATGCCCATATTAGACTATGAGGCCGAGCCACAGCGACC
 CATGCTCCCCGCTATTAGTTACTTCAATCTAACCGGCGGAGATGACTGACCACTGGCCAACAACAACGTC
 AACGACCTTCTCCTGGACATGGACGGCCGCGCCTCGGAGCAGCGACTCGCCCAACTTCGCATTCGCCAGCA
 GCAGGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGCCATCCACCAGTGCAAGAGAGGCATCTTCTGCC
 TGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAAACGACCATCGCCTCTCCTACGAGCTCCTGCAG
 CAGCGCCAGAAGTTACCTGCCTGGTCGGAGTCAACCCCATCGTCATCACCCAGCAGTCTGGCGATACCAA
 GGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGTCCACACTCTGATCAAGACCCTCTGCGGCCTCC
 GCGACCTCCTCCCCATGAATAATCACCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTAC

AGAAATAAAAAATAATCATTTTGATTTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAACAAAAA
 AATAAGAATCACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCC
 CCTCTTCCCAGCTCTGGTACTGCAGGCCCCGGCGGGCTGCAAACCTTCCCTCCACACGCTGAAGGGGATGTCA
 AATTCCCTCTGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCTCCGGGTGGATGAT
 GACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCCTTCGT
 CTCTTCAGATGGATTCCAAGAGAAGCCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCAACCACCA
 AGAACGGGGAAATCACCCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCCCTCGGGAAAACTCATCTCCAAC
 ACGGCCACCAAGGCCGCCGCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCACCCCTT
 TTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAAATATACTGAGAACAAGCATTCTAA
 ACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCT
 CCACTTACATTTGATACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGA
 TGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTTTAAAAATTGAAGATGGAGCCATAGCAACCAACATTG
 GAAATGGGTTAGAGTTTGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAA
 CTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCAC
 TTTGTGGACAACACCTGATCCATCACCAACTGTCAAATACTCGCAGAAAATGATGCAAACTAACACTTT
 GCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGTGAGTCTTAGTTGTAGGAAGTGGAAACCTAAAC
 CCCATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAAACGGTGTTCTTTTAAACAGA
 ACATTCTACACTAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATATACCAATG
 CTGTAGGATTTCATGCCCAATTTAAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTA
 GGGCAAGTATACATGAATGGAGATGTTTCAAACCTATGCTTCTCACTATAACCCTCAATGGTACTGATGA
 CAGCAACAGTACATATTCAATGTCAATTTTATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTG
 GGGCTAACTCTTATACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCT
 TCCCACCCCACTCTGTGGAACAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAA
 CAGTTTTTACAGGATTCGAGCAGTTATTTTTTCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCC
 CCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTT
 CAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAG
 CTCAACAGCTGAGGATTGTCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGG
 GAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTGCGATCAGGCCCGCAGCAGTCGCTGCCGCCGCCG
 CTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCCACGGCCCTCAGCA
 TCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATCGGGATCTCGCTCAGGTCTGCTGCAAGTACGTCGAACAC
 AGAACCACAGGTTGTTCAACAGTCCATAGTTCAACAGCTCCAGCCGAAACTCATCGCGGGAAGGATGCT
 ACCCAGTGCCCGTCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCACGT
 ACATGATCTCCTTGGGCATGTGGCGGTTTACCACCTCCCGGTACCACATCACCTCTGGTTGAACATGCAG
 CCCCAGATGATCCTGCGGAACCACAGGGCCAGCACCGCCCCCGCCGCGCATGCAGCGAAGAGACCCCGGGTC
 CCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGG
 CACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCCAG
 GGCACGGGGAACCTCTTGCAAGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACCTTACATTGTG
 CATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCT
 CACAGCGTGGAAGGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTGCGGACCGTGTC
 ATGATGCAGTTGCTTTTCGGACATTTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATC
 GCCGGCGCGGTCTCGGCGCTTGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGTGC
 AGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGT
 GGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTGGGTTTTCGGTGACGGCGGGGAGGGAAGAACAG
 GAAGAACCATGATTAACTTTTAAATCCAAACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATGGCAC
 CTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCAC
 GGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTA
 ATTCTCAATCATCATGTTACACTCCTGCACCATCCCCAGATAATTTTCATTTTCCAGCCTTGAATGATT
 CGAAGTGTCTGAGGTAAATCCAAGCCAGCTGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGCAT
 TCTTAAGCACACCTCATAAATTCGAAGATATTCTGCTCTGGTTACCTGCAGCAGATTGACAAGCGGAAT
 ATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCTCTC
 CGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAAACCGAAGTCT
 CCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATA
 ACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTA
 GAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGA
 AAAAACAAAAATGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCAG
 GCAGGCCACGGGGTCTCCGGCGCGACCTCGTAAAAATTGTGCTATGATTGAAAACCATCACAGAGAGAC
 GTTCCCGGTGGCCGGCGTGAATGATTGCACAAGATGAATACACCCCGGAACATTGGCGTCCGCGAGTGAA

AAAAAGCGCCCCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGG
ATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCCTCCTGCACAGGCAGCAAAGCCCCCG
ATCCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAG
AGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGA
CGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAACCGGTGA
CACACTCAAAAAAATACGCGCACTTCTCAAACGCCCAAACTGCCGTCAATTTCCGGGTTCCACGCTACG
TCATCAAAACACGACTTTCAAATTCCGTGCGACCGTTAAAAACGTACCCCGCCCCGCCCTAACGGTCGCCC
GTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAA
GTTTGAGGTATATTATTGATGATGG

ChAdV68.4WTnt.GFP (SEQ ID NO:11); AC_000011.1 con secuencias E1 (nt 577 a 3.403) y E3 (nt 27.816- 31.332) eliminadas; nucleótidos de ATCC VR-594 sustituidos en cuatro posiciones; GFP indicadora bajo control del promotor/potenciador de CMV insertado en el lugar de E1 eliminada

CCATCTTCAATAATATACCTCAAACCTTTTTGTGCGGTTAATATGCAAATGAGGCGTTTGAATTTGGGGAG
GAAGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGACGTGGTT
GCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTTGAACACGGAA
ATACTCAATTTTCCGCGCTCTCTGACAGGAAATGAGGTGTTTTCTGGGCGGATGCAAGTGAAAACGGGCCA
TTTTTCGCGGAAAACCTGAATGAGGAAGTGAAAATCTGAGTAATTTTCGCGTTTATGGCAGGGAGGAGTATTT
GCCGAGGGCCGAGTAGACTTTGACCGATTACGTGGGGGTTTTCGATTACCGTGTTTTTACCTAAATTTCCG
CGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTGACGTGATCGCCAGGGTATTTAAACCTGCGCTCT
CCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACAC
TTTGAAAGTAGGGATAACAGGGTAATGacattgattattgactagttGttaaTAGTAATCAATTACGGGGT
CATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
CCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAA
GTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTACATGACCTTACGG
GACTTTTCTACTTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGgTGATGCGGTTTTGGCAGTA
CACCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCAATGGGA
GTTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCTGTAATAACCCCGCCCCGTTGACGCAAATG
GGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAgcTCGTTTTAGTGAACCGTCAGATCGCCTGGAA
CGCCATCCACGCTGTTTTGACCTCCATAGAAGACAGCGATCGCGccaccATGGTGAGCAAGGGCGAGGAGC
TGTTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
GGCGAGGGCAGGGCGATGCCACCTACGGCAAGCTGACCTGAAGTTCATCTGCACCACCGGCAAGCTGCC
CGTGCCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAG
GACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCT
GAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACA
ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTTCAAGATCCGCCACAACATCGAG
GACGGCAGCGTGACGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCATATGGTCC
TGCTGGAGTTCTGTGACCGCCGCGGGGATCACTCTCGGCATGGACGAGCTtTACAAGTAGtgaGTTTAAACT
CCCATTTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGACAAACCAC
AACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTA
TAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCAATTTTATGTTTTAGGTTTACGGGGGAGATGTGG
GAGGTTTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTAAATAACTATAACGGTCCTAAGGTAGCGAGT
GAGTAGTGTTCTGGGGCGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTTCTGTGTG
TTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTTCAGCCCTTATCTGACGGGGCGTCTCC
CCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCCGGAAC
TCTTCAACCCCTGACCTATGCAACCCTGAGCTCTTCGTGCTTGGACGACGCTGCCGCCGACGCTGCTGCATC
TGCCGCCAGCCTCCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTT
CCACCAATAATCCCGCAGCCTGAACGAGGAGAAGCTGTGCTGCTGATGGCCAGCTCGAGGCCCTTGACC
CAGCGCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGT
GAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAAACACAGAGTCTGAATCT
TTATTTGATTTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTT
CCAGGACCCGGTAGAGGTGGGCTTGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAG
CTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATG
GTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGACGCCCTTTGGTGTAGGTGTTTACAAATC

TGTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATG
 TTACCGCCCAGATCCCGCCTGGGGTTTCATGTTGTGCAGGACCACCAGCACGGTGTATCCGGTGCACCTTGGG
 GAATTTATCATGCAACTTGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCCTTGTGCCCCGCCAGGTTTT
 CCATGCACTCATCCATGATGATGGCGATGGGCCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTGCG
 GACACATCATAGTTGTGGTCCCTGGGTGAGGTGCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCC
 GGACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTT
 TGAGCTCGGAGGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGAGATG
 AGCTGGGCGGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCGAGCCGGTGGGGCCGTAGATGACCCCGAT
 GACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCTCCCGGAGGAGGGGGGCCACCTCGTTCA
 TCATCTCGCGCACGTGCATGTTCTCGCGCACCAAGTTCCGCCAGGAGGCGCTCTCCCCCAGGGATAGGAGC
 TCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTGCGCCATGGGCATTTTGGAGAGGGTTTGTG
 CAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTT
 TCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCTT
 TCCAGGGTTCGACGCTCCGCGTCAGGGTGGTCTCCGTACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTT
 GCGAGGGTGCCTTTCAGGCTCATCCGGCTGGTGCAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAG
 GTAGCAATTGACCATGAGTTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGG
 AAGTCTGCCCCGAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCG
 GGGGCGTAGGCGTCCGCGCCGCGAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTGCGGCTG
 GTCGGGGTCAAAAACCAAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGT
 GTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCTTCGAGCGGT
 GTGCCGCGGTCTCTCTCGTAGAGGAACCCCGCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAA
 GGAGGCCACGTGGGACGGGTAGCGGTGCTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAACACA
 TGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGGCC
 GGGGGGGTATAAAGGGTGCGGGTCCCTGCTCGTCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAG
 CTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTGAGTTTCTAGAAACG
 AGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCCTCGTCCATCTGGTCAGAAAAG
 ACGATCTTTTTGTTGTGCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGA
 GCGCATGGTCTGGTTTTTTTTCTTGTGCGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCG
 CCACGCACTTCCATTTCGGGGAAGACGGTGGTCAGCTCGTCCGGCACGATTCTGACCTGCCAGCCCCGATTA
 TGCAGGGTGTAGAGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGGCGTCC
 CCCCCTGCGCGAGCAGAAGGGGGGCGAGGGGTCCAGCATGACCTCGTTCGGGGGGGTCCGGCATGAGTGTGA
 AGATGCCGGGCGAGGAGGTGCGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCAT
 TCGCGCACGGCCAGCGCGCTCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGA
 GGCGTACATGCCGCGAGATGTGCTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGC
 GCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCCCAGG
 TTGGTGGCACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGT
 GGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGT
 CTTGCAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATG
 ATGTCATACTTGAGCTGTCCCTTTTGTTCACAGCTCGCGGTTGAGAAGGAACCTTTCGCGGTCTTCCA
 GTACTCTTCGAGGGGGAACCCGTCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCT
 TGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCGCAGGGAGGTGTGCGTG
 AGGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACCTGGTGCTTGAAGTCGATATCGTTCGACGCCCCCTG
 CTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGA
 GGATCTTGCCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTG
 ATGACCTGGGCGGCGAGCACGATCTCGTGAAGCCGTTGATGTTGTGGCCCACGATGTAGAGTTCCACGAA
 TCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTGGGGTTCGCTGAGCC
 CGTGCTGCTCGAGCGCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCC
 AGGGCGGGTTGCGAGCGGTCCCAGTACTGACGGAAGTCTGCCCCAGCGCCATTTTTTCGGGGGTGACGCA
 GTAGAAGGTGCGAGGGGTCCCCGTGCCAGCATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGA
 CGAGCCGGTTCGTCGCCGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATC
 CAGGTGTAGGTTTTCCACATCGTAGGTGAGGAAGAGCCTTTTCGGTTCGAGGATGCGAGCCGATGGGGAAGAA
 CTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCGG
 AACACTCGTGCTTGTGTTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACG
 AGCTGTACCTGAGTTCTTTGACGAGGAATTTTCAGTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGTGCTG
 TACTACGTGCTGGTGGTTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTTCATGCTGACGAGCCCCGCGCGGGA
 GGCAGGTCCAGACCTCGGCGCGAGCGGGTTCGGAGAGCGAGGACGAGGGCGCGCAGGCCGGAGCTGTCCAGG
 GTCCTGAGACGCTGCGGAGTCAGGTGAGTGGGACGCGGCGCGCGGTTGACTTGCAGGAGTTTTTCCAG

GGGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGCGACGTGATGGCTTGCAGGG
 TCCCGTGCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGCGGTGCC
 TCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCCGGGCGGCAGGGGCGGCTCGGGGCCCCGAGGCAGG
 GGCGGCAGGGGCACGTGCGGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGC
 GACGACGCGACGGTTGACGTCTGGATCTGACGCCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACC
 TGAAAGAGAGTTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGAGGATCTCTTGACGTCG
 CCCGAGTTGTCTGGTAGGCGATCTCGGTATGAACCTGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCC
 GGCGCGCTCCACGGTGGCCGCGAGGTGCTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTCATGCCCG
 CCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCgGGCGCGCATGACCACCTGGGCGAGG
 TTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGC
 GATGTGCTCGGTGACGAAGAAATACATGATCCAGCGCGGAGCGGCATCTCGCTGACGTGCGCCAGCGCCT
 CCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGAGACGGTC
 AACTCCTCCTCCAGAAGACGGATGAGCTCGGCATGGTGGCGCGCACCTCGCGCTCGAAGGCCCGGGAG
 TTCTCTCACTTCTCTTCTTCTCTCTCTCCACTAACATCTCTTCTACTTCTCTCTCAGGCGGCAGTGGTGGCG
 GGGGAGGGGGCCTGCGTCGCGCGGCGGCGCACGGGAGACGGTTCGATGAAGCGCTCGATGGTCTCGCCGCGC
 CGGCGTCGCATGGTCTCGGTGACGGCGCGCCCGTCTCGCGGGGCCGAGCGTGAAGACGCCGCCGCGCAT
 CTCCAGGTGGCCGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCG
 TAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCG
 AGCCAGTCGCGAGTCGCAAGGTAGGCTGAGCACGGTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCG
 GCGCATGCTGCTGGTGATGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACAGGT
 CTTTGGGCCCCGGCTTGCTGGATGCGCAGACGGTTCGGCCATGCCCCAGGCGTGGTCTGACACCTGGCCAGG
 TCCTTGTAGTAGTCTGTCATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCCGTGTCATGCGCGT
 GAGCCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTGCGCGACGACGCGCTCGGCGAGGATGGCTTGCT
 GGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGTTGATGGTG
 TAGGAGCAGTTGGCCATGACGGACAGTTGACGGTCTGGTGGCCCCGGACGCACGAGCTCGTGGTACTTGAG
 GCGCGAGTAGGCGCGCTGTGAAGATGTAGTCGTTGACAGGTGCGCACCAAGGTACTGGTAGCCGATGAGGA
 AGTGCGGCGGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCTCGAGC
 ATGGTGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGG
 GAACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTTCATGGTGGGCACGGTCTGGCCCCG
 TGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTGAGCGGCTCGACTCCGTGGCC
 TGGAGGCTAAGCGAACGGTTGGGCTGCGCGTATACCCGGTTCAATCTCGAATCAGGCTGGAGCCGAG
 CTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTCGT
 TTTGCAACTTTTTTTTTGGAGGCCGATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTG
 CCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTTCGAGGCCGGCCGGATTCCGCG
 GCTAACGAGGGCGTGGCTGCCCGTCTGTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCG
 AGCCCCCTCTTTTGTGTTTGTGTTTGTGTTTGTGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCT
 CCACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGAC
 CGCCGCGGCCCGCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGAAGAGGGCGAGGGGC
 TGGCGCGCCTGGGGGCGTCTGTCGCCGAGCGGCACCCGCGCGTGCAGATGAAAAGGGACGCTCGCGAGGCC
 TACGTGCCCAAGCAGAACCTGTTTACAGAGACAGGAGCGGCGAGGAGCCGAGGAGATGCGCGCGGCCCGGTT
 CCACGCGGGGCGGGAGCTGCGGCGCGGCCTGGACCGAAAGAGGGTGCTGAGGGACGAGGATTTTCGAGGCGG
 ACGAGCTGACGGGGATCAGCCCCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAG
 ACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCACCCCTGATCGCGCGCGAGGA
 GGTGACCCTGGGCCTGATGCACCTGTGGGACCTGCTGGAGGCCATCGTGAGAACCCACCAGCAAGCCGC
 TGACGGCGCAGCTGTTTCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTTCAGGGAGGCGCTGCTGAAT
 ATCACCGAGCCCGAGGGCCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCG
 CGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTA
 GGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATCGCGCATG
 ACCGTGAAAGTGTGACCTGAGCGACGATCTGGGGGTGATCCGCAACGACAGGATGCACCTGCGCTGAG
 CGCCAGAGGCGGCGGAGCTGAGCGACCAAGGAGTGCATAGTCTGCAGCGGGCCCTGACCGGGGCCG
 GGACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCGCTTGGAG
 GCGGCGGAGGACCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTG
 ATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCATGCGGGCGGCGCT
 GCAGAGCCAGCCGTCCGGCATTAACCTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGA
 CGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTG
 GTGCCCTCGCGCTCCAACCCACGCACGAGAAGGTCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGC
 CATCCGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGCACCA

ACGTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCAC
 CGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCCGCAACGTGCCCCG
 GGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGG
 TGTACCAGTCCGGGCCGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGCAGACCGTGAACCTGAGCCAG
 GCTTTCAAGAACTTGAGGGCCTGTGGGGCGTGACGGCCCCGGTCGGGGACCGCGCGACGGTGTGAGCCT
 GCTGACGCCGAACCTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCAACCGCA
 ACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACC
 TACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAA
 CTTTTTGCTGACCAACCGGTGCGAGAAGATCCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCC
 TGCGTTACGTGCAGCAGAGCGTGGGCCTGTTCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGAC
 ATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTTCATCAATAAACTGATGGACTA
 CTTGCATCGGGCGGCCGCCATGAACTCTGACTATTTACCAACGCCATCCTGAATCCCCACTGGCTCCCGC
 CGCCGGGGTTCTACACGGGCGAGTACGACATGCCCCGACCCCAATGACGGGTTCCTGTGGGACGATGTGGAC
 AGCAGCGTGTTCCTCCCCCGACCGGGTGCTAACGAGCGCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACG
 CCCGTCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCGAGGCCGCCAGTCCTTTCCCGA
 GCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGC
 GAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGA
 AAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTCGC
 AGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCCGTGGCACGACAGGCAGCGGGGACAGATGTGG
 GACGATGAGGACTCCGCCGACGACAGCAGCGTGTGGACTTGGGTGGGAGTGGTAACCCGTTTCGCTCACCT
 GCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACC
 AGCGTGCCTTCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCCTACCCGGAGGGTCCTC
 CTCCCTCGTACGAGAGCGTGTATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCGCTGGAGGCTCCT
 TACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTTCGTTACTCGGAGCTGGCACCCCTT
 GTACGATACCAACCCGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACG
 ACCACAGCAACTTCCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCACCCAGACC
 ATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGT
 GAACGAGTTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGA
 CAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCCGAA
 GGAACCTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGG
 GCGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACCTTCAGGCTGGGCTGGG
 ACCCCGTAACCGAGCTGGTTCATGCCCCGGGGTGATACCAACAGAGGCTTTCCATCCCGATATTGTCTTGCTG
 CCCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTTCGAAGAGGCAGCCCTT
 CCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGCAACATCCCCGCGCTCCTGGATGTGACG
 CCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTC
 AGGGGCGATAATTTTGCAAGCGCCGCGAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGT
 CATTACGCCGTTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCT
 ACCGCAGCTGGTACCTAGCCTACAACCTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACC
 ACCTCGGACGTACCTGCGGCGTGGAGCAAGTCTACTGGTGCCTGCCCGACATGATGCAAGACCCGGTCAC
 CTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGA
 GCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACGTCTTCAAC
 CGCTTCCCCGAGAACCAGATCCTCGTCCGCCCCGCCGCGCCACCATTACCACCGTCAGTGAACACGTTCC
 TGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTG
 ACGCCAGACGCCGCACCTGCCCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCTCTCGAGC
 CGCACCTTCTAAATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCAGCAA
 GATGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCAGCGTGCGCGGGCACTTCCGCGCTCCCT
 GGGGCGCCCTCAAGGGCCGCGTGCGGTGCGGCACCAACCGTCGACGACGTGATCGACCAGGTGGTGGCCGAC
 GCGCGCAACTACACCCCCGCGCGCGCCCGTCTCCACCGTGACGCGCGTCATCGACAGCGTGGTGGCCGA
 CGCGCGCGGTACGCGCGCCAAGAGCCGCGCGCGCGCATCGCCCGGCGGCACCGGAGCAGCAGCCCGCCA
 TGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGA
 CGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGAGACGCGCGGCCACGGCGGCGGCAGCGGCCATCGC
 CAGCATGTCCCGCCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCGGCCACCGGTGTGCGCGTGCCG
 TGCGCACCCGCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCAGCGGCGAGGAGGA
 TGTCCAAGCGCAAATTCAAGGAAGAGATGCTCCAGGTTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTG
 AAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAGGAAGAAGAAAGTGTGTGGA
 CGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAAC
 CGGTGCTGAGACCCGGCACCCCGTGGTCTTCACGCCCCGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCC

TACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGTCTTACGG
 CAAGCGCAGCCGTTCCGCAACGAAGGAAGAGGCGGTGTCCATCCCCGTGGACCACGGCAACCCACGCCGA
 GCCTCAAGCCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCCGCGCCGGGGGTTCAAGCGCGAGGGC
 GAGGATCTGTACCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCAT
 GAAGGTGGACCCGACGTGCAGCCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCTGGGCG
 TGCAGACCGTGGACATCAAGATTCCACAGGAGCCCATGGAACGCAGACCGAGCCCATGATCAAGCCCAGC
 ACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTA
 CGGCGCGGCCAGCCTGCTGATGCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCG
 GCACGCGCTTCTACCGCGGTTCATACCAGCAGCCGCCGCCGAAGACCACCACTCGCCGCCGCCGTGCGCGC
 ACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCT
 GCCGCGCGCGCGCTACCACCCGAGCATCGCCATTTAACTTTTCGCTTCTGCTTTCGAGATCAATGGCCCTCAC
 ATGCCGCTTTCGCGTTCCCATTTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACG
 GGATGCGTCGCGACACACCGCGCGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCCCTGCCCGCG
 CTGATCCCCATCATCGCCGCGCGCATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCA
 GCGCCACTGAGACACACTTGGAAACATCTTGAATAAACCaaATGGACTCTGACGCTCCTGGTCCTGTGATG
 TGTTTTTCGTAGACAGATGGAAGACATCAATTTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTTCAT
 GGGCACCTGGAGCGACATCGGCACCAGCCAATGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCG
 GGCTTAAGAATTTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACACAGGGCAGGCG
 CTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTTCGATGGGCTCGCCTCGGGCATCAACGG
 GGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGGCGCCCGCG
 GCTCCGTGGAGATGCCGCGAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGC
 CCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCGCTACGAGGAGGCGGTGAAACTGGG
 TCTGCCACACACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCCTGG
 ACTTGCCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCCTGCCGCCGGTGGCCGTGGCCCCG
 GCGCGACCCGGGGGACCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGG
 AGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGT
 ATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGAT
 GGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGA
 GTCCGGGTCTGGTGCAGTTTGGCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCC
 ACGGTGGCGCCACGACAGATGTGACCACCGACCGCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGGA
 CGCGAGGACAACACCTACTCGTACAAAGTGCGCTACACGCTGGCCGTGGGCGACAACCCGCTGGGACA
 TGGCCAGCACCTACTTTGACATCCGCGGCGTGTGGATCGGGGCCCTAGCTTCAAACCTACTCCGGCAC
 GCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTGAGTGGACATATAAAGCCGATGGTGAAAC
 TGCCACAGAAAAAACCTATACATATGGAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTC
 AACTTGGAACCTGACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGGGT
 GATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAA
 AATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTGAGGCAAATGTGAAAACAG
 GAACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCAGGCTGCTGCTGGC
 CTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGGAACTCCAGATACCCATATTGTATACAA
 AGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTGAGCAAGCCATGCCAACAGACCTAACTACA
 TTGGTTTTAGAGACAACCTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGT
 CAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCTTACCAGCTCTTGCT
 TGACTCTCTGGGTGACAGAACCCGGTATTTTCACTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATG
 TGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGC
 AGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACCTGATCAAACCACATGGACCAAAGATGACAGTGT
 CAATGATGCTAATGAGATAGGCAAGGGTAATCCATTCGCCATGGAAATCAACATCCAAGCCAACCTGTGGA
 GGAACCTTCTCTACGCCAACGTGGCCCTGTACCTGCCGACTCTTACAAGTACACGCCGGCCAATGTTACC
 CTGCCCACCAACCAACACCTACGATTACATGAACGGCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTA
 CATCAACATCGGGGCGCGTGGTTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCAACCGCAATG
 CGGGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAG
 AAATTTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGCTCTACACCTACGAGTGGAACTTCCGCAAGGA
 CGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCTCCATCTCCTTACCA
 GCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAAACAGGCCTCCACGCTCGAGGCCATGCTGCGC
 AACGACACCAACGACCAGTCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAA
 CGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCGGCCTTCCGCGGCTGGTCTTTCACGCGTC
 TCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCC
 TACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTACG

CTGGCCCCGGCAACGACCGGCTCCTGACGCCCCAACGAGTTGAAATCAAGCGCACCGTCGACGGCGAGGGCT
 ACAACGTGGCCCAAGTGAACATGACCAAGGACTGGTTCTGGTCCAGATGCTGGCCCACTACAACATCGGC
 TACCAGGGCTTCTACGTGCCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCAT
 GAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTACCCCTGGCCTACCAGCACAACA
 ACTCGGGCTTCGTGCGCTACCTCGCGCCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACCTACCCCTAC
 CCGCTCATCGGCAAGAGCGCCGTACCCAGCGTCACCCAGAAAAAGTTCTCTGCGACAGGGTCATGTGGCG
 CATCCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCA
 ACTCCGCCCCACGCGCTAGACATGAATTTTGAAGTCGACCCCATGGATGAGTCCACCCTTCTCTATGTTGTC
 TTGAAGTCTTCGACGTGTCGAGTGACACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCAC
 CCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGA
 GCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCTGGGCACCTTCGATAAGCGCT
 TCCCGGGATTTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGC
 GAGCACTGGCTGGCCTTCGCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTCTC
 GGACGAGCGCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGAGCGCCCTGGCCACCG
 AGGACCGCTGCGTCACCCCTGGAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCTGCGGGCTC
 TTCTGCTGCATGTTCTGTCACGCCTTCGTGCACTGGCCCCGACCGCCCCATGGACAAGAACCCACCATGAA
 CTTGCTGACGGGGGTGCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGG
 AGGCGCTCTACCGCTTCTCAACTCCCACTCCGCTACTTTGCTCCACCGCGCGCGCATCGAGAAGGCC
 ACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCAC
 TTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTCTGCGGGTCTCGGCATGGC
 CCGCGGGCAGGGACACGTTGCGGAACCTGGTACTTGGCCAGCCACTTGAACCTCGGGGATCAGCAGTTTGGGC
 AGCGGGGTGTCGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCAGCAGGTGCGGGCGC
 GGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTCAGCACT
 GGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTTCGCGTCGGTGATGCTCTCCACGTGAGG
 TCCTCGGCGTTGGCCATCCCGAAGGGGGTTCATCTTGACAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGG
 CTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTTCGGCGTTTCATCCCCGGGTACA
 TGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCG
 CAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTCGTGCACGCAGCAGCGCGCGTCTGTTGTTGGC
 CAGCTGCACACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTTCGGGGTCTCTCTTCAGCGCGC
 GCTGCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGG
 CACCGACCTTGCCCTCGGCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACCTCCAGTTCTT
 GTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTCAGGGTCTTGT
 TGCTAGTGAAGGTGAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACC
 TCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTTCAAGTTCGGTCTCCACGCGGTAGCGGTCCATCAGCAT
 AGTCATGATTTCCATACCCCTTCTCCAGGCCGAGACGATGGGCAGGCTCATAGGGTCTTTCACCATCATCT
 TAGCGCTAGCAGCCGCGGCCAGGGGGTTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCTTCTCG
 GTGATCCGCACCGGGGGGTAGCTGAAGCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCTCTCGCT
 GTCCTGGCTGACGTCTTGCAGGACCACATGCTTGGTCTTGGCGGGTCTTCTTCTTGGGCGGCAGCGGCGGCG
 GAGATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCCTACTATCTCTTCTCTTCTTGGTCC
 GAGGCCACGCGCGGTAGGTATGTCTCTTTCGGGGGCGAGGCGGAGGCGACGGGCTCTCGCCGCCGCGACT
 TGGCGGATGGCTGGCAGAGCCCTTCCGCGTTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCCTC
 CGCGGCCGGCCATTGTGTTCTCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCA
 TCTGCCCCCACC GCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCCGCCAGCCCCGC
 CACCTCCGACGCGGCCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGA
 CGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCCTTTTCAAGAAGAGATACACCAAGAAGAGCCAGAG
 CAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGG
 GGAGGACGCGCTCATCAAGCATCTGGCCCCGAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCG
 AGGTGCCCTTCAGCTGGAGGAGCTCAGCCGCGCTACGAGTTGAACCTCTTCTCGCCGCGCGTGGCCCCC
 AAGCGCCACCCCAATGGCACCTGCGAGCCCCAACCCGCGCTCAACTTCTACCCGGTCTTCGCGGTGCCGA
 GGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCGCGCCAACCGCACCC
 GCGCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCTACCTGATATCGCCTCCTTGAAGAGGTTCCC
 AAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGGAACGCTCTGCAAGGAGAAGGAGGAGAGCA
 TGAGCACACAGCGCCCTGGTTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTTCG
 AGCTGACCCATTTTCGCTACCCGGCTCTGAACCTGCCCCCAAAGTCATGAGCGCGGTTCATGGACCAGGTG
 CTCATCAAGCGCGCTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGT
 CAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCTAATGCTAGTCCCCAGAGTTTGAAGAGCGGCGCAAAC
 TCATGATGGCCGTGGTCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGAGACC

CTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTC
 CAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAACGTGC
 TGCACACCACCCTGCGCGGGGAGGCCCCGGCGGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCAC
 ACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCT
 CCTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCCTCGGACCTGGCCGACC
 TCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCCGACTTTATGAGCCAAAGCATGTTG
 CAAAACCTTTGCTCTTTTCATCCTCGAACGCTCCGGAATCCTGCCCCGCCACCTGCTCCGCGCTGCCCTCGGA
 CTTCTGTGCCGCTGACCTTCCGCGAGTGCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCA
 ACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTGAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGC
 TGCAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGG
 CACCTTCGAGTTGCAAGGGCCCAGCGAAGGCGAGGGTTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGG
 GGCTGTGGACCTCGGCCCTACTTGCGCAAGTTCTGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTAC
 GAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTCCGGCTGCGTCAATCACCAGGGGGCGATCCTGGCCCA
 ATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCC
 AGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCCGAGGAAACAAGAAGCTGAAAGTGAGCT
 GCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAG
 ACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAG
 GAGGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCTGCTCTCGGCGGGGGAGAAAGCAAGCAGCACGGATAC
 CATCTCCGCTCCGGGTCCGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAACC
 CCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCTTGGCGGGGGCACAAAAACGCCATCGTC
 TCCTGCTTGCAGGCCTGCGGGGGCAACATCTCTTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAA
 CTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCCTACTACTTCCAAGAAGAGGCAGCAG
 CAGCAGAAAAAGACCAGCAGAAAAACCAGCAGCTAGAAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGAT
 CGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCACCCCTCTATGCCATCTTCC
 AGCAGAGTCGGGGGAGGAGCAGGAACCTGAAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGT
 CTGTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTG
 CGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCAGTCGCAGAAAAAGGCGGGAATTACGTACCTGTGC
 CCTTCGCCCTAGCCGCCTCCACCCATCATCATGAGCAAAGAGATTCCACGCCTTACATGTGGAGCTACCA
 GCCCCAGATGGGCCTGGCCGCCGGTGCCGCCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGC
 CCGCGATGATCTCACGGGTGAATGACATCCGCGCCACCAGAAACCAGATACTCCTAGAACAGTCAGCGCTC
 ACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCGGAAATTTCCCA
 GCCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGG
 CGGGCGGCGCCACCCTGTGTGCTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAGAGGC
 ACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACCTCGCCGG
 ATCGGGGAGATCTTCTTTCACGCCTCGTCAGGCCGTCTGACTTTGGAGAGTTCTGCTCTCGCAGCCCCGCT
 CGGGTGGCATCGGCACTCTCCAGTTCTGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGC
 TCCCCCGGCCACTACCCGGACGAGTTTCATCCCGAATTCGACGCCATCAGCGAGTCGGTGGACGGCTACGA
 TTGAATGTCCCATGGTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGGACCACTGCCGCCGCTTCCGCT
 GCTTCGCTCGGGATCTCGCCGAGTTTGCCTACTTTGAGCTGCCCGAGGAGCACCCCTCAGGGCCCCGGCCAC
 GGAGTGCGGATCGTCGTGCAAGGGGGCCTCGACTCCACCTGCTTCGGATCTTCAGCCAGCGTCCGATCCT
 GGTGAGCGCGAGCAAGGACAGACCCTTCTGACTCTGTACTGCATCTGCAACCACCCCGGCCTGCATGAAA
 GTCTTTGTTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGGACTTCCGTGTGT
 TCCTGAATCCATCAACCAGTCTTTGTTCTTACCAGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAGCCCC
 ACAAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCGACAACGACGGA
 GTCCTGCTGAGCGGCCCTGCCAACCTTACTTTTTCCACCCGAGAAAGCAAGCTCCAGCTCTTCCAACCCCTT
 CCTCCCCGGGACCTATCAGTGCGTCTCGGGACCCTGCCATCACACCTTCCACCTGATCCCGAATACCACAG
 CGTCGCTCCCCGCTACTAACAACCAAACTAACCTCCACCAACGCCACCGTCGCGACGGCCACAATACATGC
 CCATATTAGACTATGAGGCCGAGCCACAGCAGCCCATGCTCCCCGCTATTAGTTACTTCAATCTAACCGGC
 GGAGATGACTGACCCACTGGCCAACAACACGTCAACGACCTTCTCTGGACATGGACGGCCGCGCTCGG
 AGCAGCGACTCGCCCAACTTCGCATTTCGCCAGCAGCAGGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTG
 GCCATCCACCAGTGCAAGAGAGGCATCTTCTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCC
 AAACGACCATCGCCTCTCCTACGAGCTCCTGCAGCAGCGCCAGAAGTTACCTGCCTGGTGGAGTCAACC
 CCATCGTCATCACCCAGCAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGC
 GTCCACACTCTGATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATACCCCCCTTATCC
 AGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATAAAGAT
 ACAATCATATTGATGATTTGAGTTTAAACAAAAAATAAAGAATCACTTACTTGAAATCTGATACCAGGTCT
 CTGTCCATGTTTTCTGCCAACACCACCTTCACTCCCCTCTTCCAGCTCTGGTACTGCAGGCCCCGGCGGGC

TGCAAACCTTCCTCCACACGCTGAAGGGGATGTCAAATTCCTCCTGTCCCTCAATCTTCATTTTATCTTCTA
 TCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCCGTCTACCCCTACGATGCAGACAACG
 CACCGACCGTGCCCTTCATCAACCCCCCTTCGTCTCTTCAGATGGATTCCAAGAGAAGCCCCCTGGGGGTG
 TTGTCCCTGCGACTGGCCGACCCCGTCACCACCAAGAACGGGGAAATCACCTCAAGCTGGGAGAGGGGGT
 GGACCTCGATTCTCGGGAAAACCTCATCTCCAACACGGCCACCAAGGCCGCCGCCCTCTCAGTTTTTCCA
 ACAACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTTCT
 CCACCATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTTAGGACT
 CCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAAGCTTA
 CCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTTTA
 AAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTGAAGCAGTAGTACAGAAAC
 AGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCA
 TAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACACCTGATCCATCACCAAACTGTCAA
 ATACTCGCAGAAAATGATGCAAACTAACACTTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGT
 GTCAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCAATTACTGGCACCGTAAGCAGTGCTCAGGTGTTTC
 TACGTTTTGATGCAAACGGTGTTCTTTTAACAGAACATTCTACACTAAAAAATACTGGGGGTATAGGCAG
 GGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTTCATGCCCAATTTAAAAGCTTATCCAAA
 GTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAAACCTA
 TGCTTCTCACTATAACCCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCATTTTCATACACC
 TGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACATCGCCCAAGA
 ATGAACACTGTATCCACCCCTGCATGCCAACCCCTTCCACCCCACTCTGTGGAACAAACTCTGAAACACAA
 AATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTTACAGGATTCGAGCAGTTATTTTTCTCCAC
 CCTCCAGGACATGGAATACACCACCCTCTCCCCCGCACAGCCTTGAACATCTGAATGCCATTGGTGATG
 GACATGCTTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAA
 ACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAGGATTGTCTCGGTGGTCGGGATCA
 CGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTC
 GCATCAGGCCCCCGCAGCAGTGCCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGG
 GACTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCG
 GATCTCGCTCAGGTGCTGTCAGTACGTGCAACACAGAACCACAGGTTGTTCAACAGTCCATAGTTCAACA
 CGCTCCAGCCGAAACTCATCGCGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATC
 AAGTGGTGCCCCCTCCAGAACACGCTGCCCACGTACATGATCTCCTTGGGCATGTGGCGGTTTACCACCTC
 CCGGTACCACATCACCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGCCCG
 CCCC GCCGCCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCG
 TGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCAC
 TCTCAACTCCTCGGGGGTCAAAACCATATCCAGGGCACGGGGAACCTCTTGAGGACAGCGAACCCCGCAG
 AACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGA
 TCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATACGGGTGATG
 GCGGGACGCGGCTGATCGTGTTTCGCGACCGTGTCTGATGCAGTTGCTTTTCGGACATTTTTCGTACTTGCTG
 TAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGCGGGCTCTCGGCGCTTGAACGCTCGGTGTT
 GAAATTGTAAAACAGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCAT
 CATGCCTGATGGCTCTGATCACATCGACCACCGTGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGT
 TGGGTTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACTTTTAATCCAAACGGTCTCG
 GAGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGCTGTGTTGGTGGAAAATAACAGCCA
 GGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGA
 AACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCCTCAATCATGTTTACACTCCTGCACCATCCC
 CAGATAATTTTCATTTTTCCAGCCTTGAATGATTGCAACTAGTTTCGTGAGGTAAATCCAAGCCAGCCATGA
 TAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAAGATATTCTGCT
 CCTGGTTACCTGCAGCAGATTGACAAGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTC
 AGCAATAACTGTAAGTACTCTTTTCATATCCTCTCCGAAATTTTGTAGCCATAGGACCACAGGAATAAGATT
 AGGGCAAGCCACAGTACAGATAAACCGAAGTCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAG
 CATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAATCGCCAGGCAATTTTTTAAGAAAA
 TCAACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGT
 GCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGC
 GAACAGGTGGGTAAATCGTTCTCTCCAGCACAGGCAGGCCACGGGGTCTCCGGCGCGACCCCTCGTAAAAA
 TTGTGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGATTGCACAAGATGA
 ATACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAGCGCCCCGAGGAAGCAATAAGGCACTACAATGC
 TCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAA
 TTAATCCCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCAT

AGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGC
TCTCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAAT
CACACACGCCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCTCAAACGCCC
AAAAGTGCCTGTCATTTCCGGGTTCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTA
AAAACGTACCCGCCCCGCCCTAACGGTCGCCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTC
AAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG

ChAdV68.4WTnt.MAG25mer (SEQ ID NO:12); AC_000011.1 con secuencia E1 (nt 577 a 3.403) 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

CCATCTTCAATAATATACCTCAAACCTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTGGGGAG
GAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGACGTGGTT
GCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTGAACACGGAA
ATACTCAATTTTCCCGCGCTCTCTGACAGGAAATGAGGTGTTTTCTGGGCGGATGCAAGTGAAAACGGGCCA
TTTTTCGCGCGAAAACCTGAATGAGGAAGTGAAAATCTGAGTAATTTTCGCGTTTATGGCAGGGAGGAGTATTT
GCCGAGGGCCGAGTAGACTTTGACCGATTACGTGGGGGTTTTCGATTACCGTGTTTTTACCTAAATTTCCG
CGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTGACGTGATCGCCAGGGTATTTAAACCTGCGCTCT
CCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACAC
TTTGAAAGTAGGGATAACAGGGTAATgacattgattattgactagttGttaaTAGTAATCAATTACGGGGT
CATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCG
CCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCCTTGGCAGTACATCAAGTGTATCATATGCCAA
GTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTACGG
GACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGgTGATGCGGTTTTTGGCAGTA
CACCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA
GTTTTGTTTTTGGCACCAAAATCAACGGGACTTTTCCAAAATGTTCGTAATAACCCCGCCCCGTTGACGCAATG
GGCGGTAGGCGTGACGGTGGGAGGTCTATATAAGCAGAgcTCGTTTTAGTGAACCGTCAGATCGCCTGGAA
CGCCATCCACGCTGTTTTGACCTCCATAGAAGACAGCGATCGCGccaccATGGCCGGGATGTTCCAGGCAC
TGTCCGAAGGCTGCACACCCTATGATATTAACCAGATGCTGAATGTCCTGGGAGACCACCAGGTCTCTGGC
CTGGAGCAGCTGGAGAGCATCATCAACTTCGAGAAGCTGACCGAGTGGACAAGCTCCAATGTGATGCCTAT
CCTGTCCCCACTGACCAAGGGCATCCTGGGCTTCGTGTTTTACCCTGACAGTGCCTTCTGAGCGGGGCTGT
TTTGTGACAGCGAGGAGCAGCAACCACACAGGAGTCCGCCAATCTGGGCGAGGAGATCTGTCTCAGCTG
TACCTGTGGCCCCGGGTGACATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGGAGAGCCAAGTA
CAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGCTGTTTCGGCTACCCCGTGATGCTGTTTCGGCGATTGCG
TGCAGGGCGACTGGGATGCCATCCGGTTTTAGATACTGCGCACACCCTGGATATGCACTGCTGAGGTGTAAC
GACACCAATTATTCCGCCCTGCTGGCAGTGGGCGCCCTGGAGGGCCCTCGCAATCAGGATTGGCTGGGCGT
GCCAAGGCAGCTGGTGACACGCATGCAGGCCATCCAGAACGCAGGCCTGTGCACCCTGGTGGCAATGCTGG
AGGAGACAATCTTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCGGCCCAAGACAAACATCATC
GTGGATTCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGTTTGTGGACTGGGAGAACGTGAG
CCCAGAGCTGAATTCACCGATCAGCCATTCTGGCAGGCAGGAATCCTGGCAAGGAACCTGGTGCCTATGG
TGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCCAGAGCCTGGTCATCAGCGCCTCCATCATCGTG
TTTAACCTGCTGGAGCTGGAGGGCGACTATCGGGACGATGGCAACGTGTGGGTGCACACCCCACTGAGCCC
CAGAACACTGAACGCCTGGGTGAAGGCCGTGGAGGAGAAGAAGGGCATCCCAGTGCACCTGGAGCTGGCCT
CCATGACCAATATGGAGCTGATGTCTAGCATCGTGACCAGCAGGTGAGGACATACGGACCCGTGTTTCATG
TGCCTGGGAGGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGCTGTTTCAG
AGCCGCCCAGCTGGCCAACGATGTGGTGCTGCAGATCATGGAGCTGTGCGGAGCAGCCTTTTCGCCAGGTGT
GCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGTGAACAATGAGACAACACAGCCTCAG
ATCGCCAACTGTAGCGTGTACGACTTCTTCGTGTGGCTGCCTACTATAGCGTGAGGGATACCTGTGGCC
CCGCGTGACATACCAATGAATAAGTACGCCTATACATGCTGGAGAGGCGCGCCAAGTATAAGAGAGGCC
CTGGCCAGGCGCAAGTTTGTGGCAGCATGGACCCTGAAGGCCGCCCGGCCGCCCGCCCGCCAGTAT
ATCAAGGCTAACAGTAAGTTTATTGGAATCACAGAGCTGGGACCCGGACCTGGATAATGAGTTTAAACTCC
CATTTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAA
CTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATA
AGCTGCAATAAACAAGTTAACAACAACAATTGCATTCAATTTTATGTTTCAGGTTTCAGGGGAGATGTGGGA
GGTTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGGTAGCGAGTGA
GTAGTGTTCTGGGGCGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTGTT

GCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTACGCCCTTATCTGACGGGGCGTCTCCCC
 TCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCCGCAACTC
 TTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGAGCTGCTGCATCTG
 CCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACCTCGAGTTCC
 ACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCCAGCTCGAGGCCTTGACCCA
 GCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGTGA
 AATCCAAATAAAAAATGAATCAATAAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCTTT
 ATTTGATTTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTTCC
 AGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAGCT
 CCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATGGT
 GTTGACACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATCTG
 TTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATGTT
 ACCGCCAGATCCCGCTGGGGTTCATGTTGTGCGAGGACCACAGCACGGTGTATCCGGTGCACCTGGGGA
 ATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCAGGTTTTCC
 ATGCACTCATCCATGATGATGGCGATGGGCCCCTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGGGTGCGA
 CACATCATAGTTGTGGTCTGGGTGAGGTCAATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCCGG
 ACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCCTCACAGATCTGCATCTCCCAGGCTTTG
 AGCTCGGAGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCGGGGCGGGGAGATGAG
 CTGGGCCGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGATGA
 CCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCTCCCGAGGAGGGGGGCCACCTCGTTTCATC
 ATCTCGCGCACGTGCATGTTCTCGCGCACAGTTCCGCCAGGAGGCGCTCTCCCCCAGGGATAGGAGCTC
 CTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTGCGCCATGGGCATTTTGGAGAGGGTTTGTGCA
 AGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTTTC
 GCGGGTTGGGACGGCTGCGGGAGTAGGGCACAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCTTTC
 CAGGGTGCAGCGTCCGCGTCAGGGTGGTCTCCGTACGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTTGC
 GAGGGTGCCTTCAGGCTCATCCGGCTGGTTCGAAAACCGTCCCGATCGGCGCCCTGCGCGTGGGCCAGGT
 AGCAATTGACCATGAGTTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGGAA
 GTCTGCCCCGAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGACTCGGG
 GCGGTAGGCGTCCGCGCCGCAAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTGCGGCTGGT
 CCGGCTCAAAAACAGATTTCCCGCCGTTCTTTTGTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGTGT
 CCGCGTGGGTGACAAAGAGGCTGTCCGTGTCCCGTAGACCGACTTTATGGGCCGGTCTCTCGAGCGGTGT
 GCCGCGGTCTCTCGTAGAGGAACCCCGCCACTCCGAGACGAAAGCCGGGTCCAGGCCAGCACGAAGG
 AGGCCACGTGGGACGGGTAGCGGTGCTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAACACATG
 TCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAAGTGTAGGCCACGTGACCGGGGGTCCCGGCCGG
 GGGGTATAAAAGGGTGGGGTCCCTGCTCGTCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAGCT
 GTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTGAGTTTCTAGAAACGAG
 GAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCTCGTCCATCTGGTCAGAAAAGAC
 GATCTTTTTTGTGTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGAGC
 GCATGGTCTGGTTTTTTTTCTTGTGCGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCGCC
 ACGCACTTCCATTGCGGGGAAGACGGTGGTCAGCTCGTGGGCGACGATTCTGACCTGCCAGCCCCGATTATG
 CAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCCGC
 CCTTGCGCGAGCAGAAGGGGGGCGAGGGGTCCAGCATGACCTCGTGGGGGGGTGCGCATCGATGGTGAAG
 ATGCCGGGCGAGGAGGTGGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCATTC
 GCGCACGGCCAGCGCGCTCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGAGG
 CGTACATGCCGCGAGATGTGCTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGCGC
 CCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGCCAGGTT
 GGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGTGG
 GCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGATGCT
 TGCAGCTTGGCGCAGACTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCTGGATGAT
 GTCATACTTGAGCTGTCCCTTTTGTTCACAGCTCGCGGTTGAGAAGGAACCTCTCGCGGTCTTCCAGT
 ACTCTTCGAGGGGGAACCCGTCTGATCTGCACGGTAAGAGCCTAGCATGTAGAAGTGGTTGACGGCCTTG
 TAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCAGAGGAGGTGTGCGTGAG
 GGCGAAAGTGTCCCTGACCATGACCTTGAGGAAGTGGTGCTTGAAGTCGATATCGTCGACGCCCCCTGCT
 CCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTGGGCAAAGCGAAAGTAACATCGTTGAAGAGG
 ATCTTGCCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTGAT
 GACCTGGGCGGCGAGCACGATCTCGTCAAGCCGTTGATGTTGTGGCCACGATGTAGAGTTCCACGAATC
 GCGGACGGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGGTCTGTCAGCCCCG

GGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTAGG
 AAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGGTTTTACATGCGCATGAC
 CCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAGCG
 CCAGCAGGCGGCGCGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCCGGG
 ACCGAGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCCCTTGGAGGC
 GGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTGAT
 GGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCTGC
 AGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGACG
 ACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTGGT
 GCCCTCGCGCTCCAACCCACGCACGAGAAGGTCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGCCA
 TCCGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCCGCTACAACAGCACCAAC
 GTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCACCG
 CGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCTCAGCACCCAGCCCGCAACGTGCCCCGGG
 GCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGGTG
 TACCAGTCCGGGCCGACTACTTCTTCCAGACCAGTCGCCAGGGCTTGAGACCGTGAACCTGAGCCAGGC
 TTTCAAGAACTTGAGGGCCTGTGGGGCGTGAGGCCCGGTCGGGGACCGCGCGACGGTGTGAGCCTGC
 TGACGCCGAACCTCGCGCCTGCTGCTGCTGGTGGCCCCCTTACGACAGCGGCAGCATCAACCGCAAC
 TCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACCTA
 CCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAACT
 TTTTGCTGACCAACCGGTGCGAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCCTG
 CGTTACGTGCGAGCAGAGCGTGGGCCTGTTCTGATGCAGGAGGGGGCCACCCCCAGCGCCGCGCTCGACAT
 GACCGCGCGCAACATGGAGCCAGCATGTACGCCAGCAACCGCCCGTTTCATCAATAAACTGATGGACTACT
 TGCATCGGGCGGCCGCCATGAACTCTGACTATTTACCAACGCCATCCTGAATCCCCACTGGCTCCCGCCG
 CCGGGGTTCTACACGGGCGAGTACGACATGCCCCACCCCAATGACGGGTTCTGTGGGACGATGTGGACAG
 CAGCGTGTCTCCCCCGACCGGGTGCTAACGAGCGCCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACGCC
 CGTCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCCTTTCCCGAGC
 TTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGCGA
 AGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGAAA
 GCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTGCAG
 GGGGCCACGAGCGGGGCGAGCGCCGCCCGTAAACGCCGCTGGCACGACAGGCAGCGGGGACAGATGTGGGA
 GGTGAGGACTCCGCCGACGACAGCAGCGTGTGGACTTGGGTGGGAGTGTTAACCCTTCGCTACCTGCTGC
 GCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTACCAAGGCCATGGCGACCAG
 CGTGCGTTTCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGAGGGTCTCTCT
 CCCTCGTACGAGAGCGTGATGCAGCAGGCATGGCGGCGGCGGCGATGCAGCCCCGCTGGAGGCTCCTTA
 CGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTTCGTTACTCGGAGCTGGCACCTTGT
 ACGATAACACCCGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACGAC
 CACAGCAACTTCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGGAGGCCAGCACCCAGACCAT
 CAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGTGA
 ACGAGTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGACA
 GTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAAGG
 CAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGGGC
 GGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTTGACACTAGGAACCTTCAGGCTGGGCTGGGAC
 CCCGTGACCGAGCTGGTCATGCCCGGGGTGTACACCAACGAGGCTTTCCATCCCGATATTGTCTTGCTGCC
 CGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTTCGCAAGAGGCAGCCCTTCC
 AGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTGACGCC
 TATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTGAG
 GGGCGATAATTTTGCAAGCGCCGAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGTCA
 TTCAGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCTAC
 CGCAGCTGGTACTAGCCTACAACATATGGCAGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACCAC
 CTCGACGTCACCTGCGGCGTGAGCAAGTCTACTGGTTCGCTGCCCGACATGATGCAAGAGCCCGGTACCT
 TCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCCAAGAGC
 TTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTACCTCGCTTACGCACGTCTTCAACCG
 CTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCACCATTACCACCGTCAGTGAACACGTTCTCTG
 CTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTGAC
 GCCAGACGCCGACCTGCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGTCTCTCGAGCCG
 CACCTTCTAAATGTCCATTCTCATCTCGCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCAGCAAGA
 TGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCAGCGTGCGCGGGCACTTCCGCGCTCCTTG

GCGGCCCTCAAGGGCCGCGTGC GGTCGCGCACCACCGTCGACGACGTGATCGACCAGGTGGTGGCCGACGC
 GCGCAACTACACCCCCGCGCCGCGCCCGTCTCCACCGTGGACGCCGTCATCGACAGCGTGGTGGCcGACG
 CGCGCCGGTACGCCCCGCGCCAAGAGCCGGCGGGCGGCATCGCCCCGGCGGCACCGGAGCACCCCCGCCATG
 CGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGACG
 CGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGGCGGCAGCGGCCATCGCCA
 GCATGTCCCCGCGCGGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCGTG
 CGCACCCGCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGGCGAGGAGGATG
 TCCAAGCGCAAATTCAAGGAAGAGATGCTCCAGGTTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTGAA
 GGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGACG
 GATTGGTGGAGTTTGTGCGGAGTTTCGCCCCCGGGCGGCGCGTGCAGTGGCGCGGGCGGAAGGTGCAACCG
 GTGCTGAGACCCGGCACCAACCGTGGTCTTCACGCCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCCTA
 CGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGGCCGAGCGCTGGGCGAGTTTGCTTACGGCA
 AGCGCAGCCGTTCCGCAACGAAGGAAGAGGCGGTGTCATCCCGCTGGACCACGGCAACCCACGCCCGGAC
 CTCAAGCCCGTGACCTTGACAGAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGTTCAAGCGCGAGGGCGA
 GGATCTGTACCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCATGA
 AGGTGGACCCGGACGTGCAGCCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCTGGGCGTG
 CAGACCGTGGACATCAAGATTCCACGGAGCCCATGGAACGCAGACCGAGCCCATGATCAAGCCCAGCAC
 CAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTACG
 GCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCGGC
 ACGCGCTTCTACCGCGGTTCATACCAGCAGCCGCCCGCGCAAGACCACCACTCGCCGCCGCCGTGCGCCGAC
 CGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCTGC
 CGCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTTCGCCtGCTTTGCAGATCAATGGCCCTCACAT
 GCCGCCTTCGCGTTCCCATTAACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACGGG
 ATGCGTGCACCACCACCACCGGCGGGCGGCGGCCATCAGCAAGCGGTTGGGGGGAGGCTTCTGCCCGCGCT
 GATCCCCATCATCGCCGCGGGCATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCAGC
 GCCACTGAGACACACTTGGAACATCTTTGTAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTGATGTG
 TTTTCGTAGACAGATGGAAGACATCAATTTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTTCATGG
 GCACCTGGAGCGACATCGGCACCAGCCAATGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCGGG
 CTTAAGAATTTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACACAGGGCAGGCGCT
 GAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTTCGATGGGCTCGCCTCGGGCATCAACGGGG
 TGGTGGACCTGGCCAACCAAGGCCGTGCAGCGGCAGATCAACAGCCGCTGGACCCGTTGCCGCCCGCGGC
 TCCGTGGAGATGCCGCGAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGCCC
 CGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCGCGTACGAGGAGGCGGTGAAACTGGGTC
 TGCCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCCTGGAC
 TTGCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGTTGGCCGTGGCCCGCGC
 GCGACCCGGGGGACCCGCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGGAG
 TGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTAT
 GTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGATGG
 CCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGAGT
 CCGGGTCTGGTGCAGTTTGCCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCAC
 GGTGGCGCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGACC
 GCGAGGACAACACCTACTCGTACAAAGTGCCTACACGCTGGCCGTGGGCGACAACCGCGTGTGACATG
 GCCAGCACCTACTTTGACATCCGCGGCGTGTGGATCGGGGCCCTAGCTTCAAACCTACTCCGGCACCGC
 CTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTGAGTGGACATATAAAGCCGATGGTGAAACTG
 CCACAGAAAAAACCTATACATATGGAATGCACCCGTGCAGGGCATTAAACATCACAAAAGATGGTATTCAA
 CTTGGAAGTACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGGGTGA
 TGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAAAA
 TGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTGAGGCAATGTGAAAACAGGA
 ACAGGCACTACTAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCAGGCTGTGCTGGCCT
 AGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTATACAAAG
 CAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTGAGCAAGCCATGCCCAACAGACCTAACTACATT
 GGTTTTAGAGACAACCTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGTCA
 GGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCTACCAGCTCTTGCTTG
 ACTCTCTGGGTGACAGAACCCGGTATTTTCAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATGTG
 CGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGCAG
 AACAGATACTTATCAGGGAATTAAGGCTAATGGAAGTATGATCAAACCACATGGACCAAAGATGACAGTGTCA
 ATGATGCTAATGAGATAGGCAAGGGTAATCCATTGCCCATGGAATCAACATCCAAGCCAACCTGTGGAGG

AACTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACCCT
 GCCCACCAACACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTACA
 TCAACATCGGGGCGCGCTGGTTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATGCG
 GGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAGAA
 ATTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCGGTCTACACCTACGAGTGGAACCTCCGCAAGGACG
 TCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCTCCATCTCCTTCACCAGC
 ATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAACACGGCCTCCACGCTCGAGGCCATGCTGCGCAA
 CGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCAACATGCTCTACCCCATCCCGGCCAACG
 CCACCAACGTGCCCATCTCCATCCCCCTCGCGCAACTGGGCGGCCTTCCGCGGCTGGTCCCTTCACGCGTCTC
 AAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCCTA
 CCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTCACT
 GGCCCGGCAACGACCGGCTCCTGACGCCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAGGGCTAC
 AACGTGGCCCATGCAACATGACCAAGGACTGGTTCTTGGTCCAGATGCTGGCCCACTACAACATCGGCTA
 CCAGGGCTTCTACGTGCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCATGA
 GCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTACCCCTGGCCTACCAGCACAACAAC
 TCGGGCTTTCGTGCGCTACCTCGCGCCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACCTACCCCTACCC
 GCTCATCGGCAAGAGCGCCGTACCCAGCGTCACCCAGAAAAAGTTCTCTGCGACAGGGTCATGTGGCGCA
 TCCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCAAC
 TCCGCCCCACGCGCTAGACATGAATTTTGAAGTCGACCCCATGGATGAGTCCACCCCTTCTCTATGTTGTCTT
 CGAAGTCTTCGACGTCGTCCGAGTGCACCAGCCCCACCGCGGCGTCATCGAGGCCGTCTACCTGCGCACCC
 CCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGAAGCCATGGCCGCGGGCTCCGGCGAGC
 AGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCTGCGGACCTTCGATAAGCGCTTC
 CCGGGATTTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGCGA
 GCACTGGCTGGCCTTCGCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTCGG
 ACGAGCGCCTCAAGCAGATCTACCAGTTTCGAGTACGAGGGCCTGCTGCGCCGACGCGCCCTGGCCACCGAG
 GACCGCTGCGTCACCTTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCCTGCGGGCTCTT
 CTGCTGCATGTTTCTGACGCCTTTCGTGCACTGGCCCCGACCGCCCCATGGACAAGAACCCACCATGAACT
 TGCTGACGGGGGTGCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCTGCGCCGCAACCAGGAG
 GCGCTCTACCGCTTCTCAACTCCCACTCCGCCTACTTTTCGTCTCCACCGCGCGCGCATCGAGAAGGCCAC
 CGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGACTT
 TCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTTCTGCGGGTCTCGGCATGGCCC
 GCGGGCAGGACACGTTGCGGAACCTGGTACTTGGCCAGCCACTTGAACCTCGGGGATCAGCAGTTTGGGCAG
 CGGGGTGTCGGGAAGGAGTCCGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTGCGGCGCGG
 AGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAGCACTGG
 AACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTTCGCGTTCGGTGATGCTCTCCACGTCGAGGTC
 CTCGGCGTTGGCCATCCCGAAGGGGGTTCATCTTGCAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGGCT
 TGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTTCGGCGTTTCATCCCCGGGTACATG
 GCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCGCA
 GGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTTCGTGCACGCAGCAGCGCGCTCGTTGTTGGCCA
 GCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTTCGGGGTTCTCCTTCAGCGCGCGC
 TGCCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGGCA
 CCGCAGCTTGCCCTCGGCCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCCTCCAGTTCTTGT
 GGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTTCAGGGTCTTGTG
 CTAGTGAAGGTTCAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACCTC
 GCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTTCAGGTTCGGTCTCCACGCGGTAGCGGTCCATCAGCATAG
 TCATGATTTCCATACCTTCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTCTTTCACCATCATCTTA
 GCGCTAGCAGCCGCGGCCAGGGGGTTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCTCTCTCGGT
 GATCCGACACCGGGGGTAGCTGAAGCCCACGGCCCGCAGCTCCTCCTCGGCTGTCTTTCGCTCTCGCTGT
 CCTGGCTGACGTCCTGAGGACCACATGCTTGGTCTTGGCGGGTTTCTTCTTGGGCGGCGAGCGCGCGGA
 GATGTTGGAGATGGCGAGGGGAGCGAGTCTCGCTCACCCTACTATCTCTTCTCTTCTTGGTCCGA
 GGCCACGCGCGGTAGGTATGTCTCTTTCGGGGGCGAGGCGGAGGCGACGGGCTCTCGCCGCGCGCACTTG
 GCGGATGGCTGGCAGAGCCCTTCCGCGTTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTTCTCCG
 CGGCCGGCCATTGTGTTCTCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCATC
 TGCCCCCACCAGCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCGCCAGCCCCGCCA
 CCTCCGACGCGGCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGACG
 CCCGCGGAGCACGAGGAGGAGCTGGCAGTGCCTTTTACAAGAAGAGATACACCAAGAAGCAGCCAGAGCA
 GGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGGGG

AGGACGCGCTCATCAAGCATCTGGCCCCGGCAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCGAG
 GTGCCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCCAA
 GCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCCGAGG
 CCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCCGC
 GCCGACGCCCTTTTCAACCTGGGTCCCGGCGCCCGCTACCTGATATCGCCTCCTTGGAAGAGGTTCCCAA
 GATCTTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGAGCATG
 AGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCGAG
 CTGACCCATTTTCGCTACCCGGCTCTGAACCTGCCCCCAAAGTCATGAGCGCGGTTCATGGACCAGGTGCT
 CATCAAGCGCGCGTCGCCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGTCA
 GCGACGAGCAGCTGGCCCCGGTGGCTGGGTCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAACTC
 ATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCGGACGCGGAGACCCT
 GCGCAAGGTTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCTGCAAGATCTCCA
 ACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGACGAGAACCGCTGGGGCAGAACGTCGTG
 CACACCACCTGCGCGGGGAGGCCCGGCGGACTACATCCGCGACTGCGTCTACCTCTACCTCTGCCACAC
 CTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCTCC
 TGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTTCGACGAGCGCACCAACCGCCTCGGACCTGGCCGACCTC
 ATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCGACTTTATGAGCCAAAGCATGTTGCA
 AAACCTTTCGCTCTTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGACT
 TCGTGCCGCTGACCTTCCGCGAGTGCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCAAC
 TACCTGGCCTACCACTCGGACGTGATCGAGGACGTGACGGCGAGGGCCTGCTCGAGTGCCACTGCCGCTG
 CAACCTCTGCACGCCGCACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGGCA
 CCTTCGAGTTGCAAGGGCCAGCGAAGGCGAGGGTTTCAGCCGCAAGGGGGGTCTGAAACTCACCCCGGGG
 CTGTGGACCTCGGCCTACTTGCGCAAGTTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTACGA
 GGACCAATCCCATCCGCCCAAGGCCGAGCTGTGGCCTGCGTCATACCCAGGGGGCGATCCTGGCCCAAT
 TGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCCAG
 ACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCGGAGGAAACAAGAAGCTGAAAGTGGAGCTGC
 CGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAGAC
 TGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAGGA
 GGAGGTGGAAGAAGCAGCCGCCGCCAGACCGTCTCTCGCGGGGGGAGAAAGCAAGCAGCACGGATACCA
 TCTCCGCTCCGGGTCCGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCGAACCCC
 ACCACCTACCCGCTAAGAAGGAGCGGCAGGGATACAAGTCTTGCGGGGGGACAAAAACGCCCCTGCTCTC
 CTGCTTGCAAGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAACT
 TTCCCCGCAACATCTTGCACTACTACCGTCACCTCCACAGCCCCCTACTACTTCCAAGAAGAGGCAGCAGCA
 GCAGAAAAAGACCAGCAGAAAAACCAGCAGCTAGAAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGATCG
 CGGCGAACGAGCCGGCGCAACCCGGGAGCTGAGGAACCGGATCTTTCCACCCCTCTATGCCATCTTCCAG
 CAGAGTCGGGGGAGGAGCAGGAAGTGAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAGTTGTCT
 GTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTGCG
 CGCTCACTCTTAAAGAGTAGCCCGCGCCCCGCCAGTCGCAGAAAAAGGCGGGAATTACGTACCTGTGCCC
 TTCGCCCTAGCCGCTCCACCCATCATCATGAGCAAAGAGATTCCACGCCTTACATGTGGAGCTACCAGC
 CCCAGATGGGCCTGGCCGCCGGTGGCGCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGCC
 GCGATGATCTCAGGGTGAATGACATCCGCGCCACCAGAAACCAGATACTCCTAGAACAGTCAGCGCTCAC
 CGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCCGCCCTGGTGTACCAGGAAATTCCCCAGC
 CCACGACCGTACTACTTCCGCGAGACGCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGGCG
 GGCGGCGCCACCCTGTGTGCTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGTACCGGGGAGAGGCAC
 ACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACCTCGCCGGAT
 CGGGGAGATCTTCCTTACGCCTCGTCAGGCCGTCTGACTTTGGAGAGTTCGTCTCGCAGCCCCGCTCG
 GGTGGCATCGGCACCTCTCCAGTTCGTGGAGGAGTTCACCTCCCTCGGTCTACTTCAACCCCTTCTCCGGCTC
 CCCCAGCCACTACCCGAGAGTTCATCCCGAATTCGACGCCATCAGCGAGTCGGTGGACGGCTACGATT
 GAATGTCCTCCATGTTGGCGCAGCTGACCTAGCTCGGCTTCGACACCTGGACCCTGCCCGCTTCCGCTGC
 TTCGCTCGGGATCTCGCCGAGTTTGCTACTTTGAGCTGCCCGAGGAGCACCTCAGGGCCCGGCCACGG
 AGTGCGGATCGTCGTGCAAGGGGGCCTCGACTCCACCTGCTTCGGATCTTCAGCCAGCGTCCGATCCTGG
 TCGAGCGCGAGCAAGGACAGACCCTTCTGACTCTGTACTGCATCTGCAACCACCCCGCCTGCATGAAAGT
 CTTTGTGTGTCTGCTGTGTACTGAGTATAATAAAAGCTGAGATCAGCGACTACTCCGACTTCCGTGTGTTC
 CTGAATCCATCAACCAGTCTTTGTTCTTACCAGGGAACGAGACCGAGCTCCAGCTCCAGTGTAAAGCCCCAC
 AAGAAGTACCTCACCTGGCTGTTCCAGGGCTCCCCGATCGCCGTTGTCAACCACTGCGACAACGACGGAGT
 CCTGCTGAGCGGCCCTGCCAACCTTACTTTTTCCACCCGAGAAAGCAAGCTCCAGCTCTTCCAACCTTCC
 TCCCCGGGACCTATCAGTGCGTCTCGGGACCTGCCATCACACCTTCCACCTGATCCCGAATACCACAGCG

TCGCTCCCCGCTACTAACAACCAAATAACCTCCACCAACGCCACCGTCGCGACGGCCACAATACATGCC
 ATATTAGACTATGAGGCCGAGCCACAGCGACCCATGCTCCCCGCTATTAGTTACTTCAATCTAACCGGCGG
 AGATGACTGACCCACTGGCCAACAACAACGTCAACGACCTTCTCCTGGACATGGACGGCCGCGCCTCGGAG
 CAGCGACTCGCCCAACTTCGCATTTCGCCAGCAGCAGGAGAGAGCCGTCAAGGAGCTGCAGGATGCGGTGGC
 CATCCACCAGTGCAAGAGAGGCATCTTCTGCCTGGTGAAACAGGCCAAGATCTCCTACGAGGTCACTCCAA
 ACGACCATCGCCTCTCCTACGAGCTCCTGCAGCAGCGCCAGAAGTTACCTGCCTGGTCGGAGTCAACCCC
 ATCGTCATCACCCAGCAGTCTGGCGATACCAAGGGGTGCATCCACTGCTCCTGCGACTCCCCCGACTGCGT
 CCACACTCTGATCAAGACCCTCTGCGGCCTCCGCGACCTCCTCCCCATGAACTAATACCCCCCTTATCCAG
 TGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAATAATCATTTGATTTGAAATAAAGATAC
 AATCATATTGATGATTTGAGTTTAACAAAAAATAAAGAATCACTTACTTGAAATCTGATACCAGGTCTCT
 GTCCATGTTTTCTGCCAACACCACTTCACTCCCCTCTTCCCAGCTCTGGTACTGCAGGCCCGGCGGGCTG
 CAAACTTCCTCCACACGCTGAAGGGGATGTCAAATTCTCTCCTGTCCCTCAATCTTCATTTTATCTTCTATC
 AGATGTCCAAAAAGCGCTCCGGGTGGATGATGACTTCGACCCCGTCTACCCCTACGATGCAGACAACGCA
 CCGACCGTGCCCTTCATCAACCCCCCTTCGTCTCTTCAGATGGATTCCAAGAGAAGCCCTGGGGGTGTT
 GTCCCTGCGACTGGCCGACCCCGTCACCACCAAGAACGGGGAAATCACCTCAAGCTGGGAGAGGGGGTGG
 ACCTCGATTCTCGGGAAAACCTCATCTCCAACACGCCACCAAGGCCGCCGCCCTCTCAGTTTTTCCAAC
 AACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAAAGATGGAAAATTATCCTTACAAGTTTCTCC
 ACCATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTTAGGTTTTGGATCAGGTTTAGGACTCC
 GTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTGATACTGATGGAAACATAAAGCTTACC
 TTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGCAACATAAGCTGGGCTAAAGGTTTAAA
 ATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGAGTTTGAAGCAGTAGTACAGAAACAG
 GTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCCTTAGCTTTGACAGTACAGGAGCCATA
 ATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACACCTGATCCATCACCAAACGTCAAAT
 ACTCGCAGAAAATGATGCAAACTAACACTTTGCTTGACTAAATGTGGTAGTCAAATACTGGCCACTGTGT
 CAGTCTTAGTTGTAGGAAGTGGAACCTAAACCCCATTAAGCAGTGCTCAGGTGTTTCTA
 CGTTTTGATGCAAACGGTGTTCTTTTAAACAGAACATTCTACACTAAAAAATACTGGGGGTATAGGCAGGG
 AGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTTCATGCCCAATTTAAAAGCTTATCCAAAGT
 CACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACATGAATGGAGATGTTTCAAACCTATG
 CTTCTCACTATAACCCCTCAATGGTACTGATGACAGCAACAGTACATATTCAATGTCAATTTTCATACACCTG
 GACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTATACCTTCTCATACTCGCCCAAGAAT
 GAACACTGTATCCCACCCTGCATGCCAACCCCTCCCACCCACTCTGTGGAACAACTCTGAAACACAAAA
 TAAAAATAAGTTCAAGTGTTTTATTGATTCAACAGTTTTTACAGGATTTCGAGCAGTTATTTTTCTCCACCC
 TCCCAGGACATGGAATACACCACCCTCTCCCCCGCACAGCCTTGAACATCTGAATGCCATTGGTGATGGA
 CATGCTTTTTGGTCTCCACGTTCCACACAGTTTTCAGAGCGAGCCAGTCTCGGGTCGGTCAGGGAGATGAAAC
 CCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAGGATTGTCTCGGTGGTCGGGATCACG
 GTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCGCGAACGGGATCGGCCGGTGGTGTCG
 ATCAGGCCCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTGCTGCTCAGGGGGTCCGGGTCCAGGGA
 CTCCCTCAGCATGATGCCCACGGCCCTCAGCATCAGTCGTCTGGTGCGGCGGGCGCAGCAGCGCATGCGGA
 TCTCGCTCAGGTGCTGTCAGTACGTGCAACACAGAACCACAGGTTGTTCAACAGTCCATAGTTCAACACG
 CTCCAGCCGAAACTCATCGCGGAAGGATGCTACCCACGTGGCCGTCGTACCAGATCCTCAGGTAAATCAA
 GTGGTGCCCCCTCCAGAACACGCTGCCACGTACATGATCTCCTTGGGCATGTGGCGGTTACACCACCTCCC
 GGTACCACATCACCTCTGGTTGAACATGCAGCCCCGGATGATCCTGCGGAACCACAGGGCCAGCACCGCC
 CCGCCCCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAATGGAGGACCCACCGCTCGTACCCGTG
 GATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCATATGCTCATGCATCTCTTCAGCACTC
 TCAACTCCTCGGGGGTCAAAACCATATCCCAGGGCACGGGGAACCTCTTGAGGACAGCGAACCCCGCAGAA
 CAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTATCGCAATCAGGCAGCACCGGGTGATC
 CTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAAGGGGGCCGGCCGATACGGGTGATGGC
 GGGACCGGGTGATCGGTGTTTCGCGACCGGTGTCATGATGTCAGTTGCTTTTCGGACATTTTCGTACTTGCTGTA
 GCAGAACCTGGTCCGGGCGCTGCACACCGATCGCGCGCGGTCCTCGGCGCTTGAACGCTCGGTGTTGA
 AATTGTAAACAGCCACTCTCTCAGACCGTGCAGCAGATCTAGGGCCTCAGGAGTGATGAAGATCCCATCA
 TGCCTGATGGCTCTGATCACATCGACCACCGTGAATGGGCCAGACCCAGCCAGATGATGCAATTTTGTG
 GGTTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGATTAACCTTTTAATCCAAACGGTCTCGGA
 GTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCGCTGTGTTGGTGGAATAACAGCCAGG
 TCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGCAAAGCCTCCACGCGCACATCCAGAAA
 CAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCTCAATCATCATGTTACACTCCTGCACCATCCCCA
 GATAATTTTTCATTTTTCAGCCTTGAATGATTGCAACTAGTTTCGTGAGGTAAATCCAAGCCAGCCATGATA
 AAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACCCTCATAATTCCAAGATATTCTGCTCC

TGGTTCACCTGCAGCAGATTGACAAGCGGAATATCAAAATCTCTGCCGCGATCCCTGAGCTCCTCCCTCAG
CAATAACTGTAAGTACTCTTTTCATATCCTCTCCGAAATTTTTAGCCATAGGACCACCAGGAATAAGATTAG
GGCAAGCCACAGTACAGATAAACCGAAGTCCTCCCCAGTGAGCATTGCCAAATGCAAGACTGCTATAAGCA
TGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAATCGCCCAGGCAATTTTTAAGAAAATC
AACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTCGGGAACAACGATGAAGTAAATGCAAGCGGTGC
GTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATGAACATTAAACCATGCTAGCCTGGCGA
ACAGGTGGGTAAATCGTTTCTCTCCAGCACCAGGCAGGCCACGGGGTCTCCGGCGCGACCCCTCGTAAAAATT
GTGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCCGGCGTGAATGATTGACAAGATGAAT
ACACCCCCGGAACATTGGCGTCCGCGAGTGAAAAAAGCGCCCCGAGGAAGCAATAAGGCACTACAATGCTC
AGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAATTCTCAGGTGCGTACAAAATGTAATT
ACTCCCCCTCCTGCACAGGCAGCAAAGCCCCGATCCCTCCAGGTACACATACAAAGCCTCAGCGTCCATAG
CTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGGCTGAGCTCTAACCTGTCCACCCGCTC
TCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGCCAAAGTCTAAAAATACCCGCCAAATAATCA
CACACGCCAGCACACGCCCAGAAACCGGTGACACACTCAAAAAAATACGCGCACTTCTCTAAACGCCCAA
AACTGCCGTCATTTCCGGGTTCACGCTACGTCATCAAAACACGACTTTCAAATTCCGTCGACCGTTAAA
AACGTACCCGCCCGCCCCCTAACGGTCGCCCCGTCTCTCAGCCAATCAGCGCCCCGCATCCCCAAATTCAA
ACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATATTATTGATGATGG

ChAdV68.5WTnt.GFP (SEQ ID NO:13); AC_000011.1 con secuencias E1 (nt 577
a 3.403) 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

CCATCTTCAATAATATACCTCAAACCTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTTGGGGAG
GAAGGGCGGTGATTGGTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGACGTGGTT
GCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTTGAACACGGAA
ATACTCAATTTTCCGCGCTCTCTGACAGGAAATGAGGTGTTTTCTGGGCGGATGCAAGTGAAAACGGGCCA
TTTTTCGCGCGAAAACCTGAATGAGGAAGTGAAAATCTGAGTAATTTTCGCGTTTATGGCAGGGAGGAGTATTT
GCCGAGGGCCGAGTAGACTTTGACCGATTACGTGGGGGTTTTCGATTACCGTGTTTTTCACCTAAATTTCCG
CGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTGACGTGATCGCCAGGGTATTTAAACCTGCGCTCT
CCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGATCTACAC
TTTGAAAGTAGGGATAACAGGGTAATGacattgattattgactagttGttaaTAGTAATCAATTACGGGGT
CATTAGTTTATATCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTGACCG
CCCAACGACCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCA
TTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAA
GTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGACCTTACGG
GACTTTTCTACTTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGgTGATGCGGTTTTGGCAGTA
CACCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGA
GTTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGTAATAACCCCGCCCCGTTGACGCAAATG
GGCGGTAGGCGGTGACGGTGGGAGGTCTATATAAGCAGAgcTCGTTTAGTGAACCGTCAGATCGCCTGGAA
CGCCATCCACGCTGTTTTGACCTCCATAGAAGACAGCGATCGCGccaccATGGTGAGCAAGGGCGAGGAGC
TGTTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCC
GGCGAGGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCC
CGTGCCCTGGCCCCACCTCGTGACCACCTGACCTACGGCGTGAGTGCTTCAGCCGCTACCCCGACCACA
TGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAG
GACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCT
GAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACA
ACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCACAACATCGAG
GACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGCTGCTGCC
CGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCC
TGCTGGAGTTCGTGACCGCCCGGGGATCACTCTCGGCATGGACGAGCTtTACAAGTAGTgaGTTTAAACT
CCCATTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCAC
AACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACATTAA
TAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGAGATGTGG
GAGGTTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTAAATAACTATAACGGTCCTAAGGTAGCGAGT
GAGTAGTGTTCTGGGGCGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTGTGTG
TTGCAGCAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTTCAGCCCTTATCTGACGGGGCGTCTCC
CCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCCGTGCAGCCCCGCAAC

TCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTCGTTGGACGCAGCTGCCGCCGAGCTGCTGCATC
 TGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACTCGAGTT
 CCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCAGCTCGAGGCCCTTGACC
 CAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCACGGT
 GAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGAATCT
 TTATTTGATTTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATCTTTT
 CCAGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAGGTAG
 CTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTGTAAATCACCCAGTCATAGCAGGGGCGCAGGGCATG
 GTGTTGCACAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACAAATC
 TGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGCGATG
 TTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGTCAGGACCACCAGCACGGTGTATCCGGTGCACCTGGG
 GAATTTATCATGCAACTTGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCCGCCAGGTTTT
 CCATGCACTCATGATGATGGCGATGGGCCCGTGGGCGGCGGCTGGGCCAAAGACGTTTCGGGGGTGCG
 GACACATCATAGTTGTGGTCCTGGGTGAGGTATCATAGGCCATTTTAATGAATTTGGGGCGGAGGGTGCC
 GGACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCTCACAGATCTGCATCTCCCAGGCTT
 TGAGCTCGGAGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTTCCGGGGCGGGGGAGATG
 AGCTGGGCGGAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCCCGAT
 GACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCTCCCGGAGGAGGGGGGCCACCTCGTTCA
 TCATCTCGCGCACGTGCATGTTCTCGCGCACAGTTCCGCCAGGAGGCGCTCTCCCCCAGGGATAGGAGC
 TCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTGCGCCATGGGCATTTTGGAGAGGGTTTGTG
 CAAGAGTTCCAGGCGGTCCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCTCGTT
 TCGCGGGTTGGGACGGCTGCGGGAGTAGGGCACAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGGTCCT
 TCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTACAGGTGAAGGGGTGCGCGCCGGGCTGGGCGCTT
 GCGAGGGTGCGCTTCAGGCTCATCCGGCTGGTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGGCCAG
 GTAGCAATTGACCATGAGTTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCTTTGG
 AAGTCTGCCCCGAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGCGAGGAAGACGGACTCG
 GGGGCGTAGGCGTCCGCGCCGCGAGTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTGGGGCTG
 GTCGGGGTCAAAAACAGTTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGCTCGT
 GTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCCGGTCTTCGAGCGGT
 GTGCGCGGTTCTCTCGTAGAGGAACCCCGCCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCACGAA
 GAGTCCACGTGGGACGGGTAGCGGTGCTTGTCCACGAGCGGGTCCACCTTTTCCAGGGTATGCAAAACACA
 TGTCCTCCCTCGTCCACATCCAGGAAGGTGATGGCTTGTAAGTGTAGGCCACGTGACCGGGGGTCCCGGCC
 GGGGGGTATAAAGGGTGCGGGTCCCTGCTCGTCTCACTGTCTTCCGGATCGCTGTCCAGGAGCGCCAG
 CTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTGAGTTTCTAGAAACG
 AGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCTCGTCCATCTGGTCAGAAAAG
 ACGATCTTTTTGTTGTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGATGGA
 GCGCATGGTCTGGTTTTTTTTCTTGTGCGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTCGCGCG
 CCACGCACCTTCCATTCCGGGAAGACGGTGGTCAGCTCGTGGGCACGATTCTGACCTGCCAGCCCCGATTA
 TGCAGGGTGATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGCGTCC
 GCCCTTGCAGCAGCAGAAGGGGGGAGGGGGTCCAGCATGACCTCGTGGGGGGGTGCGCATCGATGGTGA
 AGATGCCGGGCAGGAGGTGCGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTGCCAT
 TCGCGCACGGCCAGCGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCGCGGA
 GGCGTACATGCCGAGATGTGCTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAGCAGC
 GCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCGAGGGGGCGAGGAGCCCCGGGGCCAGG
 TTGGTGCGACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGATGGT
 GGGCCTTTGGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAGGAGT
 CTTGACAGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTGGATG
 ATGTCTACTTGTAGCTGTCCCTTTTGTTCACAGCTCGCGGTTGAGAAGGAACCTTCGCGGTCTCTCCA
 GTACTCTTCGAGGGGGAACCCGTCTGATCTGACGCGTAAGAGCCTAGCATGTAGAACTGGTTGACGGCCT
 TGTAGGCGCAGCAGCCCTTCTCCAGGGGAGGGCGTAGGCCCTGGGCGGCCTTGCGCAGGGAGGTGTCGCTG
 AGGGCGAAAGTGTCCCTGACCATGACCTTGAGGAACCTGGTGCTTGAAGTCGATATCGTCGCAGCCCCCTG
 CTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTTGGGCAAAGCGAAAGTAACATCGTTGAAGA
 GGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGTTGTTG
 ATGACCTGGGCGGCGAGCACGATCTCGTCGAAGCCGTTGATGTTGTGGCCACGATGTAGAGTTCCACGAA
 TCGCGGACGGCCCTTGACGTGGGGCAGTTTCTTGTAGCTCCTCGTAGGTGAGCTCGTGGGGTTCGTGAGCC
 CGTGCTGCTCGAGCGCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCACGGCC
 AGGGCGGTTTGCAGACGGTCCCGGTACTGACGGAAGTGTGCCCCGACGGCCATTTTTTCGGGGGTGACGCA

GTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGCTCGA
 CGAGCCGGTTCGTCCCCGGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCCCATC
 CAGGTGTAGGTTTTCCACATCGTAGGTGAGGAAGAGCCTTTTCGGTGCGAGGATGCGAGCCGATGGGGAAGAA
 CTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGCGCCG
 AACACTCGTGCTTGTGTTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTGCACG
 AGCTGTACCTGAGTTCTTTGACGAGGAATTTTCAGTGGGAAGTGGAGTTCGTGGCGCCTGCATCTCGTGCTG
 TACTACGTGCTGGTGGTTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTTCATGCTGACGAGCCCCGCGCGGGA
 GGCAGGTCCAGACCTCGGCGCGAGCGGGTTCGGAGAGCGAGGACGAGGGCGCGCAGGCGCGGAGCTGTCCAGG
 GTCCTGAGACGCTGCGGAGTCAGGTTCAGTGGGCAGCGCGCGCGCGGTTGACTTGCAGGAGTTTTTCCAG
 GGCGCGCGGGAGGTCCAGATGGTACTTGATCTCCACCGCGCCATTGGTGGCGACGTGATGGCTTGCAGGG
 TCCCGTGCCCTGGGGTGTGACCACCGTCCCCCGTTTCTTCTTGGGCGGCTGGGGCGACGGGGCGGTGCC
 TCTTCCATGGTTAGAAGCGGCGGCGAGGACGCGCGCGCGGGCGGAGGGGCGGCTCGGGGCCCCGAGGCGAGG
 GGCGGACGGGGCAGTTCGGCGCCGCGCGCGGGTAGGTTCTGGTACTGCGCCCGGAGAAGACTGGCGTGAGC
 GACGACGCGACGGTTGACGTCTGATCTGACGCTCTGGGTGAAGGCCACGGGACCCGTGAGTTTGAACC
 TGAAAGAGAGTTTCGACAGAATCAATCTCGGTATCGTTGACGGCGGCCTGCCGACGATCTCTTGACGTGCG
 CCCGAGTTGTCTGGTAGGCGATCTCGGTTCATGAATGCTCGATCTCCTCCTCTTGAAGGTCTCCGCGGCC
 GGCGCGCTCCACGGTGGCGCGAGGTGCTTGGAGATGCGGCCCATGAGCTGCGAGAAGGCGTTTCATGCCCG
 CCTCGTTCCAGACGCGGCTGTAGACCACGACGCCCTCGGGATCGCGGGGCGCATGACCACCTGGGCGAGG
 TTGAGCTCCACGTGGCGCGTGAAGACCGCGTAGTTGCAGAGGCGCTGGTAGAGGTAGTTGAGCGTGGTGGC
 GATGTGCTCGGTGACGAAGAAATACATGATCCAGCGGCGGAGCGGCATCTCGCTGACGTGCGCCAGCGCCT
 CCAAACGTTCCATGGCCTCGTAAAAGTCCACGGCGAAGTTGAAAACTGGGAGTTGCGCGCCGAGACGGTC
 AACTCCTCCTCCAGAAGACGGATGAGCTCGGCGATGGTGGCGCGCACCTCGCGCTCGAAGGCCCGGGAG
 TTCCTCCACTTCTTCTTCTCCTCCTCCACTAACATCTCTTCTACTTCTCCTCCTCAGGCGGCAGTGGTGGCG
 GGGGAGGGGGCCTGCGTCGCCGGCGGCGCACGGGCAGACGGTTCGATGAAGCGCTCGATGGTCTCGCCGCGC
 CGGCGTTCGATGGTCTCGGTGACGGCGCGCCCGTCTCGCGGGGCGCGAGCGTGAAGACGCCGCCGCGCAT
 CTCCAGGTGGCCGGGGGGTCCCCGTTGGGCAGGGAGAGGGCGCTGACGATGCATCTTATCAATTGCCCCG
 TAGGGACTCCGCGCAAGGACCTGAGCGTCTCGAGATCCACGGGATCTGAAAACCGCTGAACGAAGGCTTCG
 AGCCAGTCGAGTCGCAAGGTAGGCTGAGCACGGTTTTCTTCTGGCGGGTCATGTTGGTTGGGAGCGGGGCG
 GGCGATGCTGCTGGTGTGAAGTTGAAATAGGCGGTTCTGAGACGGCGGATGGTGGCGAGGAGCACAGGT
 CTTTGGGCCCCGCTTGTGATGCGCAGACGGTCGGCCATGCCCCAGGCGTGGTCTGACACCTGGCCAGG
 TCCTTGTAGTAGTTCCTGATGAGCCGCTCCACGGGCACCTCCTCCTCGCCCGCGCGGCGTGCATGCGCGT
 GAGCCGAAGCCGCGCTGGGGCTGGACGAGCGCCAGGTTCGGCGACGACGCGCTCGGCGAGGATGGCTTGCT
 GGATCTGGGTGAGGGTGGTCTGGAAGTCATCAAAGTCGACGAAGCGGTGGTAGGCTCCGGTGGTGGTG
 TAGGAGCAGTTGGCCATGACGGACAGTTGACGGTCTGGTGGCCCGGACGCACGAGCTCGTGGTACTTGAG
 GCGCGAGTAGGCGCGCTGTGAAGATGTAGTCTGTCAGGTGCGCACAGGTAAGTGGTAGCCGATGAGGA
 AGTGGCGGCGCGGCTGGCGGTAGAGCGGCCATCGCTCGGTGGCGGGGGCGCCGGGCGCGAGGTCTCGAGC
 ATGGTGGCGGTGGTAGCCGTAGATGTACCTGGACATCCAGGTGATGCCGGCGGCGGTGGTGGAGGCGCGCGG
 GAACTCGCGGACGCGGTTCCAGATGTTGCGCAGCGGCAGGAAGTAGTTTCATGGTGGGCACGGTCTGGCCCG
 TGAGGCGCGCGCAGTCGTGGATGCTCTATACGGGCAAAAACGAAAGCGGTTCAGCGGCTCGACTCCGTGGCC
 TGGAGGCTAAGCGAACGGGTTGGGCTGCGCGTGTACCCCGGTTTCAATCTCGAATCAGGCTGGAGCCGCGAG
 CTAACGTGGTATTGGCACTCCCGTCTCGACCCAAGCCTGCACCAACCCTCCAGGATACGGAGGCGGGTTCGT
 TTTGCAACTTTTTTTTTGGAGGCCGGATGAGACTAGTAAGCGCGGAAAGCGGCCGACCGCGATGGCTCGCTG
 CCGTAGTCTGGAGAAGAATCGCCAGGGTTGCGTTGCGGTGTGCCCCGGTTTCGAGGCCGGCCGGATTCCGCG
 GCTAACGAGGGCGTGGCTGCCCCGTGTTTTCCAAGACCCCATAGCCAGCCGACTTCTCCAGTTACGGAGCG
 AGCCCCCTCTTTTGTGTTTTGTTTTGTTTGGCCAGATGCATCCCGTACTGCGGCAGATGCGCCCCCACCACCT
 CCACCGCAACAACAGCCCCCTCCACAGCCGGCGCTTCTGCCCCCGCCCCAGCAGCAACTTCCAGCCACGAC
 CGCCGCGGCGCCGTGAGCGGGGCTGGACAGAGTTATGATCACCAGCTGGCCTTGAAGAGGGCGAGGGGCG
 TGGCGCGCTTGGGGCGTTCGTGCGCGGAGCGGACCCGCGCGTGCAGATGAAAAGGACGCTCGCGAGGCC
 TACGTGCCCCAAGCAGAACCTGTTTCAGAGACAGGAGCGGCGAGGAGCCGAGGAGATGCGCGCGGCCGCTT
 CCACGCGGGGCGGAGCTGCGGCGCGGCTGGACCGAAAGAGGGTGTGAGGACGAGGATTTTCGAGGCGG
 ACGAGCTGACGGGGATCAGCCCCGCGCGCGCACGTGGCCGCGGCCAACCTGGTCACGGCGTACGAGCAG
 ACCGTGAAGGAGGAGAGCAACTTCCAAAAATCCTTCAACAACCACGTGCGCACCTGATCGCGCGCGAGGA
 GGTGACCCTGGGCTGATGCACCTGTGGGACCTGCTGGAGGCCATCGTGCAGAACCCACCAGCAAGCCGC
 TGACGGCGCAGCTGTTCTGGTGGTGCAGCATAGTCGGGACAACGAAGCGTTTCAGGGAGGCGCTGCTGAAT
 ATCACCGAGCCCGAGGGCGCTGGCTCCTGGACCTGGTGAACATTCTGCAGAGCATCGTGGTGCAGGAGCG
 CGGGCTGCCGCTGTCCGAGAAGCTGGCGGCCATCAACTTCTCGGTGCTGAGTTTGGGCAAGTACTACGCTA
 GGAAGATCTACAAGACCCCGTACGTGCCCATAGACAAGGAGGTGAAGATCGACGGTTTTTACATGCGCATG

ACCCTGAAAGTGCTGACCCTGAGCGACGATCTGGGGGTGTACCGCAACGACAGGATGCACCGTGCGGTGAG
 CGCCAGCAGGGCGGCGGAGCTGAGCGACCAGGAGCTGATGCATAGTCTGCAGCGGGCCCTGACCGGGGCGG
 GGACCGAGGGGGGAGAGCTACTTTGACATGGGCGCGGACCTGCACTGGCAGCCCAGCCGCCGGGCGCTTGGAG
 GCGGCGGCAGGACCCTACGTAGAAGAGGTGGACGATGAGGTGGACGAGGAGGGCGAGTACCTGGAAGACTG
 ATGGCGCGACCGTATTTTTGCTAGATGCAACAACAACAGCCACCTCCTGATCCCGCGATGCGGGCGGCGCT
 GCAGAGCCAGCCGTCCGGCATTAACTCCTCGGACGATTGGACCCAGGCCATGCAACGCATCATGGCGCTGA
 CGACCCGCAACCCCGAAGCCTTTAGACAGCAGCCCCAGGCCAACCGGCTCTCGGCCATCCTGGAGGCCGTG
 GTGCCCTCGCGCTCCAACCCACGCACGAGAAGGTCTGGCCATCGTGAACGCGCTGGTGGAGAACAAGGC
 CATCCGCGGCGACGAGGCCGGCCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCCGCTACAACAGCACCA
 ACGTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCCAGCGCGAGCGGTTCCAC
 CGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCGCCAACGTGCCCGG
 GGGCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGCGAGG
 TGTACCAGTCCGGGCGGACTACTTCTCCAGACCACTGCGCCAGGGCTTGACAGCCGTGAACCTGAGCCAG
 GCTTTCAAGAACTTGAGGGCCTGTGGGGCGTGAGGCCCGGTCGGGGACCGCGCGACGGTGTGAGCCT
 GCTGACGCCGAACTCGCGCCTGCTGCTGCTGCTGGTGGCCCCCTTCACGACAGCGGCAGCATCAACCGCA
 ACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCAGACC
 TACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCCTGAA
 CTTTTTGTGCTGACCAACCGGTGCGAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGCATCC
 TGCGTTACGTGCAGCAGAGCGTGGGCCTGTTCTGATGCAGGAGGGGGCCACCCCGAGCGCCGCGCTCGAC
 ATGACCGCGCGCAACATGGAGCCCAGCATGTACGCCAGCAACCGCCCGTTTCATCAATAAACTGATGGACTA
 CTTGCATCGGGCGGCCGCCATGAACTCTGACTATTTACCAACGCCATCCTGAATCCCCACTGGCTCCCGC
 CGCCGGGGTTCTACACGGGCGAGTACGACATGCCCCAGCCCAATGACGGGTTCTGTGGGACGATGTGGAC
 AGCAGCGTGTCTCCCCCGACCGGGTGCTAACGAGCGCCCCCTTGTGGAAGAAGGAAGGCAGCGACCGACG
 CCCGTCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCCGAGGCCGCCAGTCTTTCCCGA
 GCTTGCCCTTCTCGCTGAACAGTATCCGCAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCTGGGC
 GAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGATAGA
 AAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCGTCGC
 AGGGGGCCACGAGCCGGGGCAGCGCCGCCCCGTAAACGCCGGTGGCACGACAGGCAGCGGGGACAGATGTGG
 GACGATGAGGACTCCGCCGACGACAGCAGCGTGTGGACTTGGGTGGGAGTGGTAACCCGTTTCGCTCACCT
 GCGCCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGCGACC
 AGCGTGCCTTCGTTTCTCTGTTGTTGTATCTAGTATGATGAGGCGTGCGTACCCGGAGGGTCTCTC
 CTCCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGCGGCGGCGATGCAGCCCCGCTGGAGGGTCTCT
 TACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTTCGTTACTCGGAGCTGGCACCCCTT
 GTACGATAACCACCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAGAACG
 ACCACAGCAACTTCTGACCACCGTGGTGCAGAACAATGACTTCACCCCCACGAGGCCAGCACCCAGACC
 ATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCAACGT
 GAACGAGTTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGGGTGA
 CAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCCCGAA
 GGCAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGGTGGG
 GCGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTCGACACTAGGAACTTCAGGCTGGGCTGGG
 ACCCCGTGACCGAGCTGGTTCATGCCCCGGGTGTACACCAACGAGGCTTTCATCCCGATATTGTCTTGCTG
 CCCGGCTGCGGGGTGGACTTCACCGAGAGCCGCCTCAGCAACCTGCTGGGCATTTCGAAGAGGCAGCCCTT
 CCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGGCAACATCCCCGCGCTCCTGGATGTGACG
 CCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGAGGTC
 AGGGGCGATAATTTTGAAGCGCCGCGAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGATAGT
 CATTACGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACCGCCT
 ACCGCAGCTGGTACCTAGCCTACAACCTATGGCGACCCCGAGAAGGGCGTGCGCTCCTGGACGCTGCTCACC
 ACCTCGGACGTACCTGCGGCGTGGAGCAAGTCTACTGGTTCGCTGCCCCGACATGATGCAAGAACCCGGTAC
 CTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGTTGGTGGGCGCCGAGCTCCTGCCCTACTACTCAAGA
 GCTTCTTCAACGAGCAGGCGCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCAAGTCTTCAAC
 CGCTTCCCCGAGAACCAGATCCTCGTCCGCCCCGCGCGCCACCATTACCACCGTCAGTGAAAACGTTCC
 TGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTTACTG
 ACGCCAGACGCCGACCTGCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCGCGCGCTCCTCTCGAGC
 CGCACCTTCTAAATGTCCATTCTCATCTCGCCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCAGCAA
 GATGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCAGCGTGCGCGGGCACTTCCGCGCTCCCT
 GGGGCGCCCTCAAGGGCCGCGTGCGGTGCGCGACCAACCGTGCAGCAGCTGATCGACCAGGTGGTGGCCGAC
 GCGCGCAACTACACCCCGCGCGCGCCGCTCTCCACCGTGGACGCCGTTCATCGACAGCGTGGTGGCCGA

CGCGCGCCGGTACGCCCCGCCAAGAGCCGGCGGGCGCATCGCCCCGGCGGCACCGGAGCACCCCCGCCA
 TGCGCGCGGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGCAGGGCCATGCTCAGGGCGGCCAGA
 CGCGCGGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGGCGGCAGCGGCCATCGC
 CAGCATGTCCCGCCCGCGGGCAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTGCCCG
 TGCGCACCCGCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGCGAGGAGGA
 TGTCCAAGCGCAAATTCAGGAAGAGATGCTCCAGGTCATCGCGCCTGAGATCTACGGCCCTGCGGTGGTG
 AAGGAGGAAAGAAAGCCCCGCAAAATCAAGCGGGTCAAAAAGGACAAAAAGGAAGAAGAAAGTGATGTGGA
 CGGATTGGTGGAGTTTGTGCGGAGTTGCCCCCGGGCGGGCGGTGCAGTGGCGGGGGCGGAAGGTGCAAC
 CGGTGCTGAGACCCGGCACCCACCGTGGTCTTCACGCCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCGCTCC
 TACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGGCCGAGCGCCTGGGCGAGTTTGTCTACGG
 CAAGCGCAGCCGTTCCGCAACGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCACGCCGA
 GCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCAGCCGCGGGCGCCGCGCCGGGGGTTCAGCGCGAGGGC
 GAGGATCTGTACCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGACCAT
 GAAGGTGGACCCGGACGTGCAGCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCGCTGGGCG
 TGCAGACCGTGGACATCAAGATTCCACGAGGCCATGGAACGCAGACCGAGCCCATGATCAAGCCAGC
 ACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCAAGTA
 CGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTACCGCG
 GCACGCGCTTCTACCGCGGTTCATACCAGCAGCCGCCGCCGAAGACCACCACTCGCCGCCGCCGTGCGCGC
 ACCGCCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGACCCT
 GCCGCGCGCGCTACACCCGAGCATCGCCATTTAACTTTGCGCTGCTTTGCAGATCAATGGCCCTCAC
 ATGCCGCTTTCGCGTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGCTGGCGGGGAACG
 GGATGCGTGCACCACCACCGGCGGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCTGCCCCGCG
 CTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCTCTCA
 GCGCCACTGAGACACACTTGGAACATCTTGTAAATAAACCaATGGACTCTGACGCTCCTGGTCCTGTGATG
 TGTTTTTCGTAGACAGATGGAAGACATCAATTTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGTTTCAT
 GGGCACCTGGAGCGACATCGGCACCAGCCAATGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGGAGCG
 GGCTTAAGAATTTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACACAGGGCAGGCG
 CTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTCGATGGGCTCGCCTCGGGCATCAACGG
 GGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCCTGGACCCGGTGCCGCCGCCGCG
 GCTCCGTGGAGATGCCGCAAGTGGAGGAGAGCTGCCCTCCCCTGGACAAGCGGGGCGAGAAGCGACCCCGC
 CCTGTGCGGAGGAGACGTGCTGACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAACTGGG
 TCTGCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCGCGACCCCTGG
 ACTTGCCCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCTGCCGCCGGTGGCCGTGGCCCCG
 GCGCGACCCGGGGGACCCGCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTCTGGG
 AGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGTGTGT
 ATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCGCCGAGTTGCAAGAT
 GGCCACCCCATCGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTACCTGA
 GTCCGGGTCTGGTGCAGTTTGGCCGCGCCACAGACACCTACTTCAGTCTGGGGAACAAGTTTAGGAACCCC
 ACGGTGGCGCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCGTGGA
 CCGCGAGGACAACACCTACTCGTACAAAGTGCCTACACGCTGGCCGTGGGCGACAACCGCGTGCTGGACA
 TGGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCTACTCCGGCACC
 GCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTGAGTGGACATATAAAGCCGATGGTGAAAC
 TGCCACAGAAAAAACCTATACATATGGAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGTATTC
 AACTTGGAAGTACACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGTGGGT
 GATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATACCAA
 AATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTGAGGCAAATGTGAAAACAG
 GAACAGGCACTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGCTGGC
 CTAGCTCCAGAAATGTTTTGTATACTGAAAATGTGGATTTGGAACTCCAGATACCCATATTGTATACAA
 AGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTGAGCAAGCCATGCCCAACAGACCTACTACA
 TTGGTTTTAGAGACAACCTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGCCGGT
 CAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCTACCAGCTCTTGCT
 TGACTCTCTGGGTGACAGAACCCGGTATTTAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCTGATG
 TGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGTTGGC
 AGAACAGATACTTATCAGGGAATTAAGGCTAATGGAAGTATGAAACACATGGACCAAAGATGACAGTGT
 CAATGATGCTAATGAGATAGGCAAGGGTAATCCATTGCCATGGAAATCAACATCCAAGCCAACCTGTGGA
 GGAAGTTCCTCTACGCCAACGTGGCCCTGTACCTGCCCGACTCTTACAAGTACACGCCGGCCAATGTTACC
 CTGCCCCACCAACCAACACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACTCCTA

CATCAACATCGGGGCGCGCTGGTCGCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGCAATG
 CGGGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCCCCAG
 AAATTTTTTCGCCATCAAGAGCCTCCTGCTCCTGCCCCGGGTCTTACACCTACGAGTGGAACCTCCGCAAGGA
 CGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCTCCATCTCCTTACCA
 GCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAAACACGGCCTCCACGCTCGAGGCCATGCTGCGC
 AACGACACCAACGACCAGTCCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGGCCAA
 CGCCACCAACGTGCCCATCTCCATCCCCCTCGCGCAACTGGGCGCGCTTCCGCGGCTGGTCCTTACGCGTC
 TCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTCGACCCCTACTTCGTCTACTCGGGCTCCATCCCC
 TACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCGTGAG
 CTGGCCCCGGCAACGACCGGCTCCTGACGCCAACGAGTTTCAAGCAAGCGCACCGTCGACGGCGAGGGCT
 ACAACGTGGCCAGTGCAACATGACCAAGGACTGGTTCTTGGTCCAGATGCTGGCCCACTACAACATCGGC
 TACCAGGGCTTCTACGTGCCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGCCCAT
 GAGCCGCCAGGTGGTGACGAGGTCAACTACAAGGACTACCAGGCCGTACCCCTGGCCTACCAGCACACA
 ACTCGGGCTTTCGTGCGCTACCTCGCGCCCACCATGCGCCAGGGCCAGCCCTACCCCGCAACTACCCCTAC
 CCGCTCATCGGCAAGAGCGCCGTACCAGCGTCACCCAGAAAAAGTTCTCTGCGACAGGGTCATGTGGCG
 CATCCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTATGCCA
 ACTCGCCCCACGCGCTAGACATGAATTTTGAAGTGCACCCCATGGATGAGTCCACCCTTCTCTATGTTGTC
 TTGAAGTCTTCGACGTGTCGAGTGACACCAGCCCCACCGCGCGCTCATCGAGGCCGTCTACCTGCGCAC
 CCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGCAAGCCATGGCCGCGGGCTCCGGCGA
 GCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCTTGGGCACCTTCGATAAGCGCT
 TCCCGGGATTTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGGGGGC
 GAGCACTGGCTGGCCTTCGCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGTTCTC
 GGACGAGCGCCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGAGCGCCCTGGCCACCG
 AGGACCGCTGCGTCACCCCTGGAAAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCGCTGCGGGCTC
 TTCTGCTGCATGTTTCTGACGCCTTTCGTGCACTGGCCCCGACCGCCCCCATGGACAAGAACCCACCATGAA
 CTTGCTGACGGGGGTGCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAACCAGG
 AGGCGCTCTACCGCTTCTCAACTCCCCTCCGCTACTTTTCGCTCCCACCGCGCGCGCATCGAGAAGGCC
 ACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAAACAGCAC
 TTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTCTGCGGGGTCTCGGCATGGC
 CCGCGGGCAGGGACACGTTGCGGAACCTGGTACTTGGCCAGCCACTTGAACCTCGGGGATCAGCAGTTTGGGC
 AGCGGGGTGTCGGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCGAGGGCGCCAGCAGGTGCGGCGC
 GGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTCAGCACT
 GGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTTCGCGTTCGGTGATGCTCTCCACGTGAGG
 TCCTCGGCGTTGGCCATCCCGAAGGGGGTTCATCTTGAGGTCTGCCTTCCCATGGTGGGCACGCACCCGGG
 CTTGTGGTTGCAATCGCAGTGACGGGGATCAGCATCATCTGGGCCTGGTTCGGCGTTTCATCCCCGGGTACA
 TGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGACCCG
 CAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTTCGTGCACGCAGCAGCGCGCGTTCGTTGTTGGC
 CAGCTGCACCACGCTGCGCCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTTCGGGGTTCTCCTTCAGCGCGC
 GCTGCCCCGTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTGCAGG
 CACCGCAGCTTGCCTTCGGCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCCTCCAGTTCTT
 GTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTTCAGGGTCTTGT
 TGCTAGTGAAGGTGAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTACACC
 TCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTTCAGGTTCGGTCTCCACGCGGTAGCGGTCCATCAGCAT
 AGTCATGATTTCCATACCCCTTCTCCCAGGCCGAGACGATGGGCAGGCTCATAGGGTTCTTACCATCATCT
 TAGCGCTAGCAGCCGCGGCCAGGGGGTTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCTCCTTCG
 GTGATCCGCACCGGGGGGTAGCTGAAGCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCTCCTCGCT
 GTCCTGGCTGACGTCCTGCAGGACCACATGCTTGGTCTTTCGGGGTTTCTTCTTGGGCGGCAGCGGGCGG
 GAGATGTTGGAGATGGCGAGGGGGAGCGGAGTTCTCGCTCACCCTACTATCTCTTCTCTCTTCTTGGTCC
 GAGGCCACGCGCGGTAGGTATGTCTCTTTCGGGGGAGAGGCGGAGGCGAGGGCTCTCGCCGCGGCTACT
 TGGCGGATGGCTGGCAGAGCCCCCTTCCGCGTTTCGGGGGTGCGCTCCCGCGCGCGCTCTGACTGACTTCTC
 CGCGGCCGCGCCATTGTGTTCTCCTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTCGCCA
 TCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCGCCAGCCCCGC
 CACCTCCGACGCGGCGTCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTATGTGA
 CGCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCGCTTTTACAAGAAGAGATACACCAAGAAGAGCCAGAG
 CAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCGGGG
 GGAGGACGCGCTCATCAAGCATCTGGCCCCGAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGCACCG
 AGGTGCCCCCTCAGCGTGGAGGAGCTCAGCCGCGCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCCCCC

AAGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTCGCGGTGCCCGA
 GGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGCACCC
 GCGCCGACGCCCTTTTCAACCTGGGTCCCCGGCGCCCGCTACCTGATATCGCCTCCTTGGAAGAGGTTCCC
 AAGATCTTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCGAACGCTCTGCAAGGAGAAGGAGGAGAGCA
 TGAGCACCACAGCGCCCTGGTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACGGTCG
 AGCTGACCCATTTTCGCTACCCGGCTCTGAACCTGCCCCCAAGTCATGAGCGCGGTTCATGGACCAGGTG
 CTCATCAAGCGCGCTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCCGTGGT
 CAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCTAATGCTAGTCCCCAGAGTTTGGAAGAGCGGCGCAAAC
 TCATGATGGCCGTGGTCCTGGTGACCGTGGAGCTGGAGTGCCTGCGCCGCTTCTTCGCGGACGCGGAGACC
 CTGCGCAAGGTTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGATCTC
 CAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGACGAGAACC GCCTGGGGCAGAACGTGC
 TGCACACCACCTTGC GCGGGGAGGCCCCGGCGCAGTACATCCGCGACTGCGTCTACCTCTACCTCTGCCAC
 ACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCAAGCT
 CCTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTCGACGAGCGCACCACCGCTCGGACCTGGCCGACC
 TCATTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCGACTTTATGAGCCAAAGCATGTTG
 CAAAACCTTTTCGCTCTTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCTCGGA
 CTTTCGTGCGCTGACCTTCCGCGAGTGCCCCCGCGCTGTGGAGCCACTGCTACCTGCTGCGCCTGGCCA
 ACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTGAGCGGCGAGGGCCTGCTCGAGTGCCACTGCCGC
 TGCAACCTCTGCACGCCGACCGCTCCCTGGCCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCATCGG
 CACCTTCGAGTTGCAAGGGCCAGCGAAGGCGAGGGTTTCAGCCGCCAAGGGGGGTCTGAAACTCACCCCGG
 GGCTGTGGACCTCGGCCTACTTGCGCAAGTTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTTCTAC
 GAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTGCGCCTGCGTCATCACCCAGGGGGCGATCCTGGCCCA
 ATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGACCCCC
 AGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCGGAGGAAACAAGAAGCTGAAAGTGAGCT
 GCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAGGAAG
 ACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGCAGAG
 GAGGAGGTGGAAGAAGCAGCCGCCCGCCAGACCGTTCGTCTCGGCGGGGGAGAAAGCAAGCAGCACGGATAC
 CATCTCCGCTCCGGGTTCGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTTCCCGAACC
 CCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCTTCGCGGGGGCACAAAAACGCCATCGTC
 TCCTGCTTGCAAGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGGTGAA
 CTTTCCCCGCAACATCTTGCAATTACTACCTGCACCTCCACAGCCCTACTACTTCCAAGAAGAGGCAGCAG
 CAGCAGAAAAAGACCAGCAGAAAAACCAGCAGCTAGAAAAATCCACAGCGGCGGCAGCAGGTGGACTGAGGAT
 CGCGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCACCCCTCTATGCCATCTTCC
 AGCAGAGTCGGGGGAGGAGCAGGAACCTGAAAGTCAAGAACCCTTCTCTGCGCTCGCTCACCCGCAGTTGT
 CTGTATCACAAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGTACTG
 CGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCAGTCGCAGAAAAAGGCGGGAATTACGTACCTGTGC
 CCTTCGCCCTAGCCGCTCCACCCATCATCATGAGCAAAGAGATTCCACGCCTTACATGTGGAGCTACCA
 GCCCCAGATGGGCCTGGCCGCCGGTGCCGCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCCGGGC
 CCGCGATGATCTCACGGGTGAATGACATCCGCGCCACC GAAACCAGATACTCCTAGAACAGTCAGCGCTC
 ACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCTGGTGTACCAGGAAATTCCCCA
 GCCCACGACCGTACTACTTCCGCGAGACGCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAGCTGG
 CGGGCGGCGCCACCCTGTGTGCTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGTACCAGGAGGAGGC
 ACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACCTCGCCGG
 ATCGGGGAGATCTTCTTCACGCCTCGTCAGGCCGTCTGACTTTGGAGAGTTTCGTCTCGCAGCCCCGCT
 CGGGTGGCATCGGCACTCTCCAGTTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTCCGGC
 TCCCCCGGCCACTACCCGGACGAGTTCATCCCGAATTCGACGCCATCAGCGAGTCGGTGGACGGCTACGA
 TTGAAACTAATCACCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAAAAAT
 AATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTTTGAATTTAACAATAAATAAAGAACTACT
 TACTTTAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCCTTCACTCCCCCTTCTCCGACT
 CTGGTACTGCAGCCCCGGCGGGCTGCAAACCTTCCTCCACGCTGAAGGGGATGTCAAATTCCTCCTGTC
 CCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGACCCCG
 TCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCTTCGTCTCTTCAGATGGA
 TTCCAAGAGAAGCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCAACCACCAAGAACGGGGAAT
 CACCCTCAAGCTGGGAGAGGGGTGGACCTCGATTCTCGGGAAAACCTCATCTCCAACACGGCCACCAAGG
 CCGCCGCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAAAGAT
 GGAAAATTATCCTTACAAGTTTCTCCACCATTAAATATACTGAGAACAAGCATTCTAAACACACTAGCTTT
 AGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACATTTG

ATACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGAAAGC
 AACATAAGCTGGGCTAAAGGTTTAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGTTAGA
 GTTTGGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCTGGCC
 TTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGACAACA
 CCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAACTAACACTTTGCTTGACTAAATG
 TGGTAGTCAAATACTGGCCACTGTGTCACTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCCATTACTGGCA
 CCGTAAGCAGTGCTCAGGTGTTTTCTACGTTTTTGATGCAAACGGTGTTCTTTTAAACAGAACATTCTACACTA
 AAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGATTTCAT
 GCCCAATTTAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTATACA
 TGAATGGAGATGTTTTCAAAACCTATGCTTCTCACTATAACCCCTCAATGGTACTGATGACAGCAACAGTACA
 TATTCAATGTCAATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACTCTTA
 TACCTTCTCATACTCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCCCTTCCCACCCCACTC
 TGTGGAACAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTACAGGA
 TTCGAGCAGTTATTTTTCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCGCACAGCCTTG
 AACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAGCCAG
 TCTCGGGTCGGTCAGGGAGATGAAACCCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGCTGAG
 GATTGTCTCTCGGTGGTTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGCGGTGGGAATCATAGTCCG
 CGAACGGGATCGGCCGGTGGTGTGCGCATCAGGCCCCGAGCAGTCGCTGCCGCCGCCGCTCCGTCAAGCTG
 CTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCACGGCCCTCAGCATCAGTCGTCTGGT
 GCGGCGGGCGCAGCAGCGCATGCGGATCTCGCTCAGGTGCTGCGTACGTGCAACACAGAACCACAGGT
 TGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGAAGGATGCTACCCACGTGGCCG
 TCGTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCACGTACATGATCTCCTT
 GGGCATGTGGCGGTTACCACCTCCCGGTACCACATCACCTCTGGTTGAACATGCAGCCCCGGATGATCC
 TGCGGAACCACAGGGCCAGCACCCGCCCGCCCGCCATGCAGCGAAGAGACCCCGGGTCCCGGCAATGGCAA
 TGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACAGGCA
 TATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGTCAAAACCATATCCCAGGGCACGGGAACT
 CTTGCAGGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAGGGTA
 TCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTGGTAA
 GGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTTCGCGACCGTGTCTATGATGCAGTTGC
 TTTTCGACATTTTCTGACTTGTGTAGCAGAACTGGTCCGGGCGCTGCACACCGATCGCCGGCGGGCGGT
 CCGGCGCTTGGAACGCTCGGTGTTGAAATTGTAACACGCCACTCTCTCAGACCGTGCAGCAGATGTAGGG
 CCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGTGGGAATGGGCCAGA
 CCCAGCCAGATGATGCAATTTTGTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCATGAT
 TAACTTTTAATCCAAACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCCCCGC
 TGTGTTGGTGGAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTCCAGC
 AAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCTCAATCAT
 CATGTTACACTCCTGCACCATCCCCAGATAATTTTCAATTTTCCAGCCTTGAATGATTGCAACTAGTTCcT
 GAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGGCATTCTTAAGCACACC
 CTCATAATTCCAAGATATTCTGCTCCTGGTTACCTGCAGCAGATTGACAAGCGGAATATCAAAATCTCTG
 CCGCATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTCATATCCTCTCCGAAATTTTTAGC
 CATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAAACCGAAGTCTCCCCAGTGAGCAT
 TGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAGAAAA
 TCGCCCAGGCAATTTTTAAGAAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTLAGAGCCTCGGGAAC
 AACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAAAATG
 AACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACAGGAGAGCGTTCCCGGTGGCC
 TCTCCGGCGCGACCCCTCGTAAAAATTGTGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGTGGCC
 GCGGTGAATGATTCGACAAGATGAATACACCCCGGAACATTGGCGTCCGCGAGTGAAAAAAGCGCCCCGA
 GGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCACAAAA
 TTCTCAGGTGCGTACAAAATGTAATTACTCCCCCTCTGCACAGGCAGCAAAGCCCCCGATCCCCCAGGTA
 CACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGAAAGG
 CTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGCCAAA
 GTCTAAAAATACCCGCCAAATAATCACACACGCCAGCACACGCCAGAAACCGGTGACACACTCAAAAAA
 ATACGCGCACTTCTCCTCAAACGCCCAAACTGCCGTCAATTTCCGGGTTCCACGCTACGTATCAAAACACG
 ACTTTCAAATTCGTCGACCGTTAAAAACGTACCCGCCCGCCCTAACGGTCGCCCCGTCTCTCAGCCAA
 TCAGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGGTATAT
 TATTGATGATGG

5

10

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

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.

Se digirieron 10 µg de ADN plásmido con PaeI 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 µl de agua previamente calentada; las columnas se dejaron a 37 grados durante 0.5-1 horas antes de la etapa de elución.

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 µg de ADN purificado se utilizaron por pocillo en una transfección con dos veces el volumen en µl (2-4 µl) 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.

Los cultivos celulares transfectados se incubaron a 37 °C 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 °C 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

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.

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 ADN/50 µL.

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 centrifuga 50 mL seguido de tres rondas de congelamiento descongelamiento (a -80 °C y 37 °C). El lisado posterior, denominado solución madre de virus primaria, se clarificó mediante centrifugación a velocidad total en una centrifuga 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 $1,5 \times 10^7$ 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

La producción del virus ChAdV68 se llevó a cabo en células 293F cultivadas en medio 293 FreeStyleT[™] (ThermoFisher) en una incubadora a CO₂ 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, centrifuga 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

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.

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 °C. 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-Lyzer™ (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 °C.

Ensayos virales

La concentración de VP se llevó a cabo mediante el uso de un ensayo de OD 260 en función del coeficiente de extinción de $1,1 \times 10^{12}$ 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 $1,1 \times 10^{12}$ VP.

Se calculó una titulación de la unidad infecciosa (IU, por su sigla 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 1×10^{-7} . 100 µL de estas diluciones se agregaron a células 293A que se cultivaron al menos una hora antes de a 3×10^5 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 °C. Las células se lavaron luego con 1XPBS y luego se fijaron con metanol frío al 100 % (-20 °C). Las placas se incubaron luego a -20 °C 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 las 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 H₂O₂ 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 mm² 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

Se inyectaron ratones hembra C57BL/6J y ratones hembra Balb/c con 1x10⁸ partículas virales (VP) de ChAdV68.5WTnt.MAG25mer en una inyección intramuscular bilateral de 100 µL de volumen (50 µL por pata).

Disociación de esplenocitos

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 (NH₄Cl 150 mM, KHCO₃ 10 mM, Na₂EDTA 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 inmunospot 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 IFN γ de ratón (MABTECH). Se incubaron 5×10^4 esplenocitos con 10 μ M de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFN γ 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.

XV.B.2. Producción de partículas de administración viral ChAdV68 luego de la transfección de ADN

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

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

5 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 epítipo de MHC clase I SIINFEKL (SEQ ID NO: 57) se midieron en ratones hembra C57BL/6J y el epítipo de MHC clase I AH1-A5 (Slansky et al., 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 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 10⁶ esplenocitos en ensayos ELISpot cuando se inmunizaron ratones C57BL/6J o Balb/c con ChAdV68.5WTnt.MAG25mer, respectivamente, 10 días luego de la inmunización.

15 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

20 **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 m⁷G (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.

Para estudios in vivo: Se generó el ARN y se purificó mediante TriLink Biotechnologies y se tapó con Cap1 enzimática.

Transfección de ARN

Las células HEK293A se cultivaron a 6×10^4 células/pocillo para 96 pocillos y 2×10^5 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 μ L de lipofectamina y 10 ng de ARNm por pocillo, y para 24 pocillos, se utilizaron 0,75 μ L 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

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

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 qTower³ (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	138
	Cebador2	CGCTCGTTGTAGATGTCGTTAG	139
	Sonda	/56-FAM/TTGCAGTTC/ZEN/TTCATGCCCGTGTTG/3IABkFQ/	140
GusB	Cebador1	GTTTTTGATCCAGACCCAGATG	141
	Cebador2	GCCCATTATTCAGAGCGAGTA	142
	Sonda	/56-FAM/TGCAGGGTT/ZEN/TCACCAGGATCCAC/3IABkFQ/	143

ActB	Cebador1	CCTTGCACATGCCGGAG	144
	Cebador2	ACAGAGCCTCGCCTTTG	145
	Sonda	/56-FAM/TCATCCATG/ZEN/GTGAGCTGGCGG/3IABkFQ/	146
MAG-25mer	Cebador1	CTGAAAGCTCGGTTTGCTAATG	147
Conjunto1	Cebador2	CCATGCTGGAAGAGACAATCT	148
	Sonda	/56-FAM/CGTTTCTGA/ZEN/TGGCGCTGACCGATA/3IABkFQ/	149
MAG-25mer	Cebador1	TATGCCTATCCTGTCTCCTCTG	150
Conjunto2	Cebador2	GCTAATGCAGCTAAGTCCTCTC	151
	Sonda	/56-FAM/TGTTTACCC/ZEN/TGACCGTGCCTTCTG/3IABkFQ/	152

5

Modelo tumoral B16-OVA

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.

10

Modelo tumoral CT26

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.

15

Inmunizaciones

Para la vacuna de srRNA, los ratones recibieron la inyección de 10 µg de ARN en una inyección intramuscular bilateral de 100 µL de volumen (50 µL por pata). Para la vacuna de Ad5, los ratones recibieron la inyección de 5×10^{10} partículas virales (VP) en una inyección intramuscular bilateral de 100 µL de volumen (50 µL 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 µg, 2 veces por semana, a través de inyección intraperitoneal.

20

Imagenología luminiscente in vivo

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

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 (NH₄Cl 150 mM, KHCO₃ 10 mM, Na₂EDTA 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 inmunospot 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 IFN γ de ratón (MABTECH). Se incubaron 5x10⁴ esplenocitos con 10 μ M de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFN γ 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.

XVI.B. Vector del alfavirus

XVI.B.1. Evaluación in vitro del vector del alfavirus

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 al.* 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

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 epítipo (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 epítipo 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

5 En otro ejemplo, se evaluó la expresión del indicador de VEE-Luciferasa in vivo. Se administraron inyecciones a ratones de 10 µg 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

10 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

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

15 de VEE (VEE-UbAAY, SEQ ID NO:14) que expresa 2 epítomos tumorales de ratones del MHC clase I, SIINFEKL (SEQ ID NO: 57) y AH1-A5 (Slansky et al., 2000, Immunity 13:529-538). El epítipo SFL (SIINFEKL (SEQ ID NO: 57)) es expresado por la línea celular de melanoma B16-OVA, y el epítipo AH1-A5 (SPSYAYHQF(SEQ ID NO: 58); Slansky et al., 2000, Immunity) induce linfocitos T que se dirigen a un epítipo relacionado (AH1/

20 SPSYVYHQF (SEQ ID NO: 153); Huang et al., 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).

Se observó una fuerte respuesta de linfocitos T específicos del antígeno dirigida al SFL 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 MHC I 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.*

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 (% de CD8)	Grupo	Ratón	SFC/1e6 esplenocitos	Resultado positivo para pentámeros (% 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				

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

- 5 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

- 10 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
- 15 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 µL	día 0	IM
		Control con srRNA	10 µg	50 µL	día 14, 28, 42	IM
		Anti-PD1	250 µg	100 µL	2x / semana (día de inicio 0)	IP
2	40	Control con chAd68	1e11 vp	2x 50 µL	día 0	IM
		Control con srRNA	10 µg	50 µL	día 14, 28, 42	IM
		Anti-IgG	250 µg	100 µL	2x / semana (día de inicio 0)	IP
3	28	Vacuna contra chAd68	1e11 vp	2x 50 µL	día 0	IM
		Vacuna de srRNA	10 µg	50 µL	día 14, 28, 42	IM
		Anti-PD1	250 µg	100 µL	2x / semana (día de inicio 0)	IP
4	28	Vacuna contra chAd68	1e11 vp	2x 50 µL	día 0	IM
		Vacuna de srRNA	10 µg	50 µL	día 14, 28, 42	IM
		Anti-IgG	250 µg	100 µL	2x / semana (día de inicio 0)	IP
5	28	Vacuna de srRNA	10 µg	50 µL	día 0, 28, 42	IM
		Vacuna contra chAd68	1e11 vp	2x 50 µL	día 14	IM
		Anti-PD1	250 µg	100 µL	2x / semana (día de inicio 0)	IP
6	28	Vacuna de srRNA	10 µg	50 µL	día 0, 28, 42	IM
		Vacuna contra chAd68	1e11 vp	2x 50 µL	día 14	IM
		Anti-IgG	250 µg	100 µL	2x / semana (día de inicio 0)	IP
7	40	Vacuna de srRNA	10 µg	50 µL	día 0, 14, 28, 42	IM
		Anti-PD1	250 µg	100 µL	2x / semana (día de inicio 0)	IP
8	40	Vacuna de srRNA	10 µg	50 µL	día 0, 14, 28, 42	IM
		Anti-IgG	250 µg	100 µL	2x / semana (día de inicio 0)	IP

5

Inmunizaciones

Para la vacuna de srRNA, los ratones recibieron la inyección de 10 µg de srRNA de VEE-MAG25mer en una inyección intramuscular bilateral de 100 µL de volumen (50 µL por pata). Para la vacuna de C68, los ratones recibieron la inyección de 1×10^{11} partículas virales (VP) de ChAdV68.5WTnt.MAG25mer en una inyección intramuscular bilateral de 100 µL de volumen (50 µL por pata). Los animales recibieron la inyección de anti-PD-1 (clon RMP1-

10

14, BioXcell) o anti-IgG (clon MPC-11, BioXcell), en dosis de 250 µg, 2 veces por semana, a través de inyección intraperitoneal.

Disociación de esplenocitos

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 (NH₄Cl 150 mM, KHCO₃ 10 mM, Na₂EDTA 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 inmunospot 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 IFNγ de ratón (MABTECH). Se incubaron 5x10⁴ esplenocitos con 10 µM de los péptidos indicados durante 16 horas en placas recubiertas con anticuerpo de IFNγ 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

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 + anti-PD-1	srRNA de VEE-MAG25mer + anti-PD-1

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.

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.

5

Tabla 17 – Respuestas inmunitarias celulares en un modelo tumoral CT26

Tratamiento	Mediana de SFC/ 10^6 esplenocitos
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

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.

20

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

Se midió el desarrollo del tumor en el modelo de tumor de colon 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 ($> 2500 \text{ mm}^3$); 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).

5

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

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/		srRNA/	
			chAdV/ srRNA	+ PD1	srRNA/ chAdV	+ PD1	chAdV + PD1	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

5

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.

10

15

XVIII. Estudio en primates no humanos

Se evaluaron diversos protocolos de dosificación mediante el uso de ChAdV68 y ARN de autorreplicación (srRNA) en primates no humanos: (NHP).

XVIII.A. Materiales y métodos para el estudio de primates no humanos

Inmunizaciones

20

Se inyectó por vía intramuscular una vacuna de sensibilización en cada NHP para iniciar el estudio (vacuna de sensibilización). Se inmunizaron bilateralmente macacos rhesus indios Mamu A01 con 1×10^{12} partículas virales (5×10^{11} partículas virales por inyección) de ChAdV68.5WTnt.MAG25mer, 30 μ g de srRNA de VEE-MAG25MER, 100 μ g

de srRNA de VEE-MAG25mer o 300 µg de srRNA de VEE-MAG25mer formulados en LNP-1 o LNP-2. Se administraron por vía intramuscular 30 µg, 100 µg o 300 µg de refuerzos de vacuna con srRNA de VEE-MAG25mer 4 semanas después de la vacunación de sensibilización. En los grupos de estudio adicionales, se administraron 30 µg, 100 µg o 300 µg de vacunas de srRNA de VEE-MAG25mer como un segundo refuerzo por vía intramuscular 8 semanas después de la vacunación de sensibilización inicial. Se administró anti-CTLA-4 por vía SC cerca del sitio de inmunización con vacuna o por vía IV a grupos especificados. Se administran inyecciones bilaterales por dosis de acuerdo con los grupos indicados en las Tablas 21 y 23.

Supervisión de la inmunidad

Se aislaron PBMC 7, 14, 28 o 35 días 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 ELISpot o citometría de flujo. Se supervisaron las respuestas de linfocitos T de las PBMC a 6 epítomos 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 inmunospot 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 µM 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.

Se supervisan las respuestas de linfocitos T CD4 y CD8 específicas de las PBMC a 6 epítomos 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 inducen con especificidad de antígeno en respuesta a los epítomos.

XVIII.B. Evaluación de la inmunogenicidad en primates no humanos (dosificación de srRNA de alcance corto y medio)

Este estudio se diseñó para (a) evaluar la inmunogenicidad y seguridad preliminar de una inmunización de sensibilización con ChAdV68.5WTnt.MAG25mer seguida por una combinación de sensibilización/refuerzo heterólogos en dosis de 100 µg de srRNA de VEE-MAG25mer; (b) evaluar la cinética de las respuestas de linfocitos T a la combinación de sensibilización/refuerzo con ChAdV68.5WTnt.MAG25mer/srRNA de VEE-MAG25mer. Este grupo de estudio se llevó a cabo en macacos rhesus indios A01 con el fin de 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 A01 se distribuyeron aleatoriamente en los diferentes grupos de estudio (6 macacos por grupo) y recibieron la administración de una inyección IM del vector de srRNA de VEE-MAG25mer o ChAdV68.5WTnt.MAG25mer que codifica modelos de antígenos que incluye múltiples epítomos con restricción de mamu A01. Los grupos de estudio son como se describen en la Tabla 21.

Este estudio también está diseñado para evaluar la inmunogenicidad, la seguridad preliminar y la cinética de la respuesta de linfocitos T de dosis de 30 µg y 100 µg de srRNA de VEE-MAG25mer como una sensibilización/un refuerzo homólogos, así como comparar las respuestas inmunitarias del srRNA de VEE-MAG25mer en nanopartículas lipídicas mediante el uso de LNP1 con respecto a LNP2. Estos grupos de estudio se llevaron a cabo de manera similar a la sensibilización/el cebado con ChAdV68/srRNA que se describieron anteriormente. Los grupos de estudio son como se describen en la Tabla 21.

Tabla 21 - Grupos de estudio de inmunogenicidad en NHP con dosificación de srRNA de alcance corto y medio

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)

Se recolectaron PBMC antes de la inmunización y cada semana después de la inmunización inicial durante las primeras 6 semanas para la supervisión de la inmunidad. Además, se recolectaron PBMC 8 y 10 semanas después de la inmunización inicial, para la supervisión de la inmunidad.

Se midieron las respuestas inmunitarias celulares específicas del antígeno en células mononucleares de sangre periférica (PBMC) a seis epítomos con restricción de mamu A01 antes de la inmunización y 7, 14, 21, 28 o 35 días después de la inmunización de sensibilización inicial con ChAdV68.5WTnt.MAG25mer. Se graficaron las respuestas inmunitarias combinadas a los seis epítomos para cada momento de supervisión de la inmunidad (Figura 34 y Tabla 22). Se observaron respuestas inmunitarias específicas del antígeno en todas las mediciones con 1256, 1823, 1905, 987 SFC por cada 10^6 PBMC (seis epítomos combinados) 7, 14, 21 o 28 días después de la inmunización de sensibilización inicial con ChAdV68.5WTnt.MAG25mer, respectivamente. La respuesta inmunitaria mostró el perfil esperado con respuestas inmunitarias en su punto máximo medidas 7-14 días después de la inmunización de sensibilización, seguidas por una contracción en la respuesta inmunitaria después de 28 días.

También se midieron respuestas inmunitarias celulares específicas del antígeno combinadas de 1851 SFC por cada 10^6 PBMC (seis epítomos combinados) 7 días después del primer refuerzo con srRNA de VEE-MAG25mer (es decir, 35 días después de la inmunización inicial con ChAdV68.5WTnt.MAG25mer). La respuesta inmunitaria medida 7

días después del primer refuerzo con srRNA de VEE-MAG25mer (día 35) fue comparable a la respuesta inmunitaria máxima medida para la inmunización de sensibilización con ChAdV68.5WTnt.MAG25mer (día 14) y ~2 veces más alta que la medida 28 días después de la inmunización de sensibilización con ChAdV68.5WTnt.MAG25mer.

5

Tabla 22 – Respuestas inmunitarias celulares con srRNA de alcance corto y medio

Día	Antígeno					
	Tat TL8	Gag CM9	Env TL9	Env CL9	Gag LW9	Pol SV9
0	6,6±4,4	5,5±5,1	7,9±6,7	3,5±2,8	10,2±7,0	4,8±4,1
7	570,1±178,2	226,7±119,3	214,4±101,0	181,5±67,8	25,5±14,2	38,1±29,9
14	628,0±224,2	350,1±112,7	286,7±102,0	314,7±165,4	56,5±19,0	186,5±80,2
21	556,0±117,2	473,7±106,3	367,5±88,5	280,8±100,9	51,9±13,8	174,6±60,2
28	328,8±48,3	214,4±43,9	167,7±48,6	143,5±46,6	36,7±13,1	95,9±32,4
35	545,0±90,2	548,1±140,8	414,5±92,5	159,1±61,6	45,2±14,5	139,0±52,8

XVIII.C. Evaluación de la inmunogenicidad en primates no humanos (dosificación alta de srRNA y anti-CTLA4)

Este estudio se diseñó para evaluar el impacto de la vía de administración de anti-CTLA4 sobre las respuestas inmunitarias inducidas mediante vacunación (p. ej., comparar el suministro local (SC) de anti-CTLA4 en cercanía estrecha de los ganglios linfáticos de drenaje de vacuna con la administración sistémica (IV)). Este grupo de estudio se llevó a cabo en macacos rhesus indios A01 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 A01 se distribuyeron aleatoriamente en los diferentes grupos de estudio (6 macacos por grupo) y recibieron la administración de una inyección IM de ChAdV68.5WTnt.MAG25mer que codifica modelos de antígenos que incluye múltiples antígenos con restricción de mamu A01. Se administró anti-CTLA-4 por vía SC cerca del sitio de inmunización con vacuna o por vía IV a grupos especificados. Los grupos de estudio se describen en la Tabla 23.

Este estudio también está diseñado 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. Estos grupos de estudio se llevan a cabo en macacos rhesus indios mamu A01 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 A01 se distribuyen aleatoriamente en los diferentes grupos de estudio (6 macacos por grupo) y recibieron la administración de una inyección IM de ChAdV68.5WTnt.MAG25mer o srRNA de VEE-MAG25mer que codifica modelos de antígenos que incluye múltiples epítomos con restricción de mamu A01. Se administra anti-CTLA-4 por vía SC cerca del sitio de inmunización con vacuna o por vía IV a grupos especificados. Los grupos de estudio se describen en la Tabla 23.

Tabla 23 - Grupos de estudio de inmunogenicidad en NHP con dosificación de srRNA de alcance alto

Grupo	Sensibilización	Refuerzo 1	Refuerzo 2
1	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) +anti-CTLA-4 (SC)	VEE-MAG25mer srRNA - LNP2 (300 µg) +anti-CTLA-4 (SC)	VEE-MAG25mer srRNA -LNP2 (300 µg) +anti-CTLA-4 (SC)
3	VEE-MAG25mer srRNA-LNP1 (300 µg)	VEE-MAG25mer srRNA - LNP1 (300 µg)	VEE-MAG25mer srRNA -LNP1 (300 µg)
4	ChAdV68.5WTnt.MAG25mer	VEE-MAG25mer srRNA - LNP2 (300 µg)	VEE-MAG25mer srRNA -LNP2 (300 µg)
5	ChAdV68.5WTnt.MAG25mer +anti-CTLA-4 (SC)	VEE-MAG25mer srRNA - LNP2 (300 µg) +anti-CTLA-4 (SC)	VEE-MAG25mer srRNA -LNP2 (300 µg) +anti-CTLA-4 (SC)
6	ChAdV68.5WTnt.MAG25mer +anti-CTLA-4 (IV)	VEE-MAG25mer srRNA - LNP2 (300 µg) +anti-CTLA-4 (IV)	VEE-MAG25mer srRNA -LNP2 (300 µg) +anti-CTLA-4 (IV)

Los macacos rhesus indios Mamu A01 fueron inmunizados con ChAdV68.5WTnt.MAG25mer con o sin anti-CTLA-4 administrado por vía IV o SC. Se midieron las respuestas inmunitarias celulares específicas de antígenos en células mononucleares de sangre periférica (PBMC) a seis epítomos con restricción de mamu A01 diferentes 14 días después de la inmunización inicial y se graficaron las respuestas inmunitarias combinadas a los seis epítomos (Figura 35 y Tabla 24). Se observaron respuestas inmunitarias específicas del antígeno combinadas de 2257, 5887 o 3984 SFC por cada 10^6 PBMC (seis epítomos combinados) después de una sola inmunización con ChAdV68.5WTnt.MAG25mer, ChAdV68.5WTnt.MAG25mer con anti-CTLA-4 (IV) o ChAdV68.5WTnt.MAG25mer (SC), respectivamente.

Tabla 24 - Respuestas inmunitarias celulares con ChAdV68 y anti-CTLA-4

Grupo	Antígeno					
	Tat TL8	Gag CM9	Env TL9	Env CL9	Gag LW9	Pol SV9
chAdV	608,6±132,6	556,7±136,4	478,7±147,6	297,2±98,3	79,8±29,9	236,4±66,5
chAdV + anti-CTLA4 IV	899,8±287,6	1081±178,7	1234±166,2	1360±139,8	567,6±265,5	744,1±235,6
chAdV + anti-CTLA4 SC	995,5±236,8	1149±158,7	629,8±205,6	595,2±192,6	236,1±71,7	378,8±91,3

Determinadas secuencias

A continuación, se muestran secuencias para vectores, casetes y anticuerpos.

Tremelimumab VL (SEQ ID NO:16)
PSSLSASVGDRVTITCRASQSINSYLDWYQQKPGKAPKLLIYAASSLQSGVPSRFSGSGSGTDFT LTISSLQPEDFATYYCQQYYSTPFTFGPGTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPRE AKV
Tremelimumab VH (SEQ ID NO:17)
GVVQPGRSLRLSCAASGFTFSSYGMHWVRQAPGKGLEWVAVIWDGSGNKYYADSVKGRFTISRDN SKNTLYLQMNSLRAEDTAVYYCARDPRGATLYYYYYGMDVWGQGTITVTVSSASTKGPSVFPLAPCSRSTSE STAALGCLVKDYFPEPVTVSWNSGALTSGVH
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)
EIVLTQSPGTLISLSPGERATLSCRASQRVSSSYLAWYQQKPGQAPRLLIYDASSRATGIPDRFSGSGSGTD FTLTISRLEPEDFAVYYCQQYGSLPWTFGQGTKVEIK
MEDI4736 VH (SEQ ID NO:25)
EVQLVESGGGLVQPGGSLRLSCAASGFTFSRYWMSWVRQAPGKGLEWVANIKQDGSEKYYVDSVKGRFTIS RDNAKNSLYLQMNSLRAEDTAVYYCAREGGWFGELAFDYWGQGTITVTVSS
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)
GCCCGGGCATTAAATGCGATCGCATCGATtacgactctagaatagtctagtcgcgagggccacca tgCAGATCTTTCGTGAAGACCCTGACCGCAAGACCATCACCTAGAGGTGGAGCCCAGTGACACCATCGAG AACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAA GCAGCTGGAAGATGGCCGACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTC GCCTGAGAGGTGcCatgttttcaggcgctgagcgaaggctgcaccccgatgatattaaccagatgctgaac gtgctggggcgatcatcagggtctcaggccttgagcagcttgagagtataatcaactttgaaaaactgactga atggaccagttctaatgttatgCCTATCCTGTCTCCTCTGACAAAGGGCATCCTGGGCTTCGTGTTTACCC TGACCGTGCCTTCTGAGAGAGGACTTagctgcattagcgaagcggatgcgaccaccccgaaagcgcggaac ctggggaagaaattctgagccagctgtatctttggccaagggtagcctaccattcccctagttatgctta ccaccaatttgaaagacgagccaaatataaaagaCACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCT ACCCTGTGTACGTGTTTCGGCGATTGCGTGCAGGGCGATtgggatgcgattcgctttcgctattgcgcgcgcg

ccgggctatgcgctgctgcgctgcaacgataccaactatagcgctctgctggctgtgggggcccctagaagg
 acccaggaatcaggactggcttggtgtcccaagacaacttgtaactCGGATGCAGGCTATTTCAGAATGCCG
 GCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTTCTGGCTGCAAgcgtttctgatggcgctgacc
 gatagcggccccgaaaaccaacattattgtggatagccagtatgtgatgggcattagcaaaccgagctttca
 ggaatttgtggattgggaaaacgtgagcccggaactgaacagcaccgatcagccgtttTGGCAAGCCGGAA
 TCCTGGCCAGAAATCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTACCAGggtcagtca
 ctagtcatctctgcttctatcattgtcttcaacctgCtggaaactggaaggtgattatcgagatgatggcaa
 cgtgtgggtgcataccccgctgagcccgcgccacctgaacgcgtgggtgaaagcgggtggaagaaaaaaaag
 gtattccagttcacctagagctggccagtatgaccaacaTggagctcatgagcagttattgtgcatcagcag
 gtcAGAACATACGGCCCCGTGTTTCATGTGTCTCGGCGGACTGCTTACAATGGTGGCTGGTGTGTGGCT
 GACAGTGCgagtgctcgagctgttccgggcccgcgagctggccaacgacgtggtcctccagatcatggagc
 tttgtggtgagcgttttcgccaggtgtgccataccacctgcccgtggccgaacgcgagcctgacccccgaaa
 tgaacaacgaaaccacccagccccagatcgccaactgcagcgtgtatgacttttttgtgtggtcctcatta
 ttattctgttcgagacacactttggccaaggtgacctaccatatagaacaatatgcgtatcatatgctgg
 aaagacgagccaaatataaaagaGGACCAGGACCTGGCGCTAAATTTGTGGCCGCTGGACACTGAAAGCC
 GCTGCTGGTCTCGACCTGGCCAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTCGGACC
 CGGACCAGGCTGATGATTTTCGAAATTTAAATAAGCTTGCGGCCGCTAGGGATAACAGGGTAATtatacgc
 ccaaacatttacagccgcggtgtcaaaaaccgctgg

Polipéptido de UbA76-25merPDDT (SEQ ID NO:33)

MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQORLIFAGKQLEDGRTLSDYNIQKES
 TLHLVLRRLRGAMFQALSEGCTPYDINQMLNLVLDGHQVSGLEQLESIIINFEKLTWETSSNVMPILSPLTKGI
 LGFVFTLTVPSEGLSCISEADATTPESANLGEIILSQLYLWPRVITYHSPSYAYHQFERRAKYKRHPFGFG
 QSLLFGYPVYVFGDCVQGDWDAIRFRYCAPPGYALLRCNDTNYALLAVGALEGPRNQDWLGVPRLVTRM
 QAIQNAGLCTLVAMLEETIFWLQAFMLALTDSPKNTNIIVDSQYVMGISKPSFQEFVDWENVSPELNSTDQ
 PFWQAGILARNLVPVATVQGNLKYQGQSLVISASIIIVFNLLLELEGDYRDDGNVWVHTPLSPRTLNAWVK
 AVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLGGLLTMVAGAVWLTVRVLELFRAAQLANDV
 VLQIMELCGAAFRQVCHTTVPWPNASLTPKWNNETTQPQIANCSVYDFFVWLHYYSVRDTLWPRVITYHMK
 YAYHMLERRAKYKRGPGPGAKFVAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG

Nucleótido de MAG-25merPDDT (SEQ ID NO:34)

ATGGCCGGGATGTTCCAGGCACTGTCCGAAGGCTGCACACCCTATGATATTAACCAGATGCTGAATGTCCT
 GGGAGACCACCAGGTCTCTGGCCTGGAGCAGCTGGAGAGCATCATCAACTTCGAGAAGCTGACCGAGTGGA
 CAAGCTCCAATGTGATGCCTATCCTGTCCCCACTGACCAAGGGCATCCTGGGCTTCGTGTTTACCCTGACA
 GTGCCCTTCTGAGCGGGCCCTGTCTTGCATCAGCGAGGAGCAGCAACCACACAGAGATCCGCGCAATCTGGG
 CGAGGAGATCCTGTCTCAGCTGTACCTGTGGCCCCGGGTGACATATCACTCCCCCTTCTTACGCCTATCACC
 AGTTCGAGCGGAGAGCCAAGTACAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGTCTGTTTCGGCTACCCC
 GTGTACGTGTTTCGGCGATTGCGTGCAGGGCGACTGGGATGCCATCCGGTTTTAGATACTGCGCACCACTGG
 ATATGCACTGCTGAGGTGTAACGACACCAATTATTCCGCCCTGCTGGCAGTGGGCGCCCTGGAGGGCCCTC
 GCAATCAGGATTGGCTGGGCGTGCCAAGGCAGCTGGTGACACGCATGCAGGCCATCCAGAACGCAGGCCTG
 TGCACCCTGGTGGCAATGCTGGAGGAGACAATCTTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAG
 CGGCCCCAAGACAAACATCATCGTGGATTCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGT
 TTGTGGACTGGGAGAACGTGAGCCCAGAGCTGAATTCCACCGATCAGCCATTCTGGCAGGCAGGAATCCTG
 GCAAGGAACCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCCAGAGCCTGGT
 CATCAGCGCCTCCATCATCGTGTTTAACCTGCTGGAGCTGGAGGGCGACTATCGGGACGATGGCAACGTGT
 GGGTGCACACCCCACTGAGCCCCAGAACACTGAACGCCTGGGTGAAGGCCGTGGAGGAGAAGAAGGGCATC
 CCAGTGCACCTGGAGCTGGCCTCCATGACCAATATGGAGCTGATGTCTAGCATCGTGCACCAGCAGGTGAG
 GACATACGGACCCGTGTTTCATGTGCCTGGGAGGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAG
 TGCGGGTGTGAGGCTGTTTCAGAGCCGCCAGCTGGCCAACGATGTGGTGTGTCAGATCATGGAGCTGTGC
 GGAGCAGCCTTTTCGCCAGGTGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGTGGAA
 CAATGAGACAACACAGCCTCAGATCGCCAACCTGTAGCGTGTACGACTTCTTCGTGTGGCTGCACTACTATA
 GCGTGAGGGATACCTGTGGCCCCGCGTGACATACCACATGAATAAGTACGCCTATCACATGCTGGAGAGG
 CGCGCCAAGTATAAGAGAGGCCCTGGCCCCAGGCGCAAAGTTTGTGGCAGCATGGACCCTGAAGGCCCGCGC
 CGCCCCCGCCCCGGCCAGTATATCAAGGCTAACAGTAAGTTCATTGGAATCACAGAGCTGGGACCCGGAC
 CTGGA

Polipéptido de MAG-25merPDDT (SEQ ID NO:35)
MAGMFQALSEGCTPYDINQMLNVLGDHGVSGLEQLESIIINFEKLTETWTSNVMPILSPLTKGILG FVFTLTVPSEGLSCISEADATTPESANLGEIILSQLYLWPRVITYHSPSYAYHQFERRAKYKRHPGFGQS LLFGYPVYVFGDCVQGDWDIAIRFRYCAPPGYALLRCNDTNYALLAVGALEGPQNQDWLGVPRLVTRMQA IQNAGLCTLVAMLEETIFWLQAFMLALTDSPGPKTNIIVDSQYVMGISKPSFQEFVDWENVSPELNSTDQPF WQAGILARNLVPMVATVQGNLKYQGQSLVISASIIIVFNLLLELEGDYRDDGNVWVHTPLSPRTLNAWVKAV EEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLGGLLTMVAGAVWLTVRVLELFRAAQLANDVVL QIMELCGAAFRQVCHTTVPWPNASLTPKWNNETTQPQIANCSVYDFFVWLHYYSVRDTLWPRVITYHMNKYA YHMLERRAKYKRGPGPGAKFVAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG
Nucleótido Ub7625merPDDT NoSFL (SEQ ID NO:36)
CCCCGGGCATTTAAATGCGATCGCATCGATtaccgactctagaatagtctagtccgcagggccacca tgCAGATCTTTCGTGAAGACCCTGACCGGCAAGACCATCACCTAGAGGTGGAGCCCAGTGACACCATCGAG AACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAA GCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTC GCCTGAGAGGTGcCatgttttcaggcgctgagcgaaggctgcaccccgatgatattaaccagatgctgaac gtgctggggcgatcatcagtttaagcacatcaaagcctttgaccggacatttgctaacaaccagggtcccat ggttgtgtttggccacacctgggCCTATCCTGTCTCCTCTGACAAAGGGCATCCTGGGCTTCGTGTTTACCC TGACCGTGCCTTCTGAGAGAGGACTTagctgcattagcgaagcggatgcgaccaccccgaaagcgcggaac ctgggcgaagaaattctgagccagctgtatctttggccaagggtgacctaccattcccctagtattgtctta ccaccaatttgaaagacgagccaaatataaaaagaCACTTCCCCGGCTTTGGCCAGAGCCTGCTGTTTGGCT ACCCTGTGTACGTGTTTCGGCGATTGCGTGCAGGGCGATtgggatgcgattcgcttttcgctattgctgcgcgcg ccgggctatgctgctgctgctgcaacgataccaactatagcgctctgctggctgtgggggcccctagaagg accaggaatcaggactggccttggtgtcccaagacaacttgtaactCGGATGCAGGCTATTTCAGAATGCCG GCCTGTGTACCCTGGTGGCCATGCTGGAAGAGACAATCTTCTGGCTGCAAgcgtttctgatggcgctgacc gatagcggccccgaaaaccaacattattgtggatagccagtatgtgatgggcattagcaaaccgagctttca ggaatttgtggattgggaaaacgtgagcccggaactgaacagcaccgatcagccgtttTGGCAAGCCGGAA TCCTGGCCAGAAATCTGGTGCCTATGGTGGCCACAGTGCAGGGCCAGAACCTGAAGTACCAGgggtcagtca ctagtcatctctgcttctatcattgtcttcaacctgCtggaaactggaagggtgattatcgagatgatggcaa cgtgtgggtgcatacccccgctgagcccgcgccacctgaacgcgtgggtgaaagcgggtggaagaaaaaaaag gtattccagttcacctagagctggccagtatgaccaacaTggagctcatgagcagttattgtgcatcagcag gtcAGAACATACGGCCCCGTGTTTCATGTGTCTCGGCGGACTGCTTACAATGGTGGCTGGTGTGTGGCT GACAGTGcagtgctcgagctgttccggggccgcgcagctggccaacgacgtgggtcctccagatcatggagc tttgtaggtgcagcgtttccgaggtgtgccataccaccgtgcccgtggccgaacgcgagcctgaccccgaaa tggaaacaagcaaaacccagcccgatcgccaactgcagcgtgtatgactttttgtgtggtccatta ttattctgttcgagacacactttggccaagggtgacctaccatgaacaaatagcgtatcatatgtctgg aaagacgagccaaatataaaaagaGGACCAGGACCTGGCGCTAAATTTGTGGCCGCTGGACACTGAAAGCC GCTGCTGGTCTGGACCTGGCCAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTCGGACC CGGACCAGGCTGATGATTTTCGAAATTTAAATAAGCTTGCGGCCGCTAGGGATAACAGGGTAATtatacgcg ccaaacattttacagccgcggtgtcaaaaaccgcgtgg
Polipéptido de Ub7625merPDDT NoSFL (SEQ ID NO:37)
MQIFVKTLTGKTTITLEVEPSDTIENVKAKIQDKEGIPPDQQLIFAGKQLEDGRTLSDYNIQKES TLHLVLRLRGAMFQALSEGCTPYDINQMLNVLGDHGFKHIAFDRTFANNPGPMVVFATPGPILSPLTKGI LGFVFTLTVPSEGLSCISEADATTPESANLGEIILSQLYLWPRVITYHSPSYAYHQFERRAKYKRHPGFG QSLLFGYPVYVFGDCVQGDWDIAIRFRYCAPPGYALLRCNDTNYALLAVGALEGPQNQDWLGVPRLVTRM QAIQNAGLCTLVAMLEETIFWLQAFMLALTDSPGPKTNIIVDSQYVMGISKPSFQEFVDWENVSPELNSTDQ PFWQAGILARNLVPMVATVQGNLKYQGQSLVISASIIIVFNLLLELEGDYRDDGNVWVHTPLSPRTLNAWVK AVEEKKGIPVHLELASMTNMELMSSIVHQQVRTYGPVFMCLGGLLTMVAGAVWLTVRVLELFRAAQLANDV VLQIMELCGAAFRQVCHTTVPWPNASLTPKWNNETTQPQIANCSVYDFFVWLHYYSVRDTLWPRVITYHMNK YAYHMLERRAKYKRGPGPGAKFVAAWTLKAAAGPGPGQYIKANSKFIGITELGPGPG
ChAdV68.5WTnt.MAG25mer (SEQ ID NO:2); AC_000011.1 con las secuencias E1 (de nt 577 a 3.403) y E3 (de nt 27.125 - 31.825) eliminadas; nucleótidos de ATCC VR-594 correspondientes sustituidos en cinco posiciones; modelo de casete de neoantígeno bajo control del

promotor/potenciador del CMV insertado en lugar de la E1 eliminada;
SV40 polyA 3' del casete

CCATCTTCAATAATATACCTCAAACCTTTTTGTGCGCGTTAATATGCAAATGAGGCGTTTGAATTT
GGGGAGGAAGGGCGGTGATTGGTTCGAGGGATGAGCGACCGTTAGGGGCGGGGCGAGTGACGTTTTGATGAC
GTGGTTGCGAGGAGGAGCCAGTTTGCAAGTTCTCGTGGGAAAAGTGACGTCAAACGAGGTGTGGTTTTGAAC
ACGGAAATACTCAATTTTCCGCGCTCTCTGACAGGAAATGAGGTGTTTCTGGGCGGATGCAAGTGAAAAC
GGGCCATTTTCGCGCGAAAACCTGAATGAGGAAGTGAAAATCTGAGTAATTTTCGCGTTTTATGGCAGGGAGGA
GTATTTGCCGAGGGCCGAGTAGACTTTTGACCGATTACGTGGGGGTTTTCGATTACCGTGTTTTTACCTAAA
TTTCCGCGTACGGTGTCAAAGTCCGGTGTTTTTACGTAGGTGTCAGCTGATCGCCAGGGTATTTAAACCTG
CGCTCTCCAGTCAAGAGGCCACTCTTGAGTGCCAGCGAGAAGAGTTTTCTCCTCCGCGCCGCGAGTCAGAT
CTACACTTTGAAAGTAGGGATAACAGGGTAATGacattgattattgactagttGttaaTAGTAATCAATTA
CGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCGCCCATTTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGAC
TTTCCATTGAGTCAATGGGTGGAGTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATA
TGCCAAGTCCGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACC
TTACGGGACTTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGgTGATGCGGTTTTG
GCAGTACACCAATGGGCGTGGATAGCGGTTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTTGACGTCA
ATGGGAGTTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTGTAATAACCCCGCCCCGTTGACG
CAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAgcTCGTTTTAGTGAACCGTCAGATCGC
CTGGAACGCCATCCACGCTGTTTTGACCTCCATAGAAGACAGCGATCGCGccaccATGGCCGGGATGTTCC
AGGCACTGTCCGAAGGCTGCACACCCTATGATATTAACCAGATGCTGAATGTCCTGGGAGACCACCAGGTC
TCTGGCCTGGAGCAGCTGGAGAGCATCATCAACTTCGAGAAGCTGACCGAGTGGACAAGCTCCAATGTGAT
GCCTATCCTGTCCCCACTGACCAAGGGCATCCTGGGCTTCGTGTTTACCCTGACAGTGCCTTCTGAGCGGG
GCCTGTCTTGCATCAGCGAGGCAGACGCAACCACACCAGAGTCCGCCAATCTGGGCGAGGAGATCCTGTCT
CAGCTGTACCTGTGGCCCCGGGTGACATATCACTCCCCTTCTTACGCCTATCACCAGTTCGAGCGGAGAGC
CAAGTACAAGAGACACTTCCCAGGCTTTGGCCAGTCTCTGCTGTTTCGGCTACCCCGTGTACGTGTTTCGGCG
ATTGCGTGCAGGGCGACTGGGATGCCATCCGGTTTTAGATACTGCGCACCACTGGATATGCACTGCTGAGG
TGTAACGACACCAATTATTCCGCCCTGCTGGCAGTGGGCGCCCTGGAGGGCCCTCGCAATCAGGATTGGCT
GGGCGTGCCAAGGCAGCTGGTGACACGCATGCAGGCCATCCAGAACGCAGGCCTGTGCACCCTGGTGGCAA
TGCTGGAGGAGACAATCTTCTGGCTGCAGGCCTTTCTGATGGCCCTGACCGACAGCGGCCCAAGACAAAC
ATCATCGTAGATTCCAGTACGTGATGGGCATCTCCAAGCCTTCTTTCCAGGAGTTTGTGGAGTGGGAGAA
CGTGAGCCCGAGATGAATTCACCGATCAGCCATTCTGGCAGGCAGGAATCCTGGCAAGGAACCTGGTG
CTATGGTGGCCACAGTGCAGGGCCAGAATCTGAAGTACCAGGGCCAGAGCCTGGTCATCAGCGCCTCCATC
ATCGTGTTTTAACTGCTGGAGCTGGAGGGCGACTATCGGGACGATGGCAACGTGTGGGTGCACACCCCACT
GAGCCCCAGAACACTGAACGCCTGGGTGAAGGCCGTGGAGGAGAAGAAGGCATCCCAGTGCACCTGGAGC
TGGCCTCCATGACCAATATGGAGCTGATGTCTAGCATCGTGCAACAGCAGGTGAGGACATACGGACCCGTG
TTCATGTGCCTGGGAGGCCTGCTGACCATGGTGGCAGGAGCCGTGTGGCTGACAGTGCGGGTGCTGGAGCT
GTTTCAGAGCCGCCCAGCTGGCCAACGATGTGGTGTGTCAGATCATGGAGCTGTGCGGAGCAGCCTTTTCGCC
AGGTGTGCCACACCACAGTGCCATGGCCCAATGCCTCCCTGACCCCCAAGTGAACAATGAGACAACACAG
CCTCAGATCGCCAACCTGTAGCGTGTACGACTTCTTCGTGTGGCTGCACTACTATAGCGTGAGGGATACCCT
GTGGCCCCGCGTGACATACCACATGAATAAGTACGCCTATCACATGCTGGAGAGGCGCGCCAAGTATAAGA
GAGGCCCTGGCCCAGGCGCAAAGTTTGTGGCAGCATGGACCCTGAAGGCCGCCGCCGCCGCCGCCGCCGCC
CAGTATATCAAGGCTAACAGTAAGTTTATTGGAATCACAGAGCTGGGACCCGGACCTGGATAATGAGTTTA
AACTCCCATTTAAATGTGAGGGTTAATGCTTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAA
CCACAAC TAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACC
ATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGAT
GTGGGAGGTTTTTTAAAGCAAGTAAACCTCTACAAATGTGGTAAAATAACTATAACGGTCCTAAGGTAGC
GAGTGAGTAGTGTTCTGGGGCGGGGGAGGACCTGCATGAGGGCCAGAATAACTGAAATCTGTGCTTTTCTG
TGTGTTGACAGCATGAGCGGAAGCGGCTCCTTTGAGGGAGGGGTATTCAGCCCTTATCTGACGGGCGT
CTCCCCCTCCTGGGCGGGAGTGCGTCAGAATGTGATGGGATCCACGGTGGACGGCCGGCCGTGACGCCCGC
GAACCTCTTCAACCCTGACCTATGCAACCCTGAGCTCTTCGTGCTTGGACGCAGCTGCCGCCGCAGCTGCTG
CATCTGCCGCCAGCGCCGTGCGCGGAATGGCCATGGGCGCCGGCTACTACGGCACTCTGGTGGCCAACCTCG
AGTTCCACCAATAATCCCGCCAGCCTGAACGAGGAGAAGCTGTTGCTGCTGATGGCCAGCTCGAGGCCTT
GACCCAGCGCCTGGGCGAGCTGACCCAGCAGGTGGCTCAGCTGCAGGAGCAGACGCGGGCCGCGGTTGCCA
CGGTGAAATCCAAATAAAAAATGAATCAATAAATAAACGGAGACGGTTGTTGATTTTAACACAGAGTCTGA
ATCTTTATTTGATTTTTTCGCGCGCGGTAGGCCCTGGACCACCGGTCTCGATCATTGAGCACCCGGTGGATC
TTTTCCAGGACCCGGTAGAGGTGGGCTTGGATGTTGAGGTACATGGGCATGAGCCCGTCCCGGGGGTGGAG

GTAGCTCCATTGCAGGGCCTCGTGCTCGGGGGTGGTGTGTGTAATCACCCAGTCATAGCAGGGGCGCAGGG
 CATGGTGTTCACAAATATCTTTGAGGAGGAGACTGATGGCCACGGGCAGCCCTTTGGTGTAGGTGTTTACA
 AATCTGTTGAGCTGGGAGGGATGCATGCGGGGGGAGATGAGGTGCATCTTGGCCTGGATCTTGAGATTGGC
 GATGTTACCGCCCAGATCCCGCCTGGGGTTCATGTTGTGTCAGGACCACCAGCACGGTGTATCCGGTGCACCT
 TGGGGAATTTATCATGCAACTTGGAAGGGAAGGCGTGAAAGAATTTGGCGACGCCTTTGTGCCCGCCAGG
 TTTTCCATGCACTCATCCATGATGATGGCGATGGGCCCCGTGGGCGGCGGCCTGGGCAAAGACGTTTCGGGG
 GTCGGACACATCATAGTTGTGGTCCTGGGTGAGGTGCATCATAGGCCATTTTAATGAATTTGGGGCGGAGGG
 TGCCGGACTGGGGGACAAAGGTACCCTCGATCCCGGGGGCGTAGTTCCCTCACAGATCTGCATCTCCCAG
 GCTTTGAGCTCGGAGGGGGGATCATGTCCACCTGCGGGGCGATAAAGAACACGGTTTCCGGGGCGGGGGA
 GATGAGCTGGGCGCAAAGCAAGTTCCGGAGCAGCTGGGACTTGCCGCAGCCGGTGGGGCCGTAGATGACCC
 CGATGACCGGCTGCAGGTGGTAGTTGAGGGAGAGACAGCTGCCGTCTCCCGGAGGAGGGGGGCCACCTCG
 TTCATCATCTCGCGCACGTGCATGTTCTCGCGCACCACTTCCGCCAGGAGGCGCTCTCCCCCAGGGATAG
 GAGCTCCTGGAGCGAGGCGAAGTTTTTCAGCGGCTTGAGTCCGTCCGCCATGGGCATTTTGGAGAGGGTTT
 GTTGCAAGAGTTCCAGGCGGTCCAGAGCTCGGTGATGTGCTCTACGGCATCTCGATCCAGCAGACCTCCT
 CGTTTCGCGGGTGGGACGGCTGCGGGAGTAGGGCACCAGACGATGGGCGTCCAGCGCAGCCAGGGTCCGG
 TCCTTCCAGGGTCGCAGCGTCCGCGTCAGGGTGGTCTCCGTACGGTGAAGGGGTGCGCGCCGGGCTGGGC
 GCTTGCGAGGGTGCCTTCAGGCTCATCCGGTGGTTCGAAAACCGCTCCCGATCGGCGCCCTGCGCGTCGG
 CCAGGTAGCAATTGACCATGAGTTTCGTAGTTGAGCGCCTCGGCCGCGTGGCCTTTGGCGCGGAGCTTACCT
 TTGGAAGTCTGCCCCGAGGCGGGACAGAGGAGGGACTTGAGGGCGTAGAGCTTGGGGGCGAGGAAGACGGA
 CTCGGGGGCGTAGGCGTCCGCGCCGCACTGGGCGCAGACGGTCTCGCACTCCACGAGCCAGGTGAGGTTCGG
 GCTGGTTCGGGGTCAAAAACAGTTTCCCGCCGTTCTTTTTGATGCGTTTCTTACCTTTGGTCTCCATGAGC
 TCGTGTCCCCGCTGGGTGACAAAGAGGCTGTCCGTGTCCCCGTAGACCGACTTTATGGGCGGGTCCTCGAG
 CGGTGTGCCGCGGTCTCCTCGTAGAGGAACCCCGCCACTCCGAGACGAAAGCCCGGGTCCAGGCCAGCA
 CGAAGGAGGCCACGTGGGACGGGTAGCGGTGCTTGTCCACCAGCGGGTCCACCTTTTCCAGGGTATGCAAA
 CACATGTCCCCCTCGTCCACATCCAGGAAGGTGATTGGCTTGTAAAGTGTAGGCCACGTGACCGGGGGTCCC
 GGCCGGGGGGGTATAAAAGGGTGCGGGTCCCTGCTCGTCCCTCACTGTCTTCCGGATCGCTGTCCAGGAGCG
 CCAGCTGTTGGGGTAGGTATTCCCTCTCGAAGGCGGGCATGACCTCGGCACTCAGGTTGTGAGTTTCTAGA
 AACGAGGAGGATTTGATATTGACGGTGCCGGCGGAGATGCCTTTCAAGAGCCCCCTCGTCCATCTGGTCAGA
 AAAGACGATCTTTTTGTTGTGTCGAGCTTGGTGGCGAAGGAGCCGTAGAGGGCGTTGGAGAGGAGCTTGGCGA
 TGGAGCGCATGGTCTGGTTTTTTTCTTGTGCGCGCGCTCCTTGGCGGCGATGTTGAGCTGCACGTACTION
 TCGGCCACGCACTTCCATTTCGGGGAAGACGGTGGTCAGCTCGTCCGGCACGATTCTGACCTGCCAGCCCCG
 ATTATGCAGGGTATGAGGTCCACACTGGTGGCCACCTCGCCGCGCAGGGGCTCATTAGTCCAGCAGAGGC
 GTCCGCCCTTGCGCGAGCAGAAGGGGGGAGGGGGTCCAGCATGACCTCGTCCGGGGGGTTCGGCATCGATG
 GTGAAGATGCCGGGAGGAGGTGCGGGTCAAAGTAGCTGATGGAAGTGGCCAGATCGTCCAGGGCAGCTTG
 CCATTGCGGCACGGCCAGCGCGCTCGTAGGGACTGAGGGGCGTGCCCCAGGGCATGGGATGGGTAAGCG
 CGGAGGCGTACATGCCGAGATGTGCTAGACGTAGAGGGGCTCCTCGAGGATGCCGATGTAGGTGGGGTAG
 CAGCGCCCCCGCGGATGCTGGCGCGCACGTAGTCATACAGCTCGTGCAGGGGGGCGAGGAGCCCCGGGCC
 CAGGTTGGTTCGCACTGGGCTTTTCGGCGCGGTAGACGATCTGGCGGAAAATGGCATGCGAGTTGGAGGAGA
 TGGTGGGCTTTTGAAGATGTTGAAGTGGGCGTGGGGCAGTCCGACCGAGTCGCGGATGAAGTGGGCGTAG
 GAGTCTTGACGCTTGGCGACGAGCTCGGCGGTGACTAGGACGTCCAGAGCGCAGTAGTCGAGGGTCTCCTG
 GATGATGTCACTTGTAGCTGTCCCTTTTGTTCACAGCTCGCGGTTGAGAAGGAACCTCTTCGCGGTCTC
 TCCAGTACTCTTCGAGGGGGAACCCGTCTGATCTGCACGGTAAGAGCCTAGCATGTAGAACTGGTTGACG
 GCCTTGTAGGCGCAGCAGCCCTTCTCCACGGGGAGGGCGTAGGCCTGGGCGGCCTTGCAGGGGAGGTGTG
 CGTGAGGGGCGAAAGTGTCCCTGACCATGACCTTGAAGTGGTGTGAAAGTGCATATCGTTCGACGCCCC
 CCTGCTCCCAGAGCTGGAAGTCCGTGCGCTTCTTGTAGGCGGGGTGGGCAAAGCGAAAGTAACATCGTTG
 AAGAGGATCTTGCCCGCGCGGGGCATAAAGTTGCGAGTGATGCGGAAAGGTTGGGGCACCTCGGCCCGGT
 GTTGTGACCTGGGCGGCGAGCACGATCTCGTGAAGCCGTTGATGTTGTGGCCACGATGTAGAGTTCCA
 CGAATGCGGACGCGCCCTTGACGTGGGGCAGTTTCTTGAGCTCCTCGTAGGTGAGCTCGTCGGGTCGCTG
 AGCCCGTGTGCTCGAGCGCCAGTCGGCGAGATGGGGGTTGGCGCGGAGGAAGGAAGTCCAGAGATCCAC
 GGCCAGGGCGGTTTGACAGCGGTCCCGGTACTGACGGAACGTGCTGCCCGACGGCCATTTTTTCGGGGGTGA
 CGCAGTAGAAGGTGCGGGGGTCCCCGTGCCAGCGATCCCATTTGAGCTGGAGGGCGAGATCGAGGGCGAGC
 TCGACGAGCCGGTCTGTCGCCGAGAGTTTCATGACCAGCATGAAGGGGACGAGCTGCTTGCCGAAGGACCC
 CATCCAGGTGTAGGTTTCCACATCGTAGGTGAGGAAGAGCCTTTCCGGTGCAGGATGCGAGCCGATGGGGA
 AGAACTGGATCTCCTGCCACCAATTGGAGGAATGGCTGTTGATGTGATGGAAGTAGAAATGCCGACGGCGC
 GCCGAACACTCGTGCTTGTGTTTATACAAGCGGCCACAGTGCTCGCAACGCTGCACGGGATGCACGTGCTG
 CACGAGCTGTACCTGAGTTCTTTGACGAGGAATTTTCACTGGGAAGTGGAGTCGTGGCGCCTGCATCTCGT
 GCTGTACTACGTGCTGGTGGTTCGGCCTGGCCCTCTTCTGCCTCGATGGTGGTCATGCTGACGAGCCCCGCGC

CGTGGTGCCCTCGCGCTCCAACCCACGCACGAGAAGGTCTGGCCATCGTGAACGCGCTGGTGGAGAACA
 AGGCCATCCGCGGCGACGAGGCCGGCTGGTGTACAACGCGCTGCTGGAGCGCGTGGCCCGCTACAACAGC
 ACCAACGTGCAGACCAACCTGGACCGCATGGTGACCGACGTGCGCGAGGCCGTGGCCAGCGCGAGCGGTT
 CCACCGCGAGTCCAACCTGGGATCCATGGTGGCGCTGAACGCCTTCCTCAGCACCCAGCCCCCAACGTGC
 CCCGGGGCCAGGAGGACTACACCAACTTCATCAGCGCCCTGCGCCTGATGGTGACCGAGGTGCCCCAGAGC
 GAGGTGTACAGTCCGGGCCGACTACTTCTTCCAGACAGTCGCCAGGGCTTGACAGACCGTGAACCTGAG
 CCAGGCTTTCAAGAACTTGAGGGCCTGTGGGGCGTGAGGCCCGGTGCGGGACCGCGCGACGGTGTGCA
 GCCTGCTGACGCCGAACCTGCGCCTGCTGCTGCTGGTGGCCCCCTTCACGGACAGCGGCAGCATCAAC
 CGCAACTCGTACCTGGGCTACCTGATTAACCTGTACCGCGAGGCCATCGGCCAGGCGCACGTGGACGAGCA
 GACCTACCAGGAGATCACCCACGTGAGCCGCGCCCTGGGCCAGGACGACCCGGGCAACCTGGAAGCCACCC
 TGAACCTTTTGTGCTGACCAACCGGTGCGAGAAGATCCCGCCCCAGTACGCGCTCAGCACCGAGGAGGAGCGC
 ATCCTGCGTTTACGTGCAGCAGAGCGTGGGCCCTGTTCTGATGACAGGAGGGGGCCACCCCCAGCGCCGCGCT
 CGACATGACCGCGCGCAACATGGAGCCAGCATGTACGCCAGCAACCGCCGTTTCATCAATAAACTGATGG
 ACTACTTGCATCGGGCGGCCCATGAACCTCTGACTATTTTACCAACGCCATCCTGAATCCCCACTGGCTC
 CCGCCCGGGGTTCTACACGGGCGAGTACGACATGCCGACCCCAATGACGGGTTCTGTGGGACGATGT
 GGACAGCAGCGTGTCTCCCCCGACCGGGTGCTAACGAGCGCCCCCTTGTGGAAGAAGGAAGGCAGCGACC
 GACGCCCGTCTCGGCGCTGTCCGGCCGCGAGGGTGCTGCCGCGGCGGTGCCGAGGCCGCCAGTCTTTTC
 CCGAGCTTGCCCTTCTCGCTGAACAGTATCCGCGAGCAGCGAGCTGGGCAGGATCACGCGCCCGCGCTTGCT
 GGGCGAAGAGGAGTACTTGAATGACTCGCTGTTGAGACCCGAGCGGGAGAAGAACTTCCCCAATAACGGGA
 TAGAAAGCCTGGTGGACAAGATGAGCCGCTGGAAGACGTATGCGCAGGAGCACAGGGACGATCCCCGGGCG
 TCGCAGGGGGCCACGAGCCGGGGCAGCGCCGCCCGTAAACGCCGGTGGCACGACAGGCAGCGGGGACAGAT
 GTGGGACGATGAGGACTCCGCCGACGACAGCAGCGTGTGGACTTGGGTGGGAGTGGTAACCGTTTCGCTC
 ACCTGCGCCCCGTATCGGGCGCATGATGTAAGAGAAACCGAAAATAAATGATACTCACCAAGGCCATGGC
 GACCAGCGTGCCTTCGTTTCTTCTCTGTTGTTGTTGTATCTAGTATGATGAGGCGTGCCTACCCGGAGGGT
 CCTCTCCCTCGTACGAGAGCGTGATGCAGCAGGCGATGGCGGCGGCGGCGATGCAGCCCCCGCTGGAGGC
 TCCTTACGTGCCCCCGCGGTACCTGGCGCCTACGGAGGGGCGGAACAGCATTCGTTACTCGGAGCTGGCAC
 CCTTGTACGATACCACCCGGTTGTACCTGGTGGACAACAAGTCGGCGGACATCGCCTCGCTGAACTACCAG
 AACGACCACAGCAACTTCCTGACCACCGTGGTGCAGAACATGACTTCACCCCCACGGAGGCCAGCACCCA
 GACCATCAACTTTGACGAGCGCTCGCGGTGGGGCGGCCAGCTGAAAACCATCATGCACACCAACATGCCCA
 ACGTGAACGAGTTTCATGTACAGCAACAAGTTCAAGGCGCGGGTGATGGTCTCCCGCAAGACCCCCAATGGG
 GTGACAGTGACAGAGGATTATGATGGTAGTCAGGATGAGCTGAAGTATGAATGGGTGGAATTTGAGCTGCC
 CGAAGGCAACTTCTCGGTGACCATGACCATCGACCTGATGAACAACGCCATCATCGACAATTACTTGGCGG
 TGGGGCGGCAGAACGGGGTGCTGGAGAGCGACATCGGCGTGAAGTTTCGACACTAGGAACCTTCAGGCTGGGC
 TGGGACCCCGTGACCGAGCTGGTCATGCCCGGGTGATACCAACGAGGCTTTCCATCCCGATATTGTCTT
 GCTGCCCGGCTGCGGGGTGGAATTCACCGAGAGCGCCTCAGCAACCTGCTGGGCATTTCGAAGAGGCAGC
 CCTTCCAGGAAGGCTTCCAGATCATGTACGAGGATCTGGAGGGGGCAACATCCCCGCGCTCCTGGATGTC
 GACGCCTATGAGAAAAGCAAGGAGGATGCAGCAGCTGAAGCAACTGCAGCCGTAGCTACCGCCTCTACCGA
 GGTGAGGGGCGATAATTTTGAAGCGCCGCGAGCAGTGGCAGCGGCCGAGGCGGCTGAAACCGAAAGTAAGA
 TAGTCATTACGCCGGTGGAGAAGGATAGCAAGAACAGGAGCTACAACGTACTACCGGACAAGATAAACACC
 GCCTACCGCAGCTGGTACCTAGCCTACAACATATGGCGACCCGAGAAGGGCGTGCGCTCCTGGACGCTGCT
 CACCACCTCGGACGTACCTGCGGCGTGGAGCAAGTCTACTGGTCGCTGCCCGACATGATGCAAGACCCGG
 TCACCTTCCGCTCCACGCGTCAAGTTAGCAACTACCCGGTGGTGGGCGCCGAGCTCCTGCCCGTCTACTCC
 AAGAGCTTCTTCAACGAGCAGGCCGTCTACTCGCAGCAGCTGCGCGCCTTCACCTCGCTTACGCACGTCTT
 CAACCGCTTCCCCGAGAACCAGATCCTCGTCCGCCCGCCCGCGCCACCATTACCACCGTCAGTGAACACG
 TTCCTGCTCTCACAGATCACGGGACCCTGCCGCTGCGCAGCAGTATCCGGGGAGTCCAGCGCGTGACCGTT
 ACTGACGCCAGACGCCGACCTGCCCCCTACGTCTACAAGGCCCTGGGCATAGTCGCGCCGCGCGCTCCTCTC
 GAGCCGCACCTTCTAAATGTCCATTCTCATCTCGCCAGTAATAACACCGGTTGGGGCCTGCGCGCGCCCA
 GCAAGATGTACGGAGGCGCTCGCCAACGCTCCACGCAACACCCCGTGCGCGTGCGCGGCGCACTTCCGCGCT
 CCTGGGGCGCCCTCAAGGGCCGCGTGCGGTCGCGCACACCCGTCGACGACGTGATCGACAGGTGGTGGC
 CGACGCGCGCAACTACACCCCGCCGCGCGCCGCTTCCACCGTGGACGCCGTATCGACAGCGTGGTGG
 CcGACGCGCGCGGTACGCCCGCGCCAAGAGCCGGCGGCGGCGCATCGCCCGGCGGCACCGGAGCACCCCC
 GCCATGCGCGCGGCGCGAGCCTTGCTGCGCAGGGCCAGGCGCACGGGACGACGGGCCATGCTCAGGGCGGC
 CAGACGCGCGGCTTCAGGCGCCAGCGCCGGCAGGACCCGGAGACGCGCGGCCACGGCGGCGGCAGCGGCCA
 TCGCCAGCATGTCCCGCCCGCGGCGAGGGAACGTGTACTGGGTGCGCGACGCCGCCACCGGTGTGCGCGTG
 CCCGTGCGCACCCGCCCCCCCTCGCACTTGAAGATGTTCACTTCGCGATGTTGATGTGTCCCAGCGGCGAGG
 AGGATGTCCAAGCGCAAATTCAGGAAGAGATGCTCCAGGTATCGCGCCTGAGATCTACGGCCCTGCGGT
 GGTGAAGGAGGAAAGAAAGCCCCGAAAATCAAGCGGGTCAAAAAGGACAAAAGGAAGAAGAAAGTGATG

TGGACGGATTGGTGGAGTTTGTGCGCGAGTTCGCCCCCGGGCGGCGCTGCAGTGGCGCGGGCGGAAGGTG
 CAACCGGTGCTGAGACCCGGCACCACCGTGGTCTTCACGCCCGGCGAGCGCTCCGGCACCGCTTCCAAGCG
 CTCCTACGACGAGGTGTACGGGGATGATGATATTCTGGAGCAGGCGGCCGAGCGCCTGGGCGAGTTTGTCTT
 ACGGCAAGCGCAGCCGTTCCGCACCGAAGGAAGAGGCGGTGTCCATCCCGCTGGACCACGGCAACCCACG
 CCGAGCCTCAAGCCCGTGACCTTGCAGCAGGTGCTGCCGACCGCGGCGCCGCGCCGGGGGTTCAAGCGCGA
 GGGCGAGGATCTGTACCCACCATGCAGCTGATGGTGCCCAAGCGCCAGAAGCTGGAAGACGTGCTGGAGA
 CCATGAAGGTGGACCCGGACGTGCAGCCCCGAGGTCAAGGTGCGGCCCATCAAGCAGGTGGCCCCGGGCGCTG
 GGGGTGCAGACCGTGGACATCAAGATTCCACGAGGCCATGGAAACGCAGACCGAGCCCATGATCAAGCC
 CAGCACCAGCACCATGGAGGTGCAGACGGATCCCTGGATGCCATCGGCTCCTAGTCGAAGACCCCGGCGCA
 AGTACGGCGCGGCCAGCCTGCTGATGCCCAACTACGCGCTGCATCCTTCCATCATCCCCACGCCGGGCTAC
 CGGGCACGCGCTTCTACCGCGGTACATACCAGCAGCCGCCGCCGAAGACCACCACTCGCCGCCGCCGCTCG
 CCGCACCGCGCTGCAACCACCCCTGCCGCCCTGGTGCGGAGAGTGTACCGCCGCGGCCGCGCACCTCTGA
 CCCTGCCGCGCGCGCTACCACCCGAGCATCGCCATTTAAACTTTTCGCCtGCTTTGCAGATCAATGGCCC
 TCACATGCCGCTTTCGCTTCCCATTACGGGCTACCGAGGAAGAAAACCGCGCCGTAGAAGGTGGCGGGG
 AACGGGATGCGTCGCCACCACCACCGGCGGCGCGCCATCAGCAAGCGGTTGGGGGGAGGCTTCTGCC
 CGCGCTGATCCCCATCATCGCCGCGGCGATCGGGGCGATCCCCGGCATTGCTTCCGTGGCGGTGCAGGCCT
 CTCAGCGCCACTGAGACACACTTGGAAACATCTTGTAAATAAACCaATGGACTCTGACGCTCCTGGTCCTGT
 GATGTGTTTTTCGTAGACAGATGGAAGACATCAATTTTTTCGTCCCTGGCTCCGCGACACGGCACGCGGCCGT
 TCATGGGCACCTGGAGCGACATCGGCACCAGCCAACTGAACGGGGGCGCCTTCAATTGGAGCAGTCTCTGG
 AGCGGGCTTAAGAATTTTCGGGTCCACGCTTAAAACCTATGGCAGCAAGGCGTGGAACAGCACCACAGGGCA
 GGGCTGAGGGATAAGCTGAAAGAGCAGAACTTCCAGCAGAAGGTGGTGCATGGGCTCGCCTCGGGCATCA
 ACGGGGTGGTGGACCTGGCCAACCAGGCCGTGCAGCGGCAGATCAACAGCCGCTGGACCCGGTGCCGCC
 GCCGGCTCCGTGGAGATGCCGAGGTGGAGGAGGAGCTGCCTCCCCTGGACAAGCGGGGCGAGAAGCGACC
 CCGCCCCGATGCGGAGGAGACGCTGCTGACGCACACGGACGAGCCGCCCCCGTACGAGGAGGCGGTGAAAC
 TGGGTCTGCCCACCACGCGGCCCATCGCGCCCCTGGCCACCGGGGTGCTGAAACCCGAAAAGCCCCGCGACC
 CTGGACTTGCTCCTCCCCAGCCTTCCCGCCCCTCTACAGTGGCTAAGCCCCCTGCCGCCGGTGGCCGTGGC
 CCGCGCGGACCCGGGGGACCGCCCCGCCCTCATGCGAACTGGCAGAGCACTCTGAACAGCATCGTGGGTG
 TGGGAGTGCAGAGTGTGAAGCGCCGCCGCTGCTATTAAACCTACCGTAGCGCTTAACTTGCTTGTCTGTGT
 GTGTATGTATTATGTCGCCGCCGCCGCTGTCCACCAGAAGGAGGAGTGAAGAGGCGCGTCCGCCAGTTGCA
 AGATGGCCACCCCATGATGCTGCCCCAGTGGGCGTACATGCACATCGCCGGACAGGACGCTTCGGAGTAC
 CTGAGTCCGGGTCTGGTGCAGTTTGCCCGCGCCACGACACCTACTTTCAGTCTGGGGAACAAGTTTAGGAA
 CCCCACGGTGGCGCCACGCACGATGTGACCACCGACCGCAGCCAGCGGCTGACGCTGCGCTTCGTGCCCG
 TGGACCGCGAGGACAACACCTACTCGTACAAAGTGCCTACACGCTGGCCGTGGGCGACAACCGCGTGCTG
 GACATGGCCAGCACCTACTTTGACATCCGCGGCGTGCTGGATCGGGGCCCTAGCTTCAAACCTACTCCGG
 CACCGCCTACAACAGTCTGGCCCCCAAGGGAGCACCCAACACTTGTGAGTGGACATATAAAGCCGATGGTG
 AAACCTGCCACAGAAAAAACCTATACATATGGAATGCACCCGTGCAGGGCATTAACATCACAAAAGATGGT
 ATTCAACTTGGAAGTGCACCGATGATCAGCCAATCTACGCAGATAAAACCTATCAGCCTGAACCTCAAGT
 GGGTGATGCTGAATGGCATGACATCACTGGTACTGATGAAAAGTATGGAGGCAGAGCTCTTAAGCCTGATA
 CCAAATGAAGCCTTGTTATGGTTCTTTTGCCAAGCCTACTAATAAAGAAGGAGGTGAGGCAAATGTGAAA
 ACAGGAACAGGCACCTACTAAAGAATATGACATAGACATGGCTTTCTTTGACAACAGAAGTGCGGCTGCTGC
 TGGCCTAGCTCCAGAAATTGTTTTGTATACTGAAAATGTGGATTTGGAAACTCCAGATACCCATATTGTAT
 ACAAAGCAGGCACAGATGACAGCAGCTCTTCTATTAATTTGGGTGAGCAAGCCATGCCCAACAGACCTAAC
 TACATTGGTTTTAGAGACAACCTTTATCGGGCTCATGTACTACAACAGCACTGGCAATATGGGGGTGCTGGC
 CGGTGAGGCTTCTCAGCTGAATGCTGTGGTTGACTTGCAAGACAGAAACACCGAGCTGTCTACCAGCTCT
 TGCTTGACTCTCTGGGTGACAGAACCCGGTATTTAGTATGTGGAATCAGGCGGTGGACAGCTATGATCCT
 GATGTGCGCATTATTGAAAATCATGGTGTGGAGGATGAACTTCCCAACTATTGTTTCCCTCTGGATGCTGT
 TGGCAGAACAGATACTTATCAGGGAATTAAGGCTAATGGAACGATCAAAACCACATGGACCAAAGATGACA
 GTGTCAATGATGCTAATGAGATAGGCAAGGGTAATCCATTTCGCATGGAAATCAACATCCAAGCCAACCTG
 TGGAGGAACCTTCTCTACGCCAACGTGGCCCTGTACTGCCGACTCTTACAAGTACACGCCGGCCAATGT
 TACCCTGCCACCAACACCAACCTACGATTACATGAACGGCCGGGTGGTGGCGCCCTCGCTGGTGGACT
 CCTACATCAACATCGGGGCGCGCTGGTGCCTGGATCCCATGGACAACGTGAACCCCTTCAACCACCACCGC
 AATGCGGGGCTGCGCTACCGCTCCATGCTCCTGGGCAACGGGCGCTACGTGCCCTTCCACATCCAGGTGCC
 CCAGAAATTTTTGCCCATCAAGAGCCTCCTGCTCCTGCCCGGGTCTACACCTACGAGTGGAACCTCCGCA
 AGGACGTCAACATGATCCTGCAGAGCTCCCTCGGCAACGACCTGCGCACGGACGGGGCCTCCATCTCCTTC
 ACCAGCATCAACCTCTACGCCACCTTCTTCCCCATGGCGCACAAACACGGCCTCCACGCTCGAGGCCATGCT
 GCGCAACGACACCAACGACCAGTCTTCAACGACTACCTCTCGGCGGCCAACATGCTCTACCCCATCCCGG
 CCAACGCCACCAACGTGCCCATCTCCATCCCCTCGCGCAACTGGGCCGCTTCCGCGGCTGGTCTCTCACG

CGTCTCAAGACCAAGGAGACGCCCTCGCTGGGCTCCGGGTTTCGACCCCTACTTCGTCTACTCGGGCTCCAT
 CCCCTACCTCGACGGCACCTTCTACCTCAACCACACCTTCAAGAAGGTCTCCATCACCTTCGACTCCTCCG
 TCAGCTGGCCCCGGCAACGACCGGCTCCTGACGCCCCAACGAGTTCGAAATCAAGCGCACCGTCGACGGCGAG
 GGCTACAACGTGGCCAGTGCAACATGACCAAGGACTGGTTCTGGTCCAGATGCTGGCCCACTACAACAT
 CGGCTACCAGGGCTTCTACGTGCCCCGAGGGCTACAAGGACCGCATGTACTCCTTCTTCCGCAACTTCCAGC
 CCATGAGCCGCCAGGTGGTGGACGAGGTCAACTACAAGGACTACCAGGCCGTACCCCTGGCCTACCAGCAC
 AACAACCTCGGGCTTTCGTGCGGTACCTCGCGCCCCACCATGCGCCAGGGCCAGCCCTACCCCGCCAACCTACCC
 CTACCCGCTCATCGGCAAGAGCGCGGTACCCAGCGTCACCCAGAAAAAGTTCTCTGCGACAGGGTTCATGT
 GGCGCATCCCCCTTCTCCAGCAACTTCATGTCCATGGGCGCGCTCACCGACCTCGGCCAGAACATGCTCTAT
 GCCAACTCCGCCCCACGCGCTAGACATGAATTTTCAAGTTCGACCCCATGGATGAGTCCACCCTTCTCTATGT
 TGTCTTTCGAAGTCTTCGACGTCTGTCGAGTGCACAGCCCCACCGCGCGTTCATCGAGGCCGTCTACCTGC
 GCACCCCTTCTCGGCCGGTAACGCCACCACCTAAGCTCTTGCTTCTTGAAGCCATGGCCGCGGGCTCCG
 GCGAGCAGGAGCTCAGGGCCATCATCCGCGACCTGGGCTGCGGGCCCTACTTCTGGGCACCTTCGATAAG
 CGCTTCCCGGGATTTCATGGCCCCGCACAAGCTGGCCTGCGCCATCGTCAACACGGCCGGCCGCGAGACCGG
 GGGCGAGCACTGGCTGGCCTTCGCTGGAACCCGCGCTCGAACACCTGCTACCTCTTCGACCCCTTCGGGT
 TCTCGGACGAGCGCTCAAGCAGATCTACCAGTTCGAGTACGAGGGCCTGCTGCGCCGAGCGCCCTGGCC
 ACCGAGGACCGCTGCGTCACCCCTGGAAGTCCACCCAGACCGTGCAGGGTCCGCGCTCGGCCGCTGCGG
 GCTCTTCTGCTGCATGTTTCTGACGCCTTCGTGCACTGGCCCCACCGCCCCATGGACAAGAACCCACCA
 TGAACCTTGCTGACGGGGGTGCCAACGGCATGCTCCAGTCGCCCCAGGTGGAACCCACCCTGCGCCGCAAC
 CAGGAGGCGCTCTACCGCTTCTCAACTCCCACTCCGCTACTTTTCGCTCCACCGCGCGCGCATCGAGAA
 GGCCACCGCCTTCGACCGCATGAATCAAGACATGTAAACCGTGTGTGTATGTTAAATGTCTTTAATAACA
 GCACCTTTCATGTTACACATGCATCTGAGATGATTTATTTAGAAATCGAAAGGGTCTGCCGGGTCTCGGCA
 TGGCCCGCGGGCAGGGACACGTTGCGGAACCTGGTACTTGGCCAGCCACTTGAACCTCGGGGATCAGCAGTTT
 GGGCAGCGGGGTGTGCGGGAAGGAGTCGGTCCACAGCTTCCGCGTCAGTTGCAGGGCGCCCAGCAGGTTCGG
 GCGCGGAGATCTTGAAATCGCAGTTGGGACCCGCGTTCTGCGCGCGGGAGTTGCGGTACACGGGGTTGCAG
 CACTGGAACACCATCAGGGCCGGGTGCTTCACGCTCGCCAGCACCGTTCGCGTTCGGTGATGCTCTCCACGTC
 GAGGTCTTCGGCGTTGGCCATCCCGAAGGGGGTTCATCTTGACAGGTCTGCCTTCCCATGGTGGGCACGCACC
 CGGGCTTGTGGTTGCAATCGCAGTGCAGGGGGATCAGCATCATCTGGGCCTGGTTCGGCGTTCATCCCCGGG
 TACATGGCCTTCATGAAAGCCTCCAATTGCCTGAACGCCTGCTGGGCCTTGGCTCCCTCGGTGAAGAAGAC
 CCCGCAGGACTTGCTAGAGAACTGGTTGGTGGCGCACCCGGCGTTCGTGCACGCAGCAGCGCGCTCGTTGT
 TGGCAGCTGCACCACGCTGCGCCCCAGCGGTTCTGGGTGATCTTGGCCCGGTTCGGGGTCTCTCTTCAGC
 GCGCGCTGCCCCTTCTCGCTCGCCACATCCATCTCGATCATGTGCTCCTTCTGGATCATGGTGGTCCCGTG
 CAGGCACCGCAGCTTGCCTTCGGCTCGGTGCACCCGTGCAGCCACAGCGCGCACCCGGTGCACCTCCAGT
 TCTTGTGGGCGATCTGGGAATGCGCGTGCACGAAGCCCTGCAGGAAGCGGCCCATCATGGTGGTTCAGGGTC
 TTGTTGCTAGTGAAGGTGAGCGGAATGCCGCGGTGCTCCTCGTTGATGTACAGGTGGCAGATGCGGCGGTA
 CACCTCGCCCTGCTCGGGCATCAGCTGGAAGTTGGCTTTTCAGGTTCGGTCTCCACGCGGTAGCGGTCCATCA
 GCATAGTCATGATTTCCATACCTTCTCCAGGCCGAGACGATGGGCAGGCTCATAGGGTCTTTCACCATC
 ATCTTAGCGCTAGCAGCCGCGGCCAGGGGGTTCGCTCTCGTCCAGGGTCTCAAAGCTCCGCTTGCCGTCTT
 CTCGGTGATCCGCACCGGGGGGTAGCTGAAGCCACGGCCGCCAGCTCCTCCTCGGCCTGTCTTTCGTCTCT
 CGCTGTCTGGCTGACGTCTGTCAGGACCACATGCTTGGTCTTTCGGGGTTTCTTCTTGGGCGGCAGCGGC
 GGCGGAGATGTTGGAGATGGCGAGGGGGAGCGCGAGTTCTCGCTCACCCTACTATCTCTTCTCTTCTTG
 GTCCGAGGCCACGCGCGGTAGGTATGTCTCTTTCGGGGGCGAGGGCGGAGGCGACGGGCTCTCGCCGCCGC
 GACTTGGCGGATGGCTGGCAGAGCCCTTCCGCGTTTCGGGGGTGCGCTCCCGGCGGCGCTCTGACTGACTT
 CCTCCGCGGCGCGCCATTGTGTTCTCTTAGGGAGGAACAACAAGCATGGAGACTCAGCCATCGCCAACCTC
 GCCATCTGCCCCCACCGCCGACGAGAAGCAGCAGCAGCAGAATGAAAGCTTAACCGCCCCGCGCCAGCC
 CCGCCACCTCCGACGCGCGCGTCCCAGACATGCAAGAGATGGAGGAATCCATCGAGATTGACCTGGGCTAT
 GTGACGCCCCGCGGAGCACGAGGAGGAGCTGGCAGTGCCTTTTCAAGAAGAGATACACCAAGAAGAGCC
 AGAGCAGGAAGCAGAGAATGAGCAGAGTCAGGCTGGGCTCGAGCATGACGGCGACTACCTCCACCTGAGCG
 GGGGGGAGGACGCGCTCATCAAGCATCTGGCCCCGCGAGGCCACCATCGTCAAGGATGCGCTGCTCGACCGC
 ACCGAGGTGCCCCCTCAGCGTGGAGGAGCTCAGCCGCGCCTACGAGTTGAACCTCTTCTCGCCGCGCGTGCC
 CCCCAGCGCCAGCCCAATGGCACCTGCGAGCCCAACCCGCGCCTCAACTTCTACCCGGTCTTTCGCGGTGC
 CCGAGGCCCTGGCCACCTACCACATCTTTTTCAAGAACCAAAAGATCCCCGTCTCCTGCCGCGCCAACCGC
 ACCCGCGCCGACGCCCTTTTCAACCTGGGTCCCGGCGCGCCGCTACCTGATATCGCCTCCTTGGAAGAGGT
 TCCCAAGATCTTCGAGGGTCTGGGCAGCGACGAGACTCGGGCCGCAACGCTCTGCAAGGAGAAGGAGGAG
 AGCATGAGCACACAGCGCCCTGGTTCGAGTTGGAAGGCGACAACGCGCGGCTGGCGGTGCTCAAACGCACG
 GTCGAGCTGACCCATTTTCGCTACCCGGCTCTGAACCTGCCCCCAAAGTCATGAGCGCGGTTCATGGACCA
 GGTGCTCATCAAGCGCGCTCGCCCATCTCCGAGGACGAGGGCATGCAAGACTCCGAGGAGGGCAAGCCG

TGGTCAGCGACGAGCAGCTGGCCCGGTGGCTGGGTCTAATGCTAGTCCCCAGAGTTTGAAGAGCGGCGC
 AAACCTCATGATGGCCGTGGTCCTGGTGACCGTGAGCTGGAGTGCCTGCGCCGCTTCTTCGCCGACGCGGA
 GACCCCTGCGCAAGGTCGAGGAGAACCTGCACTACCTCTTCAGGCACGGGTTCGTGCGCCAGGCCTGCAAGA
 TCTCCAACGTGGAGCTGACCAACCTGGTCTCCTACATGGGCATCTTGCACGAGAACCGCCTGGGGCAGAAC
 GTGCTGCACACCACCCTGCGCGGGGAGGCCCGGCGCGACTACATCCGCGACTGCGTCTACCTCTACCTCTG
 CCACACCTGGCAGACGGGCATGGGCGTGTGGCAGCAGTGTCTGGAGGAGCAGAACCTGAAAGAGCTCTGCA
 AGCTCCTGCAGAAGAACCTCAAGGGTCTGTGGACCGGGTTTCGACGAGCGCACACCACCGCCTCGGACCTGGCC
 GACCTCATTTTTCCCCGAGCGCCTCAGGCTGACGCTGCGCAACGGCCTGCCCGACTTTATGAGCCAAAGCAT
 GTTGCAAAACTTTTCGCTCTTTTCATCCTCGAACGCTCCGGAATCCTGCCCGCCACCTGCTCCGCGCTGCCCT
 CGGACTTCGTGCCGCTGACCTTCCGCGAGTGCCCCCGCCGCTGTGGAGCCACTGCTACCTGCTGCGCCTG
 GCCAACTACCTGGCCTACCACTCGGACGTGATCGAGGACGTGAGCGGCGAGGGCCTGCTCGAGTGCCACTG
 CCGCTGCAACCTCTGCACGCCGCACCGCTCCCTGGCTGCAACCCCCAGCTGCTGAGCGAGACCCAGATCA
 TCGGCACCTTTCGAGTTGCAAGGGCCAGCGAAGGCGAGGGTTTCAGCCGCCAAGGGGGGTCTGAAACTCACC
 CCGGGGCTGTGGACCTCGGCCTACTTTCGCGCAAGTTTCGTGCCCGAGGACTACCATCCCTTCGAGATCAGGTT
 CTACGAGGACCAATCCCATCCGCCCAAGGCCGAGCTGTGCGCCTGCGTCATCACCCAGGGGGCGATCCTGG
 CCCAATTGCAAGCCATCCAGAAATCCCGCCAAGAATTCTTGCTGAAAAAGGGCCGCGGGGTCTACCTCGAC
 CCCCAGACCGGTGAGGAGCTCAACCCCGGCTTCCCCCAGGATGCCCGGAGGAAACAAGAAGCTGAAAGTGG
 AGCTGCCGCCCGTGGAGGATTTGGAGGAAGACTGGGAGAACAGCAGTCAGGCAGAGGAGGAGGAGATGGAG
 GAAGACTGGGACAGCACTCAGGCAGAGGAGGACAGCCTGCAAGACAGTCTGGAGGAAGACGAGGAGGAGGC
 AGAGGAGGAGGTGGAAGAAGCAGCCGCCCGCCAGACCGTTCGTCTCGGCGGGGGAGAAAGCAAGCAGCACGG
 ATACCATCTCCGCTCCGGGTCCGGGTCCCGCTCGACCACACAGTAGATGGGACGAGACCGGACGATTCCCG
 AACCACCACCCAGACCGGTAAGAAGGAGCGGCAGGGATACAAGTCTTGGCGGGGGCACAAAAACGCCAT
 CGTCTCCTGCTTGCAGGCCTGCGGGGGCAACATCTCCTTCACCCGGCGCTACCTGCTCTTCCACCGCGGGG
 TGAACCTTTCCCCGCAACATCTTGCATTACTACCGTCACCTCCACAGCCCCCTACTACTTCCAAGAAGAGGCA
 GCAGCAGCAGAAAAAGACCAGCAGAAAAACCAGCAGCTAGAAAAATCCACAGCGGCGGCAGCAGGTGGACTGA
 GGATCGCGGGCGAACGAGCCGGCGCAAACCCGGGAGCTGAGGAACCGGATCTTTCCACCCCTCTATGCCATC
 TTCCAGCAGAGTCGGGGGCAGGAGCAGGAAGTCAAGAACCGTTCTCTGCGCTCGCTCACCCGCAG
 TTGTCTGTATCACAAGAGCGAAGACCAACTTCAGCGCACTCTCGAGGACGCCGAGGCTCTCTTCAACAAGT
 ACTGCGCGCTCACTCTTAAAGAGTAGCCCGCGCCCGCCAGTCGCAGAAAAAGGCGGGAATTACGTACCT
 GTGCCCCCTCGCCCTAGCCGCCCTCCACCCATCATCATGAGCAAGAGATTCCACGCCTTACATGTGGAGCT
 ACGCCCCAGATGGGCTGGCCGCGGTCGCCGCCAGGACTACTCCACCCGCATGAATTGGCTCAGCGCC
 GGGCCCGCATGATCTCACGGGTGAATGACATCCGCGCCACCAGAAACCAGATACTCTAGAACAGTCAGC
 GCTCACCGCCACGCCCCGCAATCACCTCAATCCGCGTAATTGGCCCGCCGCCCTGGTGTACCAGGAAATTC
 CCCAGCCACGACCGTACTACTTCCGCGAGACGCCCAGGCCGAAGTCCAGCTGACTAACTCAGGTGTCCAG
 CTGGCGGGCGGCGCCACCCTGTGTGCTCACCGCCCCGCTCAGGGTATAAAGCGGCTGGTGATCCGGGGCAG
 AGGCACACAGCTCAACGACGAGGTGGTGAGCTCTTCGCTGGGTCTGCGACCTGACGGAGTCTTCCAACCTCG
 CCGGATCGGGGAGATCTTCCCTTCACGCCTCGTCAGGCCGTCTGACTTTGGAGAGTTCGTCTCTCGCAGCCC
 CGCTCGGGTGGCATCGGCACCTCTCCAGTTTCGTGGAGGAGTTCACTCCCTCGGTCTACTTCAACCCCTTCTC
 CGGCTCCCCCGGCCACTACCCGGACGAGTTTCATCCCGAAGTTTCGACGCCATCAGCGAGTCGGTGGACGGCT
 ACGATTGAAACTAATCACCCCTTATCCAGTGAAATAAAGATCATATTGATGATGATTTTACAGAAATAAA
 AAATAATCATTTGATTTGAAATAAAGATACAATCATATTGATGATTTGAGTTTAAACAAAAAATAAAGAAT
 CACTTACTTGAAATCTGATACCAGGTCTCTGTCCATGTTTTCTGCCAACACCACTTCACTCCCCCTCTTCCC
 AGCTCTGGTACTGCAGGCCCGGCGGGCTGCAAACTTCTCCACACGCTGAAGGGGATGTCAAATTCCTCC
 TGTCCCTCAATCTTCATTTTATCTTCTATCAGATGTCCAAAAAGCGCGTCCGGGTGGATGATGACTTCGAC
 CCCGTCTACCCCTACGATGCAGACAACGCACCGACCGTGCCCTTCATCAACCCCCCTTCGTCTCTTCAGA
 TGGATTCCAAGAGAAGCCCCCTGGGGGTGTTGTCCCTGCGACTGGCCGACCCCGTCACCACCAAGAACGGGG
 AAATCACCCCTCAAGCTGGGAGAGGGGGTGGACCTCGATTCTCTGGGAAAACCTCATCTCCAACACGGCCACC
 AAGGCGCCCGCCCCCTCTCAGTTTTTCCAACAACACCATTTCCCTTAACATGGATCACCCCTTTTACACTAA
 AGATGGAAAATTATCCTTACAAGTTTCTCCACCATTAATATACTGAGAACAAGCATCTTAAACACACTAG
 CTTTAGGTTTTGGATCAGGTTTAGGACTCCGTGGCTCTGCCTTGGCAGTACAGTTAGTCTCTCCACTTACA
 TTTGATACTGATGGAAACATAAAGCTTACCTTAGACAGAGGTTTGCATGTTACAACAGGAGATGCAATTGA
 AAGCAACATAAGCTGGGCTAAAGGTTTTAAATTTGAAGATGGAGCCATAGCAACCAACATTGGAAATGGGT
 TAGAGTTTGAAGCAGTAGTACAGAAACAGGTGTTGATGATGCTTACCCAATCCAAGTTAAACTTGGATCT
 GGCCTTAGCTTTGACAGTACAGGAGCCATAATGGCTGGTAACAAAGAAGACGATAAACTCACTTTGTGGAC
 AACACCTGATCCATCACCAAACTGTCAAATACTCGCAGAAAATGATGCAAACTAACACTTTGCTTGACTA
 AATGTGGTAGTCAAATACTGGCCACTGTGTGAGTCTTAGTTGTAGGAAGTGGAAACCTAAACCCCATTA
 GGCACCGTAAGCAGTGCTCAGGTGTTTCTACGTTTTGATGCAACCGGTGTTCTTTTAAACAGAACATTCTAC

ACTAAAAAATACTGGGGGTATAGGCAGGGAGATAGCATAGATGGCACTCCATATACCAATGCTGTAGGAT
 TCATGCCCCAATTTAAAGCTTATCCAAAGTCACAAAGTTCTACTACTAAAAATAATATAGTAGGGCAAGTA
 TACATGAATGGAGATGTTTTCAAACCTATGCTTCTCACTATAACCCTCAATGGTACTGATGACAGCAACAG
 TACATATTCAATGTCATTTTCATACACCTGGACTAATGGAAGCTATGTTGGAGCAACATTTGGGGCTAACT
 CTTATACCTTCTCATACATCGCCCAAGAATGAACACTGTATCCCACCCTGCATGCCAACCTTCCCACCCC
 ACTCTGTGGAACAAACTCTGAAACACAAAATAAAATAAAGTTCAAGTGTTTTATTGATTCAACAGTTTTAC
 AGGATTTCGAGCAGTTATTTTTCTCCACCCTCCCAGGACATGGAATACACCACCCTCTCCCCCGCACAGC
 CTTGAACATCTGAATGCCATTGGTGATGGACATGCTTTTGGTCTCCACGTTCCACACAGTTTCAGAGCGAG
 CCAGTCTCGGGTCGGTCAGGGAGATGAAACCCTCCGGGCACTCCCGCATCTGCACCTCACAGCTCAACAGC
 TGAGGATTGTCCTCGGTGGTCGGGATCACGGTTATCTGGAAGAAGCAGAAGAGCGGGCGGTGGGAATCATAG
 TCCGCGAACGGGATCGGCCGGTGGTGTCGCATCAGGCCCGCAGCAGTCGCTGCCGCCGCCGCTCCGTCAA
 GCTGCTGCTCAGGGGGTCCGGGTCCAGGGACTCCCTCAGCATGATGCCACGGCCCTCAGCATCAGTCGTC
 TGGTGCGGCGGGCGCAGCGCATGCGGATCTCGCTCAGGTCGCTGCAGTACGTGCAACACAGAACCACC
 AGGTTGTTCAACAGTCCATAGTTCAACACGCTCCAGCCGAAACTCATCGCGGAAGGATGCTACCCACGTG
 GCCGTGCTACCAGATCCTCAGGTAAATCAAGTGGTGCCCCCTCCAGAACACGCTGCCACGTACATGATCT
 CTTTGGGCATGTGGCGGTTACCACCTCCCGGTACCACATCACCTCTGGTTGAACATGCAGCCCCGGATG
 ATCCTGCGGAACCACAGGGCCAGCACCGCCCCGCCGATGCAGCGAAGAGACCCCGGGTCCCGGCAATG
 GCAATGGAGGACCCACCGCTCGTACCCGTGGATCATCTGGGAGCTGAACAAGTCTATGTTGGCACAGCACA
 GGCATATGCTCATGCATCTCTTCAGCACTCTCAACTCCTCGGGGGTCAAAACCATATCCCAGGGCACGGGG
 AACTCTTGCAAGACAGCGAACCCCGCAGAACAGGGCAATCCTCGCACAGAACTTACATTGTGCATGGACAG
 GGTATCGCAATCAGGCAGCACCGGGTGATCCTCCACCAGAGAAGCGCGGGTCTCGGTCTCCTCACAGCGTG
 GTAAGGGGGCCGGCCGATACGGGTGATGGCGGGACGCGGCTGATCGTGTTTCGCGACCGTGTCATGATGCAG
 TTGCTTTTCGACATTTTCGTACTTGCTGTAGCAGAACCTGGTCCGGGCGCTGCACACCGATCGCCGGCGGC
 GGTCTCGGCGCTTGAACGCTCGGTGTTGAAATTGTAAAACAGCCACTCTCTCAGACCGTGTCAGCAGATCT
 AGGGCCTCAGGAGTGATGAAGATCCCATCATGCCTGATGGCTCTGATCACATCGACCACCGTGGAATGGGC
 CAGACCCAGCCAGATGATGCAATTTTGTGGGTTTCGGTGACGGCGGGGGAGGGAAGAACAGGAAGAACCA
 TGATTAACCTTTTAATCCAAACGGTCTCGGAGTACTTCAAAATGAAGATCGCGGAGATGGCACCTCTCGCCC
 CCGCTGTGTTGGTGGAAAATAACAGCCAGGTCAAAGGTGATACGGTTCTCGAGATGTTCCACGGTGGCTTC
 CAGCAAAGCCTCCACGCGCACATCCAGAAACAAGACAATAGCGAAAGCGGGAGGGTTCTCTAATTCTCTCAA
 TCATCATGTTTACACTCCTGCACCATCCCCAGATAATTTTTCATTTTCCAGCCTTGAATGATTTCGAAC TAGT
 TCcTGAGGTAAATCCAAGCCAGCCATGATAAAGAGCTCGCGCAGAGCGCCCTCCACCGCATTTCTTAAGCA
 CACCCTCATAATTCCAAGATATTCTGCTCCTGGTTCACTGCGCAGCAGATTGACAAGCGGAATATCAAAATC
 TCTGCCGCGATCCCTGAGCTCCTCCCTCAGCAATAACTGTAAGTACTCTTTTCATATCCTCTCCGAAATTTT
 TAGCCATAGGACCACCAGGAATAAGATTAGGGCAAGCCACAGTACAGATAAACCGAAGTCTCCCCAGTGA
 GCATTGCCAAATGCAAGACTGCTATAAGCATGCTGGCTAGACCCGGTGATATCTTCCAGATAACTGGACAG
 AAAATCGCCCAGGCAATTTTAAAGAAAATCAACAAAAGAAAAATCCTCCAGGTGGACGTTTAGAGCCTCGG
 GAACAACGATGAAGTAAATGCAAGCGGTGCGTTCCAGCATGGTTAGTTAGCTGATCTGTAGAAAAAACAAA
 AATGAACATTAAACCATGCTAGCCTGGCGAACAGGTGGGTAAATCGTTCTCTCCAGCACCCAGGCAGGCCAC
 GGGGTCTCCGGCGCGACCCTCGTAAAAATTGTGCTATGATTGAAAACCATCACAGAGAGACGTTCCCGGT
 GGCCGGCGTGATGATTGACAAGATGAATACACCCCGGAACATTGGCGTCCGCGAGTGAAAAAAGCGC
 CCGAGGAAGCAATAAGGCACTACAATGCTCAGTCTCAAGTCCAGCAAAGCGATGCCATGCGGATGAAGCAC
 AAAATTCTCAGGTGCGTACAAAATGTAATTACTCCCTCCTGCACAGGCAGCAAAGCCCCCGATCCCTCCA
 GGTACACATACAAAGCCTCAGCGTCCATAGCTTACCGAGCAGCAGCACACAACAGGCGCAAGAGTCAGAGA
 AAGGCTGAGCTCTAACCTGTCCACCCGCTCTCTGCTCAATATATAGCCCAGATCTACACTGACGTAAAGGC
 CAAAGTCTAAAAATACCCGCCAAATAATCACACACGCCCAGCACACGCCCAGAAACCGGTGACACACTCAA
 AAAATACGCGCACTTCCCTCAAACGCCCAAACTGCCGTCAATTTCCGGGTTCACACGCTACGTCATCAAAA
 CACGACTTTCAAATTCCGTCGACCGTTAAAAACGTACCCGCCCGCCCTAACGGTCGCCCCGTCTCTCAG
 CCAATCAGCGCCCCGCATCCCCAAATTCAAACACCTCATTTGCATATTAACGCGCACAAAAAGTTTGAGG

Virus de la encefalitis equina venezolana [VEE] (SEQ ID NO:3) GenBank:
 L01442.2

atgggcggcg	catgagagaa	gccagacca	attacctacc	caaaatggag	aaagtccacg
ttgacatcga	ggaagacagc	ccattcctca	gagctttgca	gcgagagctt	ccgcagtttg
aggtagaagc	caagcaggtc	actgataatg	accatgctaa	tgccagagcg	ttttcgcatac
tggcttcaaa	actgatcgaa	acggaggtgg	acccatccga	cacgatcctt	gacattggaa
gtgcgccgcg	ccgcagaatg	tattctaagc	acaagtatca	ttgtatctgt	ccgatgagat
gtgcggaaga	tccggacaga	ttgtataagt	atgcaactaa	gctgaagaaa	aactgtaagg
aaataactga	taaggaattg	gacaagaaaa	tgaaggagct	cgccgccgtc	atgagcgacc

ctgacctgga	aactgagact	atgtgcctcc	acgacgacga	gtcgtgtcgc	tacgaagggc
aagtgcgtgt	ttaccaggat	gtatacgcg	ttgacggacc	gacaagtctc	tatcaccaag
ccaataaggg	agtttagagtc	gcctactgga	taggctttga	caccacccct	tttatgttta
agaacttggc	tggagcatat	ccatcatact	ctaccaactg	ggccgacgaa	accgtgttaa
cggctcgtaa	cataggccta	tgcagctctg	acgttatgga	gcggtcacgt	agagggatgt
ccattcttag	aaagaagtat	ttgaaacat	ccaacaatgt	tctattctct	gttggctcga
ccatctacca	cgagaagagg	gacttactga	ggagctggca	cctgccgtct	gtatttctact
tacgtggcaa	gcaaaattac	acatgtcgg	gtgagactat	agtttagttgc	gacgggtacg
tcgttaaaaag	aatagctatc	agtccaggcc	tgtatgggaa	gccttcaggc	tatgctgcta
cgatgcaccg	cgagggattc	ttgtgctgca	aagtgcacaga	cacattgaac	ggggagaggg
tctcttttcc	cgtgtgcacg	tatgtgccag	ctacattgtg	tgaccaaatg	actggcatac
tggcaacaga	tgtcagtgcg	gacgacgcgc	aaaaactgct	ggttgggctc	aaccagcgta
tagtcgtcaa	cggctgcacc	cagagaaaca	ccaataccat	gaaaaattac	cttttgcccg
tagtggccca	ggcatttgct	aggtgggcaa	aggaatataa	ggaagatcaa	gaagatgaaa
ggccactagg	actacgagat	agacagttag	tcattggggtg	ttgttgggct	tttagaaggc
acaagataac	atctattttat	aagcgcccg	atacccaaac	catcatcaaa	gtgaacagcg
atttccactc	attcgtgctg	cccaggatag	gcagtaacac	attggagatc	gggctgagaa
caagaatcag	gaaaatgtta	gaggagcaca	aggagccgtc	acctctcatt	accgccgagg
acgtacaaga	agctaagtgc	gcagccgatg	aggctaagga	ggtgcgtgaa	gccgaggagt
tgcgcgcagc	tctaccacct	ttggcagctg	atggttgagga	gcccactctg	gaagccgatg
tcgacttgat	gttacaagag	gctggggccg	gctcagtggg	gacacctcgt	ggcttgataa
aggttaccag	ctacgctggc	gaggacaaga	tcggctctta	cgctgtgctt	tctccgcagg
ctgtactcaa	gagtgaaaaa	ttatcttgca	tccacctct	cgctgaacaa	gtcatagtga
taacacactc	tggccgaaaa	gggcgttatg	ccgtggaacc	ataccatggt	aaagtagtgg
tgccagaggg	acatgcaata	cccgtccagg	actttcaagc	tctgagtga	agtgccacca
ttgtgtacaa	cgaacgtgag	ttcgtaaaaca	ggtacctgca	ccatattgcc	acacatggag
gagcgtgtaa	cactgatgaa	gaatattaca	aaactgtcaa	gcccagcgag	cacgacggcg
aataacctgta	cgacatcgac	aggaaaacagt	gcgtcaagaa	agaactagtc	actgggctag
ggctcacagg	cgagctggtg	gacccctcct	tccatgaatt	cgctacgag	agtctgagaa
cacgaccagc	cgctccttac	caagtaccaa	ccataggggt	gtatggcgtg	ccaggatcag
gcaagtctgg	catcattaaa	agcgcagtc	ccaaaaaaga	tctagtgggtg	agcgccaaga
aagaaaactg	tgcagaaatt	ataagggacg	tcaagaaaat	gaaagggctg	gacgtcaatg
ccagaactgt	ggactcagtg	ctcttgaatg	gatgcaaaaca	ccccgtagag	accctgtata
ttgacgaagc	ttttgcttgt	catgcaggta	ctctcagagc	gctcatagcc	attataagac
ctaaaaaggc	agtgtctctgc	ggggatccca	aacagtgcgg	tttttttaac	atgatgtgcc
tgaagtgc	ttttaaccac	gagatttgca	cacaagtctt	ccacaaaagc	atctctcgcc
gttgactaa	atctgtgact	tcggctcgtct	caaccttggt	ttacgacaaa	aaaatgagaa
cgacgaatcc	gaaagagact	aagatttgtga	ttgacactac	cggcagtacc	aaacctaaagc
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggct	gacccgtaaa	ggtgtgtatg
ccgttcggta	caaggtgaat	gaaaatcctc	tgtacgcacc	cacctcagaa	catgtgaacg
tcctactgac	ccgcacggag	gaccgcacgc	tgtggaaaac	actagccggc	gacccatgga
taaaaaact	gactgccaag	taccctggga	atttcactgc	cacgatagag	gagtggaag
cagagcatga	tgccatcatg	aggcacatct	tggagagacc	ggaccctacc	gacgtcttcc
agaataaggc	aaacgtgtgt	tgggccaagg	ctttagtgc	ggtgctgaag	accgctggca
tagacatgac	cactgaacaa	tggaaactgc	tggattattt	tgaacaggac	aaagctcact
cagcagagat	agtattgaac	caactatgcg	tgaggttctt	tggactcgat	ctggactccg
gtctattttc	tgcacccact	gttccgttat	ccattaggaa	taatcactgg	gataactccc
cgtcgccata	ctgttacggg	ctgaataaag	aagtggtccg	tcagctctct	cgcaggtacc
cacaactgcc	tcgggcagtt	gccactggaa	gagctctatga	catgaacact	ggtacactgc
gcaattatga	tccgcgcata	aacctagtag	ctgtaaacag	aagactgcct	catgctttag
tcctccacca	taatgaacac	ccacagagtg	acttttcttc	attcgtcagc	aaattgaagg
gcagaactgt	cctgggtggc	ggggaaaagt	tgtccgtccc	aggcaaaatg	gttgactggg
tgtcagaccg	gcctgaggct	accttcagag	ctcggctgga	tttaggcata	ccaggtgatg
tgcccaaata	tgacataata	tttgttaatg	tgaggacccc	atataaatac	catcactatc
agcagtgtga	agaccatgcc	attaagctta	gcatgttgac	caagaaagct	tgtctgcata
tgaatcccg	cggaaacctgt	gtcagcatag	gttatgggtta	cgctgacagg	gccagcgaaa

gcatcattgg	tgtctatagcg	cggcagttca	agttttcccg	ggtatgcaaa	ccgaaatcct
cacttgaaga	gacggaagtt	ctgtttgtat	tcattgggta	cgatcgcaag	gcccgtacgc
acaatcctta	caagctttca	tcaaccttga	ccaacattta	tacaggttcc	agactccacg
aagccggatg	tgcaccctca	tatcatgtgg	tgcgagggga	tattgccacg	gccaccgaag
gagtgattat	aaatgctgct	aacagcaaa	gacaacctgg	cggaggggtg	tgcggagcgc
tgtataagaa	attcccggaa	agcttcgatt	tacagccgat	cgaagtagga	aaagcgcgac
tgggtcaaagg	tgcagctaaa	catatcattc	atgccgtagg	accaaacttc	aacaaagttt
cggaggttga	aggtgacaaa	cagttggcag	aggcttatga	gtccatcgct	aagattgtca
acgataacaa	ttacaagtca	gtagcgattc	cactgttgtc	caccggcacc	ttttccggga
acaaagatcg	actaacccaa	tcattgaacc	atltgtgtac	agcttttagac	accactgatg
cagatgtagc	catatactgc	agggacaaga	aatgggaaat	gactctcaag	gaagcagtg
ctaggagaga	agcagtgagg	gagatatgca	tatccgacga	ctcttcagt	acagaacctg
atgcagagct	ggtgaggggtg	catccgaaga	gttctttggc	tggaaaggaag	ggctacagca
caagcgatgg	caaaactttc	tcataatttg	aagggaacca	gtttcaccag	gcggccaaag
atatagcaga	aattaatgcc	atgtggcccg	ttgcaacgga	ggccaatgag	caggtatgca
tgtatatcct	cggagaaaagc	atgagcagta	ttaggtcgaa	atgccccgtc	gaagagtcgg
aagcctccac	accacctagc	acgtgcctt	gcttgtgcat	ccatgccatg	actccagaaa
gagtagacgc	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	ttgtgtgcag	atggcccgac	ccaccaggtg	ctgcaagtgc
aggcagacat	tcacgggccc	ccctctgtat	ctagctcatc	ctgggtccatt	cctcatgcat
ccgactttga	tgtggacagt	ttatccatac	ttgacaccct	ggagggagct	agcgtgacca
gcggggcaac	gtcagccgag	actaactcct	acttcgcaaa	gagtatggag	tttctggcgc
gaccggtgcc	tgcgcctcga	acagtattca	ggaaccctcc	acatcccgtc	ccgcgcacaa
gaacaccgtc	acttgcaccc	agcagggcct	gctcgagaac	cagcctagtt	tcaccccgc
caggcgtgaa	taggggtgatc	actagagagg	agctcgaggc	gcttaccocg	tcacgcactc
ctagcaggtc	ggtctcgcaga	accagcctgg	tctccaacct	gccaggcgta	aataggggtga
ttacaagaga	ggagtttgag	gcgttcgtag	cacaacaaca	atgacggttt	gatgcgggtg
catacatctt	ttcctccgac	accggccaag	ggcattttaca	acaaaaatca	gtaaggcaaa
cgggtgctatc	cgaagtgggtg	ttggagagga	ccgaattgga	gatttcgtat	gccccgcgcc
tcgaccaaga	aaaagaagaa	ttactacgca	agaaattaca	gttaaattccc	acacctgcta
acagaagcag	ataccagttc	aggaaggtgg	agaacatgaa	agccataaca	gctagacgta
ttctgcaagg	cctagggcat	tattttgaagg	cagaaggaaa	agtggagtgc	taccgaacct
tgcacctctg	tcctttgtat	tcacttagtg	tgaaccgtgc	cttttcaagc	cccaaggctc
cagtggaaagc	ctgtaacgcc	atgttgaaag	agaactttcc	gactgtggct	tcttactgta
ttattccaga	gtacgatgcc	tattttggaca	tggttgacgg	agcttcatgc	tgcttagaca
ctgccagttt	ttgccctgca	aagctgcgca	gctttccaaa	gaaacactcc	tattttggaac
ccacaatacg	atcggcagtg	ccttcagcga	tccagaacac	gctccagaac	gtcctggcag
ctgccacaaa	aagaaattgc	aatgtcacgc	aatgagaga	attgcccgtg	ttggattcgg
cggccttttaa	tgtggaatgc	ttcaagaaat	atgctgtgaa	taatgaatat	tgggaaacgt
ttaaagaaaa	ccccatcagg	cttactgaag	aaaacgtgg	aaattacatt	accaaattaa
aaggaccaaa	agctgctgct	ctttttgcga	agacacataa	tttgaatatg	ttgcaggaca
taccaatgga	caggtttgta	atggacttaa	agagagacgt	gaaagtgact	ccaggaacaa
aacatactga	agaacggccc	aaggtacagg	tgatccaggc	tgccgatccg	ctagcaacag
cgtatctgtg	cggaaatccac	cgagagctgg	ttaggagatt	aaatgcggtc	ctgcttccga
acattcatac	actgtttgat	atgtcggctg	aagactttga	cgctattata	gccgagcact
tccagcctgg	ggattgtggt	ctggaaactg	acatcgcgtc	gtttgataaa	agtgaggacg
acgccatggc	cttgaccgcg	ttaatgattc	tggaaagact	aggtgtggac	gcagagctgt
tgacgctgat	tgaggcggct	ttcggcgaaa	tttcatcaat	acatttgccc	actaaaacta
aattttaaatt	cggagccatg	atgaaatctg	gaatgttcc	cacactgttt	gtgaacacag
tcattaacat	tgtaatcgca	agcagagtgt	tgagagaacg	gctaaccgga	tcaccatgtg
cagcattcat	tggagatgac	aatatcgtga	aaggagtcaa	atcggacaaa	ttaatggcag
acaggtgcgc	cacctgggtg	aatatggaag	tcaagattat	agatgctgtg	gtgggcgaga
aagcgcctta	tttctgtgga	gggtttat	tgtgtgactc	cgtgaccggc	acagcgtgcc

gtgtggcaga	ccccctaaaa	aggctgttta	agcttggcaa	acctctggca	gcagacgatg
aacatgatga	tgacaggaga	agggcattgc	atgaagagtc	aacacgctgg	aaccgagtgg
gtattctttc	agagctgtgc	aaggcagtag	aatcaaggta	tgaaaccgta	ggaacttcca
tcatagttaa	ggccatgact	actctagcta	gcagtgttaa	atcattcagc	tacctgagag
gggcccctat	aactctctac	ggctaacctg	aatggactac	gacatagtct	agtccgccaa
gatgttcccc	ttccagccaa	tgtatccgat	gcagccaatg	ccctatcgca	acccgttcgc
ggccccgcgc	aggccctggt	tccccagAAC	cgaccctttt	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	atgcgggccc	atacattcaa	atacaccat	gagaaacccc	aaggctatta
cagctggcat	catggagcag	tccaatatga	aaatgggctg	ttcacggtgc	cgaaaggagt
tggggccaag	ggagacagcg	gacgacccat	tctggataac	cagggacggg	tggctcgctat
tgtgctggga	ggtgtgaatg	aaggatctag	gacagccctt	tcagtctgca	tgtggaacga
gaagggagtt	accgtgaagt	atactccgga	gaactgcgag	caatgggtcac	tagtgaccac
catgtgtctg	ctcgccaatg	tgacgttccc	atgtgctcaa	ccaccaattt	gctacgacag
aaaaccagca	gagacttttg	ccatgctcag	cgttaacggt	gacaacccgg	gctacgatga
gctgctggaa	gcagctgtta	agtgcgcccg	aagggaaaagg	agatccaccg	aggagctggt
taaggagtat	aagctaacgc	gcccttacat	ggccagatgc	atcagatgtg	cagttgggag
ctgccatagt	ccaatagcaa	tcgaggcagt	aaagagcgac	gggcacgacg	gttatgttag
acttcagact	tcctcgcagt	atggcctgga	ttcctccggc	aacttaaagg	gcaggaccat
gcggtatgac	atgcacggga	ccattaaaga	gataccacta	catcaagtgt	cactccatac
atctcgcccc	tgtcacattg	tggatgggca	cggttatttt	ctgcttgcca	ggtgcccggc
aggggactcc	atcaccatgg	aatttaagaa	agattccgtc	acacactcct	gctcgggtgc
gtatgaagtg	aaatttaatc	ctgtaggcag	agaactctat	actcatcccc	cagaacacgg
agtagagcaa	gcgtgccaag	tctacgcaca	tgatgcacag	aacagaggag	cttatgtcga
gatgcacctc	ccgggctcag	aagtggacag	cagtttgggt	tccttgagcg	gcagttcagt
caccgtgaca	ccctctgttg	ggactagcgc	cctgggtggaa	tgcgagtgtg	gcggcacaaa
gatctccgag	accatcaaca	agacaaaaca	gttcagccag	tgacaaaaga	aggagcagtg
cagagcatat	cggctgcaga	acgataagtg	ggtgtataat	tctgacaaac	tgcccaaagc
agcgggagcc	accttaaaaag	gaaaactgca	tgtcccattc	ttgctggcag	acggcaaatg
caccgtgcct	ctagcaccag	aacctatgat	aacctttggg	ttcagatcag	tgtcactgaa
actgcaccct	aagaatccca	catatctaac	caccgcgcaa	cttgctgatg	agcctcacta
cacgcacgag	ctcatatctg	aaccagctgt	taggaatttt	accgtcaccg	aaaaaggggtg
ggagtttgta	tggggaaacc	acccgccgaa	aaggtttttg	gcacaggaaa	cagcaccggtg
aaatccacat	gggctaccgc	acgaggtgat	aactcattat	taccacagat	accctatgtc
caccatcctg	ggtttgtaaa	tttgtgccgc	cattgcaacc	gtttccggtg	cagcgtctac
ctggctgttt	tgcagatcta	gagttgcgtg	cctaactcct	taccggctaa	cacctaacgc
taggatacca	ttttgtcttg	ctgtgctttg	ctgcgcccgc	actgcccggg	ccgagaccac
ctgggagtc	ttggatcacc	tatggaacaa	taaccaacag	atgttctgga	ttcaattgct
gateccctctg	gcgccttgta	tcgtagtga	tcgcctgctc	aggtgcgtgt	gctgtgtcgt
gccttttttta	gtcatggccg	gcgcgcgagg	cgccggcgcc	tacgagcacg	cgaccacgat
gccgagccaa	gcgggaatct	cgtataacac	tatagtcaac	agagcaggct	acgcaccact
ccctatcagc	ataacaccaa	caaagatcaa	gctgatacct	acagtgaact	tggagtacgt
cacctgccac	tacaaaacag	gaatggattc	accagccatc	aaatgctgcg	gatctcagga
atgcactcca	acttacaggc	ctgatgaaca	gtgcaaaagtc	ttcacagggg	tttaccggtt
catgtggggt	ggtgcatatt	gcttttgcca	cactgagaac	acccaagtca	gcaaggccta
cgtaatgaaa	tctgacgact	gccttgcgga	tcagtctgaa	gcatataaaag	cgcacacagc
ctcagtgcag	gcgttcctca	acatcacagt	gggagaacac	tctattgtga	ctaccgtgta
tgtgaatgga	gaaactcctg	tgaatttcaa	tggggtcaaa	ttaaactgcag	gtccgctttc
cacagcttgg	acaccctttg	atcgcaaaat	cgtgcagtat	gccggggaga	tctataatta
tgattttcct	gagtatgggg	caggacaacc	aggagcattt	ggagatatac	aatccagaac
agtctcaagc	tcagatctgt	atgccaatat	caacctagt	ctgcagagac	ccaaagcagg

agcgatccac	gtgccataca	ctcaggcacc	ttcgggtttt	gagcaatgga	agaaagataa
agctccatca	ttgaaattta	ccgccccctt	cggatgcgaa	atatatacaa	accccatctg
cgccgaaaac	tgtgctgtag	ggtcaattcc	attagccttt	gacattcccg	acgccttggt
caccagggtg	tcagaaacac	cgacactttc	agcggccgaa	tgcactctta	acgagtgcgt
gtattcttcc	gactttgggtg	ggatcgccac	ggtcaagtac	tcggccagca	agtcaggcaa
gtgcgcagtc	catgtgccat	cagggactgc	taccctaaaa	gaagcagcag	tcgagctaac
cgagcaaggg	tcggcgacta	tccattttct	gaccgcaa	atccaccggg	agttcagggt
ccaaatatgc	acatcatatg	ttacgtgcaa	aggtgattgt	cacccccgga	aagaccatat
tgtgacacac	cctcagtatc	acgcccacac	atttacagcc	gcgggtgtcaa	aaaccgcgtg
gacgtgggta	acatccctgc	tgggaggatc	agccgtaatt	attataattg	gcttgggtgct
ggctactatt	gtggccatgt	acgtgctgac	caaccagaaa	cataattgaa	tacagcagca
attggcaagc	tgcttacata	gaactcgccg	cgattggcat	gccgccttaa	aatttttatt
ttattttttc	ttttcttttc	cgaatcggat	tttgttttta	atatttc	
VEE-MAG25mer (SEQ ID NO:4); contiene el nucleótido de MAG-25merPDTT (bases 30-1755)					
atggggcggcgcatgagagaagccagaccaattacctacccaaaatggagaaagttcacgttgac atcgaggaagacagcccatttcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagca ggtcactgataatgaccatgctaattgccagagcgttttcgcatctggcttcaaaactgatcgaaacggagg tggacccatccgacacgatccttgacattggaagtgcgccgcgcgcagaaatgtattctaaagcacaagtat cattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaa aaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgcgctcatgagcgacc ctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctggt taccaggatgtatacgcggttgacggaccgacaagtctctatcaccaagccaataagggagttagagtcgc ctactggataggctttgacaccaccccttttatgtttaagaacttggctggagcatatccatcatactcta ccaactgggcgcgacgaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcgg tcacgtagagggatgtccattcttagaaagaagtatgtgaaacccatccaacaatgttctattctctgttgg ctcgaccatctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcaacttacgtg gcaagcaaaattacacatgtcgggtgtgagactatagtttagttgcgacgggtacgtcggttaaaagaatagct atcagtcacaggcctgtatgggaagccttcaggctatgtcgtacgatgcaccgcgaggggattcttctgtctg caaagtgcagacacattgaacggggagaggggtctcttttcccgctgtgcacgtatgtgccagctacattgt gtgaccaaattgactggcactactggcaacagatgtcagtgccgacgacgcgcaaaaactgctgggttgggctc aaccagcgtatagtcgtcaacggctcgacccagagaaacaccaataccatgaaaaattaccttttgcgcgt agtggccaggcatttgcctaggtgggcaaggaatataaggaagatcaagaagatgaaagggccactaggac tacgagatagacagtttagtcattgggtgttgggtcttttagaaggcacaagataacatctatttataag cgcccgatacccaaacccatcatcaaagtgaacagcgatttccactcattcgtgctgccaggataggcag taacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctc tcattaccgcgcgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggagggtgcgtgaagccgag gagttgcgcgcagctctaccacctttggcagctgatgttgaggagccactctggaagccgatgtcgactt gatgttacaagaggctggggccggctcagtgagacacctcgtggcttgataaaggttaccagctacgtcg gcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaaattatcttgc atccacccctctcgctgaacaagtcatagtgataacacactctggccgaaaagggcggttatgccgtggaacc ataccatggtaagtagtggtgccagagggacatgcaataaccgctccaggactttcaagctctgagtgaaa gtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatggagga gcgctgaacactgatgaagaatattacaaaactgtcaagccagcgagcagcagcggaataacctgtacga catcgacaggaaacagtgctcaagaaagaactagtcactgggctagggtcacaggcgagctgggtggatc ctcccttccatgaattcgctacgagagctctgagaacacgaccagccgctccttaccaagtaaccaaccata ggggtgtatggcgtgccaggatcaggcaagctctggcatcattaaaagcgcagtcaccaaaaaagatctagt ggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaaagggctggacgtca atgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagacctgtatattgacgaa gcttttgccttgatgcaggtactctcagagcgtcatagccattataagacctaaaaagggcagtgctctg cggggatcccaaacagtgcggttttttttaacatgatgtgcctgaaagtgcatttttaaccacagagatttgc caaagtccttccaaaaagcatctctcgccggttgactaaatctgtgacttcggctcgtctcaaccttgtt tacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaa acctaaagcaggacgatctcattctcacttgtttcagaggggtgggtgaagcagttgcaaatagattacaaag gcaacgaaataatgacggcagctgcctctcaagggctgaccgtaaaaggtgtgtatgccgttccgttacaag gtgaatgaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgaccgcacggaggaccg					

catcgtgtggaaaacactagccggcgacccatggataaaaaacactgactgccaagtaccctgggaatttca
 ctgccacgatagaggagtggcaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccct
 accgacgtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccggtgctgaagaccgctgg
 catagacatgaccactgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagaga
 tagtattgaaccaactatgcgtgaggttcttggactcgatctggactccggtctattttctgcaccact
 gttccgttatccattaggaataatcactgggataactccccgtcgcctaactgtacgggtgaataaaga
 agtgggtccgtcagctctctcgcaggtaccacaaactgcctcgggcagttgccactggaagagtctatgaca
 tgaacactggtacactgcgcaattatgatccgcgcataaacctagtagcctgtaaacagaagactgcctcat
 gcttttagtccctccaccataatgaacacccacagagtgaacttttcttcattcgtcagcaaatggaagggcag
 aactgtcctggtggtcggggaaaagtgtcctgccagggcaaatggttgactggttgctcagaccggcctg
 aggtaccttcagagctcggctggatttaggcacatcccaggtgatgtgcccaaatatgacataatatttgtt
 aatgtgaggaccccatataaataccatcactatcagcagtgatgaagaccatgccattaagcttagcatggt
 gaccaagaaagcttctgtcgtcgtgaatccccggcggaacctgtgtcagcatagggttatggttacgtgaca
 gggccagcgaagacatcattggtgctatagcgcggcagttcaagttttccgggtatgcaaacggaaatcc
 tcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtagcgcacaatcctta
 caagctttcatcaaccttgaccaacattttatacaggtttccagactccacgaagccggatgtgcacctcat
 atcatgtggtgcgaggggatattgccacggccaccgaaggagtattataaatgctgctaacagcaaaagga
 caacctggcgagggggtgtgcggagcgtgtataagaaattcccggaagcttcgatttacagccgatcga
 agtaggaaaagcgcgactggtcaaagggtgcagctaaacatatcattcatgccgtaggaccaaacttcaaca
 aagtttcggaggttgaaggtgacaaacagttggcagaggttatgagtccatcgctaagattgtcaacgat
 aacaattacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaac
 ccaatcattgaaccatttgcgtgacagcttttagacaccactgatgcagatgtagccatatactgcagggaca
 agaaatgggaaatgactctcaaggaagcagtggttaggagagaagcagtgaggagatatgcataatccgac
 gactcttcagtgacagaacctgatgcagagctggtgaggggtgcacccgaagagttctttggtggaaggaa
 gggctacagcacaagcgatggcaaaactttctcatatttggaaagggaaccaagtttaccaggcgggccaagg
 atatagcagaaattaatgccatgtggcccgttgcaacggaggccaatgagcaggtatgcattatctctc
 ggagaaagcatgagcagttattaggtcgaaatgccccgtcgaaagagtcggaagcctccacaccacctagcac
 gctgccttgcttgcacatccatgccatgactccagaaagagtacagcgcctaaaagcctcacgtccagaac
 aaattactgtgtgctcactcttccattgcccgaagtatagaatcactggtgtgcagaagatccaatgctcc
 cagcctatattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaaacaccacc
 ggtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccactta
 taaccgaggtgagaccaggactagaacgcctgagcgcgatcatcgaagaggaagaagaggatagcata
 agtttgcgtgcagatggcccgcaccaccaggtgctgcaagtcgaggcagacattcacgggcgcctctgt
 atctagctcactcctggtccattcctcatgcacccgactttgatgtggacagtttatccatacttgacaccc
 tggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtatggag
 tttctggcgcgaccgggtgcctgcgcctcgaaacagtattcaggaacctccacatcccgcctccgcgcacaag
 aacaccgtcacttgcacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcggtgaata
 ggggtgatcactagagaggagctcgaggcgcttaccctgcacgcactcctagcaggtcggtctcgagaacc
 agcctggtctccaacccgcccaggcgtaaatagggtgattacaagagaggagtgttgaggcggttcgtagcaca
 acaacaatgacgggttgatgcgggtgcatacatcttttctccgacaccgggtcaagggcatttacaacaaa
 aatcagtaaggcaaacgggtgctatccgaagtgtgttggagaggaccgaattggagatttcgtatgccccg
 cgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatccacacctgctaacagaag
 cagataccagtcaggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggccctagggc
 attatttgaaggcagaaggaaaagtggagtgtaccgaaccctgcacctctgttctcttctgtattcatctagt
 gtgaaccgtgccttttcaagcccccaaggtcgagtggaagcctgtaacgcocatggttgaaagagaactttcc
 gactgtggcttcttactgtattattccagagtagcatgcctatttggacatggttgacggagcttcatgct
 gcttagacactgccagtttttgccttgcaagctgcgcagctttccaaagaaacactcctatttggaaacc
 acaatacgatcggcagtgcttccagcgatccagaacacgctccagaacgtcctggcagctgccacaaaaag
 aaattgcaatgtcacgcaaatgagagaattgcccgtatttggttgcggcgcccttaattgtggaatgcttca
 agaaatagcgtgtaataatgaatattgggaaacgttttaagaaaaccccatcaggtctactgaagaaaac
 gtggttaattacattaccaaattaaaaggaccaaagctgctgctcttttgcgaagacacataatttgaa
 tatgttgcaggacataccaatggacaggtttgtaatggacttaagagagacgtgaaagtgactccaggaa
 caaacatactgaagaacggcccaaggtacaggtgatccaggctgcgcgatccgctagcaacagcgtatctg
 tgcggaatccaccgagagctggttaggagattaaatgcggctcctgcttccgaacattcatacactgtttga
 tatgtcgggtgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactg
 acatcgcgtcgtttgataaaagtggagcgcagcgcctatggctctgaccgcgttaatgattctggaagactta
 ggtgtggacgcagagctgttgacgctgattgaggcggttttccggcgaaatttcatcaatacatttgcaccac

atgggcggcg	catgagagaa	gccagacca	attacctacc	caaaatggag	aaagttcacg
ttgacatcga	ggaagacagc	ccattcctca	gagctttgca	gcggagcttc	ccgcagtttg
aggtagaagc	caagcaggtc	actgataatg	accatgctaa	tgccagagcg	ttttcgcata
tggcttcaaa	actgatcgaa	acggagggtg	acccatccga	cacgatacct	gacattggaa
gtgcgccgcg	ccgcagaatg	tattctaagc	acaagtatca	ttgtatctgt	ccgatgagat
gtgcggaaga	tccggacaga	ttgtataagt	atgcaactaa	gctgaagaaa	aactgtaagg
aaataactga	taaggaattg	gacaagaaaa	tgaaggagct	cgccgcgcgc	atgagcgacc
ctgacctgga	aactgagact	atgtgcctcc	acgacgacga	gtcgtgtcgc	tacgaagggc
aagtgcgtgt	ttaccaggat	gtatacgcg	ttgacggacc	gacaagtctc	tatcaccaag
ccaataaagg	agttagagtc	gcctactgga	taggctttga	caccacccct	tttatgttta
agaacttggc	tggagcatat	ccatcatact	ctaccaactg	ggccgacgaa	accgtgttaa
cggctcgtaa	cataggccta	tgcagctctg	acgttatgga	gcggtcacgt	agagggatgt
ccattcttag	aaagaagtat	ttgaaacct	ccaacaatgt	tctattctct	gttggctcga
ccatctacca	cgagaagagg	gacttactga	ggagctggca	cctgccgtct	gtatttctact
tacgtgcaaa	gcaaaattac	acatgtctgg	gtgagactat	agttagtgtg	gacgggtacg
tcgttaaaag	aatagctatc	agtcagggcc	tgtatggtaa	gccttcaggc	tatgtgtcta
cgatgcaccg	cgaaggattc	ttgtgctgca	aaqtgacaga	cacattqaac	qgggaagagg

tctcttttcc	cgtgtgcacg	tatgtgccag	ctacattgtg	tgaccaaagt	actggcatac
tggcaacaga	tgtcagtgcg	gacgacgcgc	aaaaactgct	ggttggggctc	aaccagcgta
tagtcgtcaa	cggctgcacc	cagagaaaca	ccaataccat	gaaaaattac	cttttgcccg
tagtggccca	ggcatttgct	aggtgggcaa	aggaatataa	ggaagatcaa	gaagatgaaa
ggccactagg	actacgagat	agacagttag	tcatgggggtg	ttgttgggct	tttagaaggc
acaagataac	atctatttat	aagcgcccg	atacccaaac	catcatcaaa	gtgaacagcg
atttccactc	attcgtgctg	cccaggatag	gcagtaaacac	attggagatc	gggctgagaa
caagaatcag	gaaaatgtta	gaggagcaca	aggagccgtc	acctctcatt	accgccgagg
acgtacaaga	agctaagtgc	gcagccgatg	aggctaagga	ggtgcggtgaa	gccgaggagt
tgcgcgcagc	tctaccacct	ttggcagctg	atggttagga	gcccactctg	gaagccgatg
tcgacttgat	gttacaagag	gctggggccg	gctcagtggg	gacacctcgt	ggcttgataa
aggttaccag	ctacgctggc	gaggacaaga	tcggctctta	cgctgtgctt	tctccgcagg
ctgtactcaa	gagtgaaaaa	ttatcttgca	tccacctct	cgctgaacaa	gtcatagtga
taacacactc	tggccgaaaa	gggcgttatg	ccgtggaacc	ataccatggt	aaagttagtg
tgccagaggg	acatgcaata	cccgccag	actttcaagc	tctgagtga	agtgccacca
ttgtgtacaa	cgaacgtgag	ttcgtaaaca	ggtacctgca	ccatattgcc	acacatggag
gagcgctgaa	cactgatgaa	gaatattaca	aaactgtcaa	gccagcgag	cacgacggcg
aatacctgta	cgacatcgac	aggaaacagt	gcgtaagaa	agaactagtc	actgggctag
ggctcacagg	cgagctggtg	gatectccct	tccatgaatt	cgctacgag	agtctgagaa
cacgaccagc	cgctccttac	caagtaccaa	ccataggggt	gtatggcggtg	ccaggatcag
gcaagtctgg	catcattaaa	agcgcagtc	ccaaaaaaga	tctagtgggtg	agcgccaaga
aagaaaactg	tgcagaaatt	ataagggacg	tcaagaaaat	gaaagggctg	gacgtcaatg
ccagaactgt	ggactcagtg	ctcttgaatg	gatgcaacaa	ccccgtagag	accctgtata
ttgacgaagc	ttttgcttgt	catgcaggta	ctctcagagc	gctcatagcc	attataagac
ctaaaaaggc	agtgtctctgc	ggggatccca	aacagtgcgg	tttttttaac	atgatgtgcc
tgaagtgc	ttttaaccac	gagatttgca	cacaagtctt	ccacaaaagc	atctctcgcc
gttgcaactaa	atctgtgact	tcggctcgtct	caaccttggt	ttacgacaaa	aaaatgagaa
cgacgaatcc	gaaagagact	aagattgtga	ttgacactac	cggcagtacc	aaacctaacg
aggacgatct	cattctcact	tgtttcagag	ggtgggtgaa	gcagttgcaa	atagattaca
aaggcaacga	aataatgacg	gcagctgcct	ctcaagggct	gaccggtaaa	ggtgtgtatg
ccgttcggta	caaggtgaat	gaaaatcctc	tgtacgcacc	cacctcagaa	catgtgaacg
tctactgac	ccgcacggag	gaccgcacgt	tgtggaaaac	actagccggc	gaccctgga
taaaaacact	gactgccaa	tacctggga	atttactgc	cacgatagag	gagtggaag
cagagcatga	tgccatcatg	aggcacatct	tggagagacc	ggaccctacc	gacgtcttcc
agaataaggc	aaacgtgtgt	tgggccaagg	ctttagtgc	ggtgctgaag	accgctggca
tagacatgac	cactgaacaa	tggaaactg	tggattattt	tgaaacggac	aaagctcact
cagcagagat	agtattgaac	caactatgcg	tgaggttctt	tggactcgat	ctggactccg
gtctattttc	tgcacccact	gttccgttat	ccattaggaa	taatcactgg	gataactccc
cgctgcctaa	catgtacggg	ctgaataaag	aagtgggtccg	tcagctctct	cgcagggtacc
cacaactgcc	tcgggcagtt	gccactggaa	gagtcctatga	catgaacact	ggtacactgc
gcaattatga	tccgcgcata	aacctagtac	ctgtaaacag	aagactgcct	catgctttag
tcctccacca	taatgaacac	ccacagagtg	acttttcttc	attcgtcagc	aaattgaagg
gcagaactgt	cctggtggtc	ggggaaaagt	tgtccgtccc	aggcaaaatg	ggtgactggg
tgtcagaccg	gcctgaggct	accttcagag	ctcggtgga	tttaggcatac	ccagggtgatg
tgccccaaata	tgacataata	tttgttaatg	tgaggacccc	atataaatac	catcactatc
agcagtgtga	agaccatgcc	attaagctta	gcatgttgac	caagaaagct	tgtctgcatac
tgaatcccgg	cggaaacctgt	gtcagcatag	gttatgggtta	cgctgacagg	gccagcgaaa
gcatcattgg	tgtatagcg	cggcagttca	agttttcccg	ggtatgcaaa	ccgaaatcct
cacttgaaga	gacggaagtt	ctgtttgtat	tcatgggta	cgatcgcaag	gcccgtacgc
acaatcctta	caagctttca	tcaaccttga	ccaacattta	tacaggttcc	agactccacg
aagccggatg	tgcacctca	tatcatgtgg	tgcgagggga	tattgccacg	gccaccgaag
gagtgattat	aaatgctgct	aacagcaaag	gacaacctgg	cggaggggtg	tgcggagcgc
tgtataagaa	attcccggaa	agcttcgatt	tacagccgat	cgaagtagga	aaagcgcgac
tgggtcaaagg	tgcagctaaa	catatcatte	atgccgtagg	accaaacttc	aacaaagttt
cggagggttga	agggtgacaaa	cagttggcag	aggcttatga	gtccatcgct	aagattgtca
acgataacaa	ttacaagtca	gtagcgattc	cactgttgct	caccggcatc	ttttccggga
acaaagatcg	actaacccaa	tcattgaacc	atttgcgtgac	agcttttagac	accactgatg

cagatgtagc	catatactgc	agggacaaga	aatgggaaat	gactctcaag	gaagcagtgg
ctaggagaga	agcagtggag	gagatatgca	tatccgacga	ctcttcagtg	acagaacctg
atgcagagct	ggtgaggggtg	catccgaaga	gttctttggc	tggaggaag	ggctacagca
caagcgatgg	caaaactttc	tcatatttgg	aagggacca	gtttcaccag	gcggccaaagg
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	tagaatcaact	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	tcacgggccc	ccctctgtat	ctagctcatc	ctgggtccatt	cctcatgcat
ccgactttga	tgtggacagt	ttatccatac	ttgacaccct	ggagggagct	agcgtgacca
gcggggcaac	gtcagccgag	actaactcct	acttcgcaaa	gagtatggag	tttctggcgc
gaccggtgcc	tgcgcctcga	acagtattca	ggaaccctcc	acatcccgtc	ccgcgcacaa
gaacaccgtc	acttgcaccc	agcagggcct	gctcgagAAC	cagcctagtt	tcacccccgc
caggcgtgaa	taggggtgatc	actagagagg	agctcgaggc	gcttaccctg	tcacgcactc
ctagcaggtc	ggtctcagaga	accagcctgg	tctccaaccc	gccaggcgta	aataggggtga
ttacaagaga	ggagtttgag	gcgttcgtag	cacaacaaca	atgacgggtt	gatgcgggtg
catacatcct	ttcctccgac	accggtcaag	ggcatttaca	acaaaaatca	gtaaggcaaa
cgggtgctatc	cgaagtgggtg	ttggagagga	ccgaattgga	gatttcgtat	gccccgcgcc
tcgaccaaga	aaaagaagaa	ttactacgca	agaaattaca	gttaaattccc	acacctgcta
acagaagcag	ataccagtc	aggaaggtgg	agaacatgaa	agccataaca	gctagacgta
ttctgcaagg	cctagggcat	tattttgaagg	cagaaggaaa	agtggagtgc	taccgaaccc
tgcatectgt	tcctttgtat	tcattctagt	tgaaccgtgc	cttttcaage	cccaaggctg
cagtggaaagc	ctgtaacgcc	atggttgaag	agaactttcc	gactgtggct	tcttactgta
ttattccaga	gtacgatgcc	tattttggaca	tgggttgacgg	agcttcatgc	tgcttagaca
ctgccagttt	ttgccctgca	aagctgcgca	gctttccaaa	gaaacactcc	tattttggaac
ccacaatacg	atcggcagtg	ccttcagcga	tccagaacac	gctccagaac	gtcctggcag
ctgccacaaa	aagaaattgc	aatgtcacgc	aaatgagaga	attgcccgta	ttggattcgg
cggcctttta	tgtggaatgc	ttcaagaaat	atgcgtgtaa	taatgaatat	tgggaaacgt
ttaaagaaaa	ccccatcagg	cttactgaag	aaaacgtggg	aaattacatt	accaaattaa
aaggacaaaa	agctgctgct	ctttttgcga	agacacataa	tttgaatatg	ttgcaggaca
taccaatgga	caggtttgta	atggacttaa	agagagacgt	gaaagtgact	ccagggaaca
aacatactga	agaacggccc	aagggtacagg	tgatccaggc	tgccgatccg	ctagcaacag
cgtatctgtg	cggaaatccac	cgagagctgg	ttaggagatt	aaatgcggtc	ctgcttccga
acattcatatc	actgtttgat	atgtcggctg	aagactttga	cgctattata	gccgagcact
tccagcctgg	ggatttgtgt	ctggaaactg	acatcgctgc	gtttgataaa	agtgaggacg
acgccatggc	tctgaccggc	ttaatgattc	tggaaagact	aggtgtggac	gcagagctgt
tgacgctgat	tgaggcggct	ttcggcgaaa	tttcatcaat	acatttgccc	actaaaacta
aattttaaatt	cggagccatg	atgaaatctg	gaatgttcct	cacactgttt	gtgaacacag
tcattaacat	tgtaatcgca	agcagagtgt	tgagagaacg	gctaaccgga	tcaccatgtg
cagcattcat	tggagatgac	aatatcgtga	aaggagtcaa	atcggacaaa	ttaatggcag
acagggtgcgc	cacctgggtg	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	tgaaacgta	ggaacttcca
tcatagttat	ggccatgact	actctagcta	gcagtgttaa	atcattcagc	tacctgagag
gggcccctat	aactctctac	ggctaacctg	aatggactac	gacatagtct	agtccgccaa
gatgtttccg	ttccagccaa	tgtatccgat	gcagccaatg	ccctatcgca	acccgttcgc
ggccccgcgc	aggccctggg	tccccagAAC	cgaccctttt	ctggcgatgc	aggtgcagga
attaaccgcgc	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
ttgtgtgggtc	ggaggggaagt	tattcaggcc	gatgcatgtg	gaaggcaaga	tcgacaacga
cgttctggcc	gcgcttaaga	cgaagaaagc	atccaaatac	gatcttgagt	atgcagatgt
gccacagaac	atgcggggccg	atacattcaa	atacaccat	gagaaacccc	aaggctatta
cagctggcat	catggagcag	tccaatatga	aaatgggctg	ttcacgggtg	cgaaaggagt
tggggccaag	ggagacagcg	gacgacccat	tctggataac	cagggacggg	tggctcgctat
tgtgctggga	ggtgtgaatg	aaggatctag	gacagccctt	tcagtcgtca	tgtggaacga
gaagggaagt	accgtgaagt	atactccgga	gaactgcgag	caatgggtcac	tagtgaccac
catgtgtctg	ctcgccaatg	tgacgttccc	atgtgctcaa	ccaccaat	gctacgacag
aaaaccagca	gagacttttg	ccatgctcag	cgttaacggt	gacaaccggg	gctacgatga
gctgctggaa	gcagctgtta	agtgcgccgg	aaggaaaagg	agatccaccg	aggagctggt
taaggagtat	aagctaacgc	gcccttacat	ggccagatgc	atcagatgtg	cagttgggag
ctgccatagt	ccaatagcaa	tcgaggcagt	aaagagcgac	gggcacgacg	gttatgttag
acttcagact	tcctcgagct	atggcctgga	ttcctccggc	aacttaaagg	gcaggaccat
gcggtatgac	atgcacggga	ccattaaaga	gataccacta	catcaagtgt	cactccatac
atctcgcccg	tgtcacattg	tggatgggca	cggttatttc	ctgcttgcca	ggtgcccggc
aggggactcc	atcaccatgg	aatttaagaa	agattccgtc	acacactcct	gctcgggtgc
gtatgaagtg	aaatttaatc	ctgtaggcag	agaactctat	actcatcccc	cagaacacgg
agtagagcaa	gcgtgccaag	tctacgcaca	tgatgcacag	aacagaggag	cttatgtcga
gatgcacctc	ccgggctcag	aagtggacag	cagtttggtt	tccttgagcg	gcagttcagt
caccgtgaca	cctcctgttg	ggactagcgc	cctgggtggaa	tgcgagtgtg	gcggcacaaa
gatctccgag	accatcaaca	agacaaaaca	gttcagccag	tgcacaaaga	aggagcagtg
cagagcatat	cggctgcaga	acgataagtg	ggtgtataat	tctgacaaac	tgcccaaagc
agcgggagcc	accttaaaaag	gaaaactgca	tgtcccattc	ttgctggcag	acggcaaagt
caccgtgcct	ctagcaccag	aacctatgat	aacctttggt	ttcagatcag	tgtcactgaa
actgcacctc	aagaatccca	catatctaac	caccgcgcaa	cttgcctgat	agcctcacta
cacgcacgag	ctcatatctg	aaccagctgt	taggaat	accgtcaccg	aaaaagggtg
ggagtttgta	tggggaaacc	acccgcgcaa	aaggtttttg	gcacaggaaa	cagcaccggg
aaatccacat	gggctaccgc	acgaggtgat	aactcattat	taccacagat	accctatgtc
caccatcctg	ggtttgtcaa	tttgtgcgcg	cattgcaacc	gtttccgttg	cagcgtctac
ctgggtgttt	tgcagatcta	gagttgcgtg	cctaactcct	taccggctaa	acctaacgc
taggatacca	ttttgtctgg	ctgtgctttg	ctgcgccgcg	actgcccggg	ccgagaccac
ctgggagtc	ttggatcacc	tatggaacaa	taaccaacag	atgttctgga	ttcaattgct
gatccctctg	gcgccttga	tcgtagtga	tcgcctgctc	aggtgcgtgt	gctgtgtcgt
gcctttttta	gtcatggccg	gcgcgcgagg	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	tttaccggtt
catgtggggt	ggtgcatatt	gcttttgoga	cactgagaac	acccaagtca	gcaaggccta
cgtaatgaaa	tctgacgact	gccttgcgga	tcatgctgaa	gcatataaag	cgcacacagc
ctcagtgag	gcgttcctca	acatcacagt	gggagaacac	tctattgtga	ctaccgtgta
tgtgaatgga	gaaactcctg	tgaatttcaa	tggggtcaaa	ttactgcag	gtccgctttc
cacagcttgg	acaccctttg	atcgcaaaat	cgtgcagtat	gcgggggaga	tctataatta
tgattttcct	gagtatgggg	caggacaacc	aggagcattt	ggagatatac	aatccagaac
agtctcaagc	tcagatctgt	atgccaatat	caacctagt	ctgcagagac	ccaaagcagg
agcgatccac	gtgccatata	ctcaggcacc	ttcggtttt	gagcaatgga	agaaagataa
agctccatca	ttgaaattta	ccgccccttt	cggatgcgaa	atatatacaa	accccattcg
cgccgaaaac	tgtgctgtag	ggtcaatttc	attagccttt	gacattcccg	acgccttggt
caccaggggtg	tcagaaacac	cgacactttc	agcggccgaa	tgactctta	acgagtgcgt
gtattctttc	gacttttggtg	ggatcgccac	ggtcaagtac	tcggccagca	agtcaggcaa
gtgcgcagtc	catgtgccat	cagggaactgc	taccctaaaa	gaagcagcag	tcgagctaac
cgagcaaggg	tcggcgacta	tccattttctc	gaccgcaaat	atccacccgg	agttcaggct
ccaaatatgc	acatcatatg	ttacgtgcaa	aggtgattgt	cacccccgga	aagaccatat
tgtgacacac	cctcagtatc	acgcccacaa	atttacagcc	gcgggtgtcaa	aaaccgcgtg
gacgtgggtta	acatccctgc	tgggaggatc	agccgtaatt	attataattg	gcttgggtgct

ggctactatt gtggccatgt acgtgctgac caaccagaaa cataattgaa tacagcagca
attggcaagc tgcttacata gaactcgagg cgattggcat gccgccttaa aatttttatt
ttattttttc ttttcttttc cgaatcggat ttgttttta atatttc

Vector de suministro del VEE (SEQ ID NO:6); genoma del VEE con
nucleótidos 7.544-11.175 eliminados [proteínas estructurales del
alfavirus removidas]

ATGggcgggcgcgatgagagaagcccagaccaattacctacccaaaATGgagaaagttcacgttgacatcgag
gaagacagcccatttcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagcaggtcac
tgataatgaccatgctaattgccagagcgttttcgcactctggcttcaaaactgatcgaaacggaggtggacc
catccgacacgatcccttgacattggaagtgcgcccgcgcgagaatgtattctaagcacaagtatcattgt
atctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactg
taaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgcggtcatgagcgacccctgacc
tggaactgagactatgtgcctccacgacgacgagctcgtgtcgctacgaagggcaagtcgctgtttaccag
gatgtatacgcgggttgacgggaccgacaagtcctatcaccaagccaataagggagtttagagtcgcctactg
gataaggctttgacaccaccccttttatgtttaagaacttggtcggagcatatccatcatactctaccaact
gggcccagcaaaaccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgt
agagggatgtccattcttagaagaagtatattgaaaccatccaacaatgttctattctctgttggctcgac
catctaccacgagaagagggacttactgaggagctggcacctgccgtctgtatttcacttacgtggcaagc
aaaattacacatgtcgggtgtgagactatagtttagttgcgacgggtacgtcgtaaaagaatagctatcagt
ccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgcaaggt
gacagacacattgaacggggagaggggtctcttttcccggtgtgcacgtatgtgccagctacattgtgtgacc
aaatgactggcactactggcaacagatgtcagtgccgacgacgcgcaaaaactgctgggtgggctcaaccag
cgtatagtcgtcaacgggtcgacccagagaaaacaccaataccatgaaaaattaccttttgcgcgtagtggc
ccaggcattttgctaggtgggcaaggaatataaggaagatcaagaagatgaaaggccactaggactacgag
atagacagtttagtcatgggggtgttgttgggcttttagaaggcacaagataacatctatttataagcgccg
gatacccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgccaggataggcagtaaacac
attggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcatta
ccgcgcgaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgagggagttg
cgcgagctctaccacctttggcagctgatgttgaggagccactctggaagccgatgtcgacttgatgtt
acaagaggctggggccggctcagtgaggacacctcgtggcttgataaagggtaccagctacgctggcgagg
acaagatcgggtcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaaattatcttgcacccac
cctctcgctgaacaagtcatagtgataacacactctggccgaaaagggcggttatgccgtggaaccatacca
tggtaaagtagtggtgccagagggacatgcaatacccgctccaggactttcaagctctgagtgaagtgcca
ccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcaccatattgccacacatggaggagcgctg
aacactgatgaagaatattacaaaactgtcaagcccagcgagcagcagcggaataacctgtacgacatcga
caggaaacagtgctcaagaaagaactagtcactgggctagggctcacaggcgagctgggtggatccctccct
tccatgaattcgctacgagagctctgagaacacgaccagccgctccttaccagtagccaaccataggggtg
tatggcggtgccaggatcaggcaagctctggcatcattaaaagcgagtcacccaaaaagatctagtggtag
cgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaaagggctggacgtcaatgcc
gaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagacctgtatattgacgaagctttt
gcttgtcatgcaggactctcagagcgctcatagccattataagacctaaaagggcagtgctctgcgggga
tcccaaacagtgcggtttttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgacacaaag
tcttccacaaaagcatctctcgccggttgactaaatctgtgacttcgggtcgtctcaacctgttttacgac
aaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaa
gcaggacgatctcattctcacttgtttcagaggggtgggtgaagcagttgcaaatagattacaaaggcaacg
aaataatgacggcagctgcctctcaagggtgacccgtaagggtgtgtatgccgttcgggtacaagggtgaat
gaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggacccgcatcgt
gtggaaaacactagccggcgacccatggataaaaacactgactgccaagtaccctgggaatttcactgcca
cgatagaggagtggaagcagagcatgatgccatcatgaggcacatcttggagagaccggacccctaccgac
gtcttccagaataaggcaaacgtgtgttgggccaaggcttttagtgccgggtgctgaagaccgctggcataga
catgaccactgaacaatggaacactgtggattattttgaaacggacaaagctcactcagcagagatagtat
tgaaccaactatgcgtgaggttctttggactcgactcgactccgggtctattttctgcacccactgttccg
ttatccattaggaataatcactgggataactccccgctcgccataacatgtacgggctgaataaagaagtggt
ccgtcagctctctcgaggtaccacaaactgcctcgggcagttgccactggaagagtctatgacatgaaca
ctgggtacactgcgcaattatgatccgcgcataaaacctagtacctgtaaacagaagactgcctcatgcttta
gtcctccaccataatgaacacccacagagtgacttttcttcattcgtcagcaaatgaaaggcgagaactgt
cctggtggctcggggaaaagttgtccgtcccaggcaaaatgggttactgggtgtcagacccggcctgaggcta

ccttcagagctcggttggtatttaggcaccccaggtgatgtgccccaaatatgacataaatatttgttaatgtg
 aggaccccatataaaataccatcactatcagcagtgatgaagaccatgccattaagcttagcatgttgacca
 gaaagcttgtctgcatctgaatcccgccggaacctgtgtcagcataggttatggttacgctgacagggcca
 gcgaaagcatcattggtgctatagcgcggcagttcaagttttcccggtatgcaaaccgaaatcctcactt
 gaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtagcgacaatccttacaagct
 tcatcaaccttgaccaacattttatacaggttcagactccacgaagccggatgtgcaccctcatatcatg
 tgggtgcgaggggatatttgccacggccaccgaaggagtgtattataaatgctgctaacagcaaaggacaacct
 ggcgaggggtgtgctgagcgctgtataagaaattcccggaagcttcgatttacagccgatcgaagtagg
 aaaagcgcgactggtcaaaggtgcagctaaacatatcattcatgcccgtaggaccaaacttcaacaaagttt
 cggaggttgaaggtgacaaacagttggcagaggttatgagtcctatcgctaagattgtcaacgataacaat
 tacaagtcagtagcgattccactgttgtccaccggcatcttttccgggaacaaagatcgactaacccaatc
 attgaaccatttgtgacagcttttagacaccactgatgcagatgtagccatatactgcaggggacaagaat
 gggaaatgactctcaaggaagcagtggttaggagagaagcagtgaggagatatgcataccgacgactct
 tcagtgcagagaacctgatgcagagctggtgaggggtgcacccgaagagttctttggctggaaggaagggcta
 cagcacaagcgatggcaaaactttctcatatttgggaagggaccaagtttcaccaggcgccaaggatatag
 cagaaattaatgccatgtggcccggttgcaacggaggccaatgagcaggtatgcatgtatatcctcggagaa
 agcatgagcagttattaggtcgaaatgccccgtcgaaagtcggaagcctccacaccacctagcagctgcc
 ttgcttgtgcatccatgccatgactccagaaagagtacagcgccctaaaagcctcagctccagaacaaatta
 ctgtgtgctcatcctttccattgcccgaagtatagaatcactggtgtgcagaagatccaatgctcccagcct
 atattgttctcaccgaaagtgcctgcttatattcatccaaggaagtatctcgtggaaacaccaccggtaga
 cgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccg
 aggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttg
 ctgtcagatggcccgaccaccagggtgctgcaagtcgaggcagacattcacgggcccgcctctgtatctag
 ctcatcctggtccattcctcatgcatccgactttgatgtggacagtttatccatacttgacaccctggagg
 gagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtatggagtttctg
 gcgcgaccggtgcctgcgccctcgaaacagtattcaggaaccctccacatcccgcctccgcgcacaagaacacc
 gtcaacttgcaaccagcagggcctgctcgagaaccagcctagtttccaccccgcagggcgtgaataggggtga
 tcaactagagaggagctcgagggcgttaccccgtcagcactccttagcaggtcggtctcgagaaccagcctg
 gtctccaaccgcagggcgttaaataaggggtgattacaagagaggagtttgaggcgttcgtagcacaacaaca
 atgacgggttgatgcggtgcatatcatcttttcccgacaccgggtcaagggcattttacaacaaaaatcag
 taaggcaaacgggtgctatccgaagtgggtgttgagagaggaccgaattggagatttcgtatgccccgcgcctc
 gaccaagaaaaagaagaattactacgcaagaattacagttaaatccacacctgtaacacagagcagata
 ccagtcaggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatt
 tgaaggcagaaggaaggaagtggtgctaccgaaccctgcatcctgttccctttgtattcatctagtgtgaac
 cgtgccttttcaagcccccaaggtcgagtggaagcctgtaacgccatgttgaaagagaactttccgactgt
 ggcttcttactgtattattccagagtagcatgacctatttggacatgggttgacggagcttcatgctgcttag
 acactgccagtttttgccttgcaaagctgcgcagctttccaaagaaacactcctattttggaaccacacaata
 cgatcggcagtgcttccagcgatccagaacacgctccagaacgctcctggcagctgccacaaaaagaaattg
 caatgtcacgcaaatgagagaattgcccgtattggattcggcgccctttaatgtggaatgcttcaagaaat
 atgctgtgaataatgaatattgggaacgctttaagaaaaaccccatcaggcttactgaagaaaaacgtggtta
 aattacattaccaaattaaaaggaccaaaagctgctgctctttttgcaagacacataatttgaatatgtt
 gcaggacataccaatggacaggtttgtaatggacttaagagagacgtgaaagtgactccaggaacaaaac
 atactgaagaacggcccaaggtacaggtgatccaggtgcccgatccgctagcaacagcgtatctgtgcgga
 atccaccgagagctggttaggagattaaatgcggctcctgcttccgaacattcatacactgtttgatatgtc
 ggctgaagactttgacgctattatagccgagcacttccagcctggggattgtgttctggaaactgacatcg
 cgtcgtttgataaaagtgaggacgacgccatggctctgaccgcgttaatgattctggaagacttaggtgtg
 gacgcagagctgttgacgctgattgaggcggctttcggcgaaatttcatcaatacatttggccactaaaac
 taaatttaaattcggagccatgatgaaatctggaatgttctcactgcttctggaacacagtcatttaaca
 ttgtaatcgcaagcagagtggttgagagaacgggttaaccggatcaccatgtgcagcattcattggagatgac
 aatatcgtgaaaggagtc aaatcggaacaaattaatggcagacaggtgcgccacctgggttgaatatggaagt
 caagattatagatgctgtgtggcgagaaagcgcttatttctgtggaggggttattttgtgtgactccg
 tgaccggcacagcgtgcccgtgtggcagacccccctaaaaaggctgtttaagcttggcaaacctctggcagca
 gacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagtggttat
 tctttcagagctgtgcaaggcagtagaatcaagggtatgaaaccgtaggaaacttccatcatagttatggcca
 tgactactctagctagcagtggttaaatcattcagctacctgagagggggccctataactctctacggcTAA
 cctgaatggactacgactatcacgccccaaacattttacagccgggtgtcaaaaaccgcgtggacgtgggtta
 acatccctgctgggaggatcagccgtaattattataattggccttggtgctggctactattgtggccatgta

[illegible]

Vector de suministro de la TC-83 SEQ ID NO:7); genoma de la TC-83 con los nucleótidos 7.544-11.175 eliminados [proteínas estructurales alfavirus removidas]

ATAGGCGGCGCATGAGAGAGAAGCCAGACCAATTACCTACCCAAAATGGAGAAAGTTTACGTTTGACATCGAG
GAAGACAGCCCATTTCCCTCAGAGCTTTTGCAGCGGAGCTTCCCGCAGTTTGAGGTAGAAGCCAAGCAGGTTCAG
TGATAATGACCATGCTAATGCCAGAGCGTTTTTCGCATCTGGCTTCAAAACTGATCGAAACGGAGGTGGACC
CATCCGACACGATCCTTGACATTGGAAGTGC GCCCGCCCGCAGAATGTATTCTAAGCACAAAGTATCATTTGT
ATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAACTAAGCTGAAGAAAAACTG
TAAGGAAATAACTGATAAGGAATTGGACAAGAAAATGAAGGAGCTCGCCGCCGTGATGAGCGACCCGTGACC
TGGAAACTGAGACTATGTGCCCTCCACGACGACGAGTTCGTGTGCTACGAAGGGCAAGTTCGCTGTTTTACCAG
GATGTATACGCGGTTGACGGACCGACAAGTCTCTATCACCAAGCCAATAAGGGAGTTAGAGTCGCCCTACTG
GATAGGCTTTTGACACCACCCCTTTTATGTTTAAAGAACTTGGCTGGAGCATATCCATCATACTCTACCAAC
GGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCATATGCAGCTCTGACGTTATGGAGCGGTACAGT
AGAGGGATGTCCATTCTTAGAAAAGAAATTTGAAACCATCCAACAATGTTCTATTCTCTGTTGGCTCGAC
CATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCCTGCCGTCTGTATTTTCACTTACGTGGCAAGC
AAAAATTACATATCTCGGTGTGAGACTATAGTTAGTTGCGCAGGGTACGTCGTTAAAAGAATAGCTATCAGT
CCAGGCCGTGTATGGGAAGCCTTCAGGCTATAGCTGCTACGATGCACCGCAGGGATTCTTGTGCTGCAAAAGT
CAGCAGACGATTTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTGCCAGCTACATTGTGTGACC
AAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCGCAAAAACCTGCTGGTTGGGCTCAACCAG
CGTATAGTCGTCAACGGTTCGCACCCAGAGAAAACACCAATACCATGAAAAATTACCTTTTGGCCGCTAGTGGC
CCAGGCATTTTGTAGGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAAAGGCCACTAGGACTACGAG
ATAGACAGTTAGTCATGGGGTGTTGTTGGGCTTTTAGAAGGCACAAGATAACATCTATTTATAAGCGCCCG
GATACCCAAACCATCATCAAAGTGAACAGCGATTTCCTACTCATTCGTGCTGCCCAGGATAGGCAGTAACAC
ATTGGAGATCGGGCTGAGAACAAAGAATCAGGAAAATGTTAGAGGAGCACAAGGAGCCGTCACCTCTCATTTA
CCGCCGAGGACGTACAAGAAGCTAAGTGC GCAGCCGATGAGGCTAAGGAGGTGCGTGAAGCCGAGGAGTTG
CGCGCAGCTCTACCACCTTTTGGCAGCTGATGTTGAGGAGCCACTCTGGAAGCCGATGTGCACTTGTATGTT
ACAAGAGGCTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTTACCAGCTACGATGGCGAGG
ACAAGATCGGGCTCTTACGCTGTGCTTTCTCCGCAGGCTGTACTCAAGAGTGAAAAATTATCTTGCATCCAC
CCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTTATGCCGTGGAACCATAACCA
TGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCAGGACTTTCAAGCTCTGAGTGAAAGTGCCA
CCATTGTGTACAACGAACGTGAGTTTCGTAAACAGGTACCTGCACCATATTTGCCACACATGGAGGAGCGCTG
AACCATGATGAAGAATATTACAAAACTGTCAAGCCCAGCGAGCAGCAGCGGCAATACCTGTACGACATCGA
CAGGAAACAGTCGCTACAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGCAGCTGGTGATCTCCCT
TCCATGAAATTCGCCCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTTACCAAGTACCAACCATAGGGGTG
TATGGCGTGGCAGGATCAGGCAAGTCTGGCATCATTTAAAAGCGCAGTCACCAAAAAAGATCTAGTGGTGAG
CGCCAAAGAAAAGAACTGTGCAGAAATTTATAAGGGACGTCAAGAAAATGAAAGGGCTGGACGTCAATGCCA
GAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCCGTAGAGACCTGTATATTGACGAAGCTTTT
GCTTGTGTCATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAAGGCAGTGCTCTGCGGGGA
TCCCAAACAGTGCGGTTTTTTTTTAACATGATGTGCCTGAAAGTGCAATTTTAACCACGAGATTTGCACACAAG
TCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCGTCTCAACCTTGTTTTACGAC
AAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTACCGGCAGTACCAAACCTAA
GCAGGACGATCTCATTTCTCACTTGTTTTCAGAGGGTGGGTGAAGCAGTTGCAAATAGATTACAAAGGCAACG
AAATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGCCGTTCCGTACAAGGTGAAT
GAAAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCTTACTGACCCGCACGGAGGACCGCATCGT
GTGGAAAAACACTAGCCGGCGACCCATGGATAAAAAACACTGACTGCCAAGTACCTTGGGAATTTTCACTGCCA
CGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGGAGAGACCGGACCCCTACCGAC
GTCTTCCAGAATAAGGCCAAACGTGTGTTGGGCCAAGGCTTTAGTGCCGGTGCTGAAGACCGCTGGCATAGA
CATGACCACCTGAACAATGGAACACTGTGGATTATTTTGAACGGCGCAAAAGCTCACTCAGCAGAGATAGTATT
TGAACCAACTATCGGTGAGTTCTTTTGGACTGCACTGGAATCTGGAATCTATTTCTGCAACCCATGTATTCCG
TTATCCATTAGGAATAACTACTGGGATAAATCCCGCTGCCCTAACATGTACGGGCTGAATAAAGAAGTGGT
CCGTGAGCTCTCTCGCAGGTACCCACAACCTGCCCTCGGGCAGTTGCCACTGGAAGAGTCTATGACATGAACA
CTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAAACAGAAGACTGCCTCATGCTTTT

GTCCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTTCGTGAGCAAATTGAAGGGCAGAACTGT
 CCTGGTGGTCGGGGAAAAGTTGTCCGTCCCAGGCAAAATGGTTGACTGGTTGTCAGACCGGCCTGAGGCTA
 CCTTCAGAGCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGACATAATATTTGTTAATGTG
 AGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTAAGCTTAGCATGTTGACCAA
 GAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTGAGCATAGGTTATGGTTACGCTGACAGGGCCA
 GCGAAAGCATCATTGGTGCTATAGCGCGGCAGTTCAAGTTTTCCCGGGTATGCAAACCGAAATCCTCACTT
 GAAGAGACGGAAGTTCTGTTTGTATTTCATTGGGTACGATCGCAAGGCCCGTACGCACAATCCTTACAAGCT
 TTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGATGTGCACCCTCATATCATG
 TGGTGCGAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGCTAACAGCAAAGGACAACCT
 GGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTTCCCGGAAAGCTTCGATTTACAGCCGATCGAAGTAGG
 AAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATTTCATGCCGTAGGACCAAACCTTCAACAAAGTTT
 CGGAGGTTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAAGATTGTCAACGATAACAAT
 TACAAGTCAGTAGCGATTCCACTGTTGTCCACCGGCATCTTTTCCGGGAACAAAGATCGACTAACCCAATC
 ATTGAACCATTTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATATACTGCAGGGACAAGAAAT
 GGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGATATGCATATCCGACGACTCT
 TCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGAGTTCTTTGGCTGGAAGGAAGGGCTA
 CAGCACAAGCGATGGCAAACTTTCTCATATTTGGAAGGGACCAAGTTTCACCAGGCGGCCAAGGATATAG
 CAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATGCATGTATATCCTCGGAGAA
 AGCATGAGCAGTATTAGGTGCAAAATGCCCCGTGCAAGAGTCCGAAGCCTCCACACCACCTAGCACGCTGCC
 TTGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCCTCACGTCCAGAACAATA
 CTGTGTGCTCATCCTTTCCATTGCCGAAGTATAGAATCACTGGTGTGCAGAAGATCCAATGCTCCCAGCCT
 ATATTGTTCTCACCGAAAGTGCTGCGTATATTCATCCAAGGAAGTATCTCGTGGAACACCACCGGTAGA
 CGAGACTCCGGAGCCATCGGCAGAGAACCAATCCACAGAGGGGACACCTGAACAACCACCACCTTATAACCG
 AGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATCATCGAAGAGGAAGAAGAGGATAGCATAAGTTTG
 CTGTGAGATGGCCCGACCCACCAGGTGCTGCAAGTCGAGGCAGACATTCACGGGCCGCCCTCTGTATCTAG
 CTCATCCTGGTCCATTTCCTCATGCATCCGACTTTGATGTGGACAGTTTATCCATACTTGACACCCTGGAGG
 GAGCTAGCGTGACCAGCGGGGCAACGTGAGCCGAGACTAACTCTTACTTCGCAAAGAGTATGGAGTTTCTG
 GCGCGACCGGTGCCTGCGCCTCGAACAGTATTACAGGAACCTCCACATCCCGCTCCGCGCACAAAGAACC
 GTCATTGACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCCGCCAGGCGTGAATAGGGTGA
 TCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCTTAGCAGGTCGGTCTCGAGAACCAACCTG
 GTCTCCAACCCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGGCGTTCTGAGACCAACA
 ATGACGGTTTGATGCGGGTGATACATCTTTTCCGACACCGGTCAAGGGCATTTACAACAAAAATCAG
 TAAGGCAAACGGTGCTATCCGAAGTGGTGTTGGAGAGGACCGAATTGGAGATTTTCGTATGCCCCGCGCCTC
 GACCAAGAAAAAGAAGAATTACTACGCAAGAAATTACAGTTAAATCCACACCTGCTAACAGAAGCAGATA
 CCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTGCAAGGCCTAGGGCATTATT
 TGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAACCTGCATCCTGTTCTTTGTATTTCATCTAGTGTGAAC
 CGTGCCTTTTCAAGCCCCAAGGTCGAGTGGAAGCCTGTAACGCCATGTTGAAAGAGAACTTTCCGACTGT
 GGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACATGGTTGACGGAGCTTCATGCTGCTTAG
 AACTGCCAGTTTTTGGCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTCCTATTTGGAACCCACAATA
 CGATCGGCAGTGCTTTCAGCGATCCAGAACACGCTCCAGAACGTCCTGGCAGCTGCCACAAAAGAAATTG
 CAATGTCACGCAAATGAGAGAATTGCCCCGATTGGATTGCGCGGCCTTTAATGTGGAATGCTTCAAGAAAT
 ATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCTTACTGAAGAAAACGTGGTA
 AATTACATTACCAATTAAGAGGACCAAAAGCTGCTGCTCTTTTTGCGAAGACACATAATTTGAATATGTT
 GCAGGACATACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGACGTGAAAGTGACTCCAGGAACAAAAC
 ATACTGAAGAACGGCCCCAAGGTACAGGTGATCCAGGCTGCCGATCCGCTAGCAACAGCGTATCTGTGCGGA
 ATCCACCGAGAGCTGGTTAGGAGATTAAATGCGGTCTGCTTCCGAACATTTCATACACTGTTTGATATGTC
 GGCTGAAGACTTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGTGTTCTGGAACTGACATCG
 CGTCGTTTGATAAAAGTGAGGACGACGCCATGGCTCTGACCGGTTAATGATTCTGGAAGACTTAGGTGTG
 GACGCAAGAGCTGTTGACGCTGATTGAGGCGGCTTTCGGCGAAATTTTCATCAATAACATTTGCCCAATAAC
 TAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTCTCACACTGTTTGTGAACACAGTCATTAACA
 TTGTAATCGCAAGCAGAGTGTTGAGAGAACGGCTAACCAGGATCACCATGTGCAGCATTTCATTGGAGATGAC
 AATATCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCACCTGGTTGAATATGGAAGT
 CAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGGTTTATTTTGTGTGACTCCG
 TGACCGGCACAGCGTGCCGTGTGGCAGACCCCTAAAAGGCTGTTTAAAGCTTGGCAAACCTCTGGCAGCA
 GACGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACACGCTGGAACCGAGTGGGTAT
 TCTTTTCAGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAAACCGTAGGAACCTCCATCATAGTTATGGCCA
 TGACTACTCTAGCTAGCAGTGTTAAATCATTTCAGCTACCTGAGAGGGGCCCTATAACTCTCTACGGCTAA

CCTGAATGGACTACGACTATCACGCCCAAACATTTACAGCCGCGGTGTCAAAAACCGCGTGACGTGGTTA
 ACATCCCTGCTGGGAGGATCAGCCGTAATTATTATAATTGGCTTGGTGCTGGCTACTATTGTGGCCATGTA
 CGTGCTGACCAACCAGAAACATAATTGAATACAGCAGCAATTGGCAAGCTGCTTACATAGAACTCGCGGCG
 ATTGGCATGCCGCCTTAAAATTTTTATTTTATTTTCTTTTCTTTTCCGAATCGGATTTTGTTTTTAATAT
 TTCAAA
 AAAAAAAAAAAAA

Vector de producción del VEE (SEQ ID NO:8); genoma del VEE con los
 nucleótidos 7.544-11.175 eliminados, más el promotor 5' de T7, más
 sitios de restricción 3'

TAATACGACTCACTATAGGATGggcgggcgcatgagagaagcccagaccaattacctacccaaaATGGagaa
 agttcacgttgacatcgaggaagacagcccattcctcagagcgtttgcagcggagcgttcccgagtttgagg
 tagaagccaagcaggtcactgataatgaccatgctaattgccagagcgttttcgcatctggcttcaaaactg
 atcgaaacgggaggtggacccatccgacacgatccttgacattggaagtgcgcccgcgcgcagaaatgtattc
 taagcacaagtatcattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaa
 ctaagctgaagaaaaactgtaaggaaataactgataagggaattggacaagaaatgaaggagctgcgcgc
 gtcatgagcgcacctgacctggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagg
 gcaagtgcgtgtttaccaggatgtatacgcggttgacggaccgacaagtctctatcaccaagccaataagg
 gagttagagtcgcctactggataggcgtttgacaccaccccttttatgtttaagaacttggtggagcatat
 ccatcatactctaccaactgggcccagcgaacacgtgttaacggctcgtaacataggcctatgcagctctga
 cgttatggagcgggtcacgtagagggatgtccattcttagaaagaagtatttgaaaccatccaacaatgttc
 tattctctgttggctcgaccatctaccacgagaagagggacttactgaggagctggcacctgcccgtctgta
 ttctacttacgtggcaagcaaaaattacacatgtcgggtgtgagactatagttagttgcgacgggtacgtcgt
 taaaagaatagctatcagtcacggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagg
 gattcttgtgctgcaaagtgcagacacattgaacggggagaggggtctcttttcccggtgtgcacgtatgtg
 ccagctacattgtgtgaccaaagtactggcatactggcaacagatgtcagtgcgagcgcgcgcaaaaact
 gctggttgggctcaaccagcgtatagtcgtcaacggctgcacccagagaaacaccaataccatgaaaaatt
 accttttgcgcgtagtgggcccaggcatttgcgtaggtgggcaagggaatataagggaagatcaagaagatgaa
 agggcactaggactacgagatagacagttagtcagtggtgtgtgttgggcttttagaaggcacaagataac
 atctattttataagcgcccggatacccaaaccatcatcaaagtgaacagcgtattccactcattcgtgctgc
 ccaggataggcagtaacacattggagatcgggctgagaacaagaatcaggaaaaatgttagaggagcacaag
 gagccgtcacctctcattaccgcccaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggt
 cgcgtgaagccgaggagttgcgcgcagctctaccaccttggcagctgatgttgaggagcccactctggaag
 ccgatgtcgacttgatgttacaagaggctggggcgggtcagtgagagacacctcgtggcttgataaagggt
 accagctacgctggcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtga
 aaaattatcttgcacccacctctcgtgtaacaagtcagtgataaacacactctggccgaaaaggcggtt
 atgcgctggaaccataccatggtaaagttagtggtgccagagggacatgcaatacccgctccaggactttcaa
 gctctgagtgaaagtgccaccattgtgtacaacgaacgtgagttcgtaaacaggtaacctgcaccatattgc
 cacacatggaggagcgtgaaacactgatgaagaatattacaaaactgtcaagcccagcgcgcgcgcgcgcgc
 aatacctgtacgacatcgacaggaacacgtgcgtcaagaaagaactagtcaactgggctagggtcacaggc
 gagctggtggatcctcccttccatgaattcgctacgagagctctgagaacacgaccagccgctccttacca
 agtaccaaccataggggtgtatggcgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcacca
 aaaaagatctagtggtgagcgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaaa
 gggctggacgtcaatgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccct
 gtatattgacgaagcttttgccttgcagtgaggtactctcagagcgtcatagccattataagacctaaaa
 aggcagtgctctgcggggatcccaaacagtgcggttttttaacatgatgtgacctgaaagtgcattttaac
 cacgagatttgcacacaagtcttccacaaaagcatctctcgccgttgcactaaatctgtgacttcggtcgt
 ctcaaccttgttttacgacaaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacacta
 ccggcagtagcaaacctaagcaggacgatctcattctcacttgtttcagaggggtgggtgaagcagttgcaa
 atagattacaaaggcaacgaaataatgacggcagctgcctctcaagggctgaccgtaaaagggtgtgtatgc
 cgttcgggtacaaggtgaatgaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgaccc
 gcacggaggacgcctcgtgtggaaaacactagcggcgaccccatggataaaaacactgactgccaagtac
 cctgggaatttctactgccacgatagaggagtggaagcagagcatgatgccatcatgaggcacatcttggga
 gagaccggaccctaccgacgtcttccagaataaggcaaacgtgtgttgggccaaggcttttagtgccggtgc
 tgaagaccgctggcatagacatgaccactgaacaatggaacactgtggattattttgaacggacaaaagct
 cactcagcagagatagttatgaaccaactatgcgtgaggttctttggactcgatctggactccgggtctatt
 ttctgcacccactgttccgttatccattaggaataatcactgggataactccccgtgcgcctaacatgtacg

ggctgaataaagaagtgggtccgtcagctctctcgcaggtaccacaactgcctcgggcagttgccactgga
 agagtctatgacatgaacactgggtacactgcgcaattatgatccgcgcataaacctagtagctgtaaacag
 aagactgcctcatgcttttagtccctccaccataatgaacacccacagagtgaacttttcttcattcgtcagca
 aattgaagggcagaactgtcctgggtgggtcggggaaaagtgtccgtcccaggcaaaatgggtgactgggtg
 tcagaccggcctgaggctaccttcagagctcgggtggatttaggcattcccagggtgatgtgccaaatatga
 cataatatttgttaatgtgaggaccccatataaataccatcactatcagcagtgatgaagacatgccatta
 agcttagcatgttgaccaagaaagcttgtctgcatctgaatcccggcggaacctgtgtcagcataggttat
 gggtacgctgacagggccagcgaaagcatcattgggtgctatagcgcggcagttcaagttttcccggtatg
 caaaccgaaatcctcacttgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgta
 cgcacaatccttacaagctttcatcaaccttgaccaacattttatacaggttccagactccacgaagccgga
 tgtgcacctcatatcatgtggtgcgaggggatattgccacggccacgaaggagtgtattataaatgtctgc
 taacagcaaaggacaacctggcgagggggtgtgaggagcgtgtataagaaattcccggaagcttcgatt
 tacagccgatcgaagtaggaaaagcgcgactgggtcaaaggtgcagctaaacatatcattcatgcccgtagga
 ccaaacttcaacaaagtttcgagggtgaaggtgacaaacagttggcagaggcttatgagtcctatcgctaa
 gattgtcaacgataacaattacaagtcagtagcgattccactgttgtccaccggcatctttccgggaaca
 aagatcgactaaccatcattgaaccatttgcgtgacagcttttagacaccactgatgcagatgtagccata
 tactgcagggaacaagaatgggaaatgactctcaaggaagcagtggttaggagagaagcagtgaggagat
 atgcatatccgacgactcttcagtgacagaacctgatgcagagctgggtgagggtgcatccgaagagttctt
 tggctggaagggaagggtacagcacaaagcgatggcaaaactttctcatatttgggaagggaaccaagttcac
 caggcgcccaaggatatagcagaaattaatgccatgtggcccggttgcaacggaggccaatgagcaggtatg
 catgtatatcctcggagaaagcatgagcagtagtaggtcgaaatgcccgcgtcgaagagtcggaagcctcca
 caccacctagcacgctgccttgccttgcctcatccatgccatgactccagaaagagtacagcgcctaaaagcc
 tcacgtccagaacaaattactgtgtgctcatcctttccattgcccgaagtatagaatcactgggtgtgcagaa
 gatccaatgtctccagcctatattgttctcaccgaaagtgccgtgcgtatattcatccaaggaagtatctcg
 tggaaacaccaccggtagacgagactccggagccatcggcagagaaccaatccacagagggggacacctgaa
 caaccaccacttataaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaaga
 agaggatagcataagtttgcgtgcagatggcccgaaccaccagggtgctgcaagtcgaggcagacattcacg
 ggccgcccctctgtatctagctcactcctgggtccattcctcatgcatccgactttgatgtggacagtttatcc
 atacttgacacctggaggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgc
 aaagagtatggagtttctggcgcgaccgggtgcctgcgcctcgaacagtagttcaggaacctccacatcccg
 ctccgcgcacaagaacaccgctcacttgcacccagcagggcctgctcgagaaccagcctagtttccaccccg
 ccaggcgtgaatagggtgatcactagagaggagctcgaggcgcttaccocgctcagcactcctagcaggtc
 ggtctcgagaaccagcctgggtctccaaccgcgagcgtaaatagggtgattacaagagaggagtttgagg
 cgttcgtagcacacaacaatgacgggttgatgcgggtgcatacatcttttctccgacaccgggtcaaggg
 catttacaacaaaaatcagtaaggcaaacgggtgctatccgaagtgggtgttgagaggaccgaattggagat
 ttgcgtatgcccgcgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatcccacac
 ctgctaacagaagcagataaccagtcagggaagggtggagaacatgaaagccataacagctagacgtattctg
 caaggcctagggcatttattgaaggcagaaggaaaagtggagtgctaccgaacctgcatcctgttccctt
 gtattcatctagtgatgaaccgtgccttttcaagcccaagggtcgcagtggaagcctgtaacgccatgttga
 aagagaactttccgactgtggcttcttactgtattattccagagtagcatgcctatttggacatgggtgac
 ggagcttcatgctgcttagacactgccagttttgcctgcaaagctgcgcagctttccaaagaaacactc
 ctatttggaaaccacaaatcagatcggcagtgcttcagcgatccagaacacgctccagaacgctcctggcag
 ctgccacaaaaagaaattgcaatgtcacgcaaatgagagaattgcccgtattggattcgggcgcccttaat
 gtggaatgcttcaagaaatattgctgtgaataatgaatattgggaaacgtttaaagaaaaccccatcaggct
 tactgaagaaaaagctggtaattacattaccaaattaaaaggacaaaaagctgctgctctttttgcaaga
 cacataatttgaatatgttgcaggacataccaatggacaggtttgtaatggacttaaagagagacgtgaaa
 gtgactccaggaacaaaacatactgaagaacggcccaagggtacaggtgatccaggctgcccgatccgctagc
 aacagcgtatctgtgcggaatccaccgagagctgggttaggagattaaatgcggtcctgcttccgaacattc
 atacactgtttgatattgtcggctgaagactttgacgctattatagccgagcacttccagcctggggatgtg
 gttctggaactgacatcgctcgtttgataaaaagtgaggacgacgcatggctcgtgacccgcttaagatg
 tctggaagacttaggtgtggacgcagagctgttgacgctgattgaggcggttccggcgaaatttcacaa
 tacatttgcacctaaaactaaattttaaattcggagccatgatgaaatctggaatgttccctcacactgttt
 gtgaacacagtcatttaacattgtaatcgcaagcagagtggttagagagaacggctaaccggatcaccatgtgc
 agcattcattggagatgacaatatcgtgaaaggagtcaaactcggaacaaattaatggcagacaggtgcgcca
 cctgggttgaatatggaagtcaagattatagatgctgtggtggggcgagaaagcgccttatttctgtggagg
 tttatttctgtgactccgtgaccggcacagcgtgcccgtgtggcagacccccctaaaaaggctgtttaagct
 tggcaaacctctggcagcagacgatgaacatgatgatgacaggagaagggtcattgcatgaagagtcacac

gctggaaccgagtggtggtattctttcagagctgtgcaaggcagtagaatcaagggtatgaaaccgtaggaact
 tccatcatagttatggccatgactactctagctagcagtggttaaattcattcagctacctgagagggggcccc
 tataactctctacggcTAacctgaatggactacgactatcacgcccacatttacagccgcggtgtcaaa
 aaccgctggacgtggttaacatccctgctgggaggatcagccgtaattattataattggcttggtgctgg
 ctactattgtggccatgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgc
 ttacatagaactcgcggcgattggcatgcccgttaaaatttttattttttttttttttttttttttttttt
 tcggtattttgttttttaatttttAAA
 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAtacgtagtttaaac

Vector de producción de la TC-83 (SEQ ID NO:9); genoma de la TC-83 con los nucleótidos 7.544-11.175 eliminados, más el promotor 5' de T7, más sitios de restricción 3'

TAATACGACTCACTATAGGATAGGCGGCATGAGAGAAGCCCAGACCAATTACCTACCCAAAATGGAGAA
 AGTTCACGTTGACATCGAGGAAGACAGCCCATTCCTCAGAGCTTTGCAGCGGAGCTTCCCGCAGTTTGAGG
 TAGAAGCCAAGCAGGTCACGTGATAATGACCATGCTAATGCCAGAGCGTTTTTCGCATCTGGCTTCAAACTG
 ATCGAAACGGAGGTGGACCCATCCGACACGATCCTTGACATTGGAAGTGCGCCCGCCGCGAGAATGTATTC
 TAAGCACAAAGTATCATTGTATCTGTCCGATGAGATGTGCGGAAGATCCGGACAGATTGTATAAGTATGCAA
 CTAAGCTGAAGAAAACTGTAAGGAAATAACTGATAAGGAATTGGACAAGAAAAATGAAGGAGCTCGCCGCC
 GTCATGAGCGACCCTGACCTGGAACTGAGACTATGTGCCTCCACGACGACGAGTCGTGTGCTACGAAGG
 GCAAGTCGCTGTTTACCAGGATGTATACGCGGTTGACGGACCGACAAGTCTCTATCACCAGCCAATAAGG
 GAGTTAGAGTCGCCTACTGGATAGGCTTTGACACCACCCCTTTTATGTTTAAGAACTTGGCTGGAGCATAT
 CCATCATACTCTACCAACTGGGCCGACGAAACCGTGTTAACGGCTCGTAACATAGGCCTATGCAGCTCTGA
 CGTTATGGAGCGGTCACGTAGAGGGATGTCCATTCTTAGAAAGAAGTATTTGAAACCATCCAACAATGTTC
 TATTCTCTGTTGGCTCGACCATCTACCACGAGAAGAGGGACTTACTGAGGAGCTGGCACCTGCCGTCTGTA
 TTTCACTTACGTGGCAAGCAAAATTACACATGTCGGTGTGAGACTATAGTTAGTTGCGACGGGTACGTCTGT
 TAAAAGAATAGCTATCAGTCCAGGCCTGTATGGGAAGCCTTCAGGCTATGCTGCTACGATGCACCGCGAGG
 GATTCTTGTGCTGCAAGTGACAGACACATTGAACGGGGAGAGGGTCTCTTTTCCCGTGTGCACGTATGTG
 CCAGCTACATTGTGTGACCAAATGACTGGCATACTGGCAACAGATGTCAGTGCGGACGACGCGCAAAACT
 GCTGGTTGGGCTCAACCAGCGTATAGTCGTCAACGGTCGCACCCAGAGAAACACCAATACCATGAAAAATT
 ACCTTTTGGCCGTAGTGGCCCAGGCATTTGCTAGGTGGGCAAAGGAATATAAGGAAGATCAAGAAGATGAA
 AGGCCACTAGGACTACGAGATAGACAGTTAGTCATGGGGTGTGTTGGGCTTTTAGAAGGCACAAGATAAC
 ATCTATTTATAAGCGCCCGGATACCCAAACCATCATCAAAGTGAACAGCGATTTCCACTCATTTCGTGCTGC
 CCAGGATAGGCAGTAACACATTGGAGATCGGGCTGAGAACAAGAATCAGGAAAATGTTAGAGGAGCACAAG
 GAGCCGTACCTCTCATTACCGCCGAGGACGTACAAGAAGCTAAGTGCGCAGCCGATGAGGCTAAGGAGGT
 GCGTGAAGCCGAGGAGTTGCGCGCAGCTCTACCACCTTTGGCAGCTGATGTTGAGGAGCCCACTCTGGAAG
 CCGATGTGCACTTGATGTTACAAGAGGCTGGGGCCGGCTCAGTGGAGACACCTCGTGGCTTGATAAAGGTT
 ACCAGCTACGATGGCGAGGACAAGATCGGCTCTTACGCTGTGCTTTCTCCGAGGCTGTACTCAAGAGTGA
 AAAATTATCTTGCATCCACCCTCTCGCTGAACAAGTCATAGTGATAACACACTCTGGCCGAAAAGGGCGTT
 ATGCCGTGGAACCATACCATGGTAAAGTAGTGGTGCCAGAGGGACATGCAATACCCGTCCAGGACTTTCAA
 GCTCTGAGTGAAAGTGCCACCATTGTGTACAACGAACGTGAGTTTCGTAAACAGGTACCTGCACCATATTGC
 CACACATGGAGGAGCGCTGAACACTGATGAAGAATATTACAAAACCTGTCAAGCCCAGCGAGCACGACGGCG
 AATACCTGTACGACATCGACAGGAAACAGTGCGTCAAGAAAGAACTAGTCACTGGGCTAGGGCTCACAGGC
 GAGCTGGTGGATCCTCCCTTCCATGAATTCGCTACGAGAGTCTGAGAACACGACCAGCCGCTCCTTACCA
 AGTACCAACCATAGGGGTGTATGGCGTGCCAGGATCAGGCAAGTCTGGCATCATTAAGGCGCAGTCACCA
 AAAAGATCTAGTGGTGAGCGCCAAGAAAGAAAACCTGTGCAGAAATTATAAGGGACGTCAAGAAAATGAAA
 GGGCTGGACGTCAATGCCAGAACTGTGGACTCAGTGCTCTTGAATGGATGCAAACACCCCGTAGAGACCCT
 GTATATTGACGAAGCTTTTGTCTGTGTCATGCAGGTACTCTCAGAGCGCTCATAGCCATTATAAGACCTAAAA
 AGGCAGTGCTCTGCGGGGATCCCAAACAGTGCGGTTTTTTTTAATCATGATGTGCCTGAAAGTGCATTTTAA
 CACGAGATTTGCACACAAGTCTTCCACAAAAGCATCTCTCGCCGTTGCACTAAATCTGTGACTTCGGTCTGT
 CTCAACCTTTGTTTTACGACAAAAAATGAGAACGACGAATCCGAAAGAGACTAAGATTGTGATTGACACTA
 CCGGCAGTACCAAACCTAAGCAGGACGATCTCATTTCTCACTTGTTCAGAGGGTGGGTGAAGCATTGCAA
 ATAGATTACAAAGGCAACGAAATAATGACGGCAGCTGCCTCTCAAGGGCTGACCCGTAAAGGTGTGTATGC
 CGTTCCGTACAAGGTGAATGAAAATCCTCTGTACGCACCCACCTCAGAACATGTGAACGTCCTACTGACCC
 GCACGGAGGACCGCATCGTGTGGAACAACTAGCCGGCGACCCATGGATAAAAAACTGACTGCCAAGTAC
 CCTGGGAATTTCACTGCCACGATAGAGGAGTGGCAAGCAGAGCATGATGCCATCATGAGGCACATCTTGA
 GAGACCGGACCCCTACCGACGTCTTCCAGAATAAGGCAAACGTGTGTTGGGCCAAGGCTTTAGTGCCGGTGC
 TGAAGACCGCTGGCATAGACATGACCACTGAACAATGGAACACTGTGGATTATTTTGAACGGACAAAGCT

CACTCAGCAGAGATAGTATTGAACCAACTATGCGTGAGGTTCTTTGGACTCGATCTGGACTCCGGTCTATT
 TTCTGCACCCACTGTTCCGTTATCCATTAGGAATAATCACTGGGATAACTCCCCGTCGCCTAACATGTACG
 GGCTGAATAAAGAAGTGGTCCGTCAGCTCTCTCGCAGGTACCCACAACCTGCCTCGGGCAGTTGCCACTGGA
 AGAGTCTATGACATGAACACTGGTACACTGCGCAATTATGATCCGCGCATAAACCTAGTACCTGTAAACAG
 AAGACTGCCTCATGCTTTAGTCTCCACCATAATGAACACCCACAGAGTGACTTTTCTTCATTCGTCAGCA
 AATTGAAGGGCAGAAGTGTCTGGTGGTTCGGGAAAAGTTGTCCGTCCCAGGCAAAATGGTTGACTGGTTG
 TCAGACCGGCCTGAGGCTACCTTCAGAGCTCGGCTGGATTTAGGCATCCCAGGTGATGTGCCCAAATATGA
 CATAATATTTGTTAATGTGAGGACCCCATATAAATACCATCACTATCAGCAGTGTGAAGACCATGCCATTA
 AGCTTAGCATGTTGACCAAGAAAGCTTGTCTGCATCTGAATCCCGGCGGAACCTGTGTGTCAGCATAGGTTAT
 GGTTACGCTGACAGGGCCAGCGAAAGCATCATTGGTGTCTATAGCGCGGCAGTTCAAGTTTTCGCGGTATG
 CAAACCGAAATCCTCACTTGAAGAGACGGAAGTTCTGTTTGTATTTCATTGGGTACGATCGCAAGGCCCGTA
 CGCACAACTCTTACAAGCTTTTCATCAACCTTGACCAACATTTATACAGGTTCCAGACTCCACGAAGCCGGA
 TGTGCACCCCTCATATCATGTGGTGCAGGGGATATTGCCACGGCCACCGAAGGAGTGATTATAAATGCTGC
 TAACAGCAAAGGACAACCTGGCGGAGGGGTGTGCGGAGCGCTGTATAAGAAATTCCCGGAAAGCTTCGATT
 TACAGCCGATCGAAGTAGGAAAAGCGCGACTGGTCAAAGGTGCAGCTAAACATATCATTTCATGCCGTAGGA
 CCAAACCTTCAACAAAGTTTTCGGAGGTTGAAGGTGACAAACAGTTGGCAGAGGCTTATGAGTCCATCGCTAA
 GATTGTCAACGATAACAATTACAAGTCAGTAGCGATTCCACTGTTGTCCACCGGCATCTTTTTCGGGAACA
 AAGATCGACTAACCCAATCATTGAACCATTGCTGACAGCTTTAGACACCACTGATGCAGATGTAGCCATA
 TACTGCAGGGACAAGAAATGGGAAATGACTCTCAAGGAAGCAGTGGCTAGGAGAGAAGCAGTGGAGGAGAT
 ATGCATATCCGACGACTCTTCAGTGACAGAACCTGATGCAGAGCTGGTGAGGGTGCATCCGAAGAGTTCTT
 TGGCTGGAAGGAAGGGCTACAGCACAAAGCGATGGCAAACTTTCTCATATTTGGAAGGGACCAAGTTTCAC
 CAGGCGGCCAAGGATATAGCAGAAATTAATGCCATGTGGCCCGTTGCAACGGAGGCCAATGAGCAGGTATG
 CATGTATATCCTCGGAGAAAGCATGAGCAGTATTAGGTGCAAAATGCCCCGTGCAAGAGTCGGAAGCCTCCA
 CACCACCTAGCACGCTGCCTTGCTTGTGCATCCATGCCATGACTCCAGAAAGAGTACAGCGCCTAAAAGCC
 TCACGTCCAGAACAAATTACTGTGTGCTCATCTTTCCATTGCCCAGATATAGAATCACTGGTGTGCAGAA
 GATCCAATGCTCCCAGCCTATATTGTTCTCACCGAAAGTGCCTGCGTATATTTCATCCAAGGAAGTATCTCG
 TGGAAACACCACCGGTAGACGAGACTCCGGAGCCATCGGCAGAGAACCAATCCACAGAGGGGACACCTGAA
 CAACCACCACTTATAACCGAGGATGAGACCAGGACTAGAACGCCTGAGCCGATCATCATCGAAGAGGAAGA
 AGAGGATAGCATAAGTTTGTCTGTGATGGCCCCGACCCACAGGTGCTGCAAGTCGAGGCAGACATTACAG
 GGCCGCCCTCTGTATCTAGCTCATCCTGGTCCATTCTCATGCATCCGACTTTGATGTGGACAGTTTATCTC
 ATACTTGACACCTTGAGGGGAGCTAGCGTGACCAGCGGGGCAACGTCAGCCGAGACTAAGCTTTACTTTCGC
 AAAGAGTATGGAGTTTCTGGCGCGACCGGTGCCTGCGCCTCGAACAGTATTTCAGGAACCTCCACATCCCG
 CTCCGCGCACAGAACACCGTCACTTGACCCAGCAGGGCCTGCTCGAGAACCAGCCTAGTTTCCACCCCG
 CCAGGCGTGAATAGGGTGATCACTAGAGAGGAGCTCGAGGCGCTTACCCCGTCACGCACTCCTAGCAGGTC
 GGTCTCGAGAACCAGCCTGGTCTCCAACCCGCCAGGCGTAAATAGGGTGATTACAAGAGAGGAGTTTGAGG
 CGTTTCGTAGCACACAACAATGACGGTTTGTATGCGGGTGCATACATCTTTTCTCCGACACCGGTCAAGGG
 CATTTACAACAAAAATCAGTAAGGCAAACGGTGTCTATCCGAAGTGGTGTGGAGAGGACCGAATTGGAGAT
 TTCGTATGCCCCGCGCCTCGACCAAGAAAAAGAAGAATTACTACGCAAGAAATTACAGTTAAATCCCACAC
 CTGCTAACAGAAGCAGATACCAGTCCAGGAAGGTGGAGAACATGAAAGCCATAACAGCTAGACGTATTCTG
 CAAGGCCTAGGGCATTATTTGAAGGCAGAAGGAAAAGTGGAGTGCTACCGAACCCCTGCATCCTGTTCTTT
 GTATTTCATCTAGTGTGAACCGTGCCTTTTCAAGCCCCAAGGTGCGAGTGGAAGCCTGTAACGCCATGTTGA
 AAGAGAACTTTCCGACTGTGGCTTCTTACTGTATTATTCCAGAGTACGATGCCTATTTGGACATGGTTGAC
 GGAGCTTCATGCTGCTTAGACACTGCCAGTTTTTGGCCTGCAAAGCTGCGCAGCTTTCCAAAGAAACACTC
 CTATTTGGAACCCACAATACGATCGGCAGTGCCTTCAGCGATCCAGAACACGCTCCAGAACGTCTTGGCAG
 CTGCCACAAAAAGAAATTGCAATGTCACGCAAATGAGAGAATTGCCCGTATTGGATTTCGGCGGCCTTTAAT
 GTGGAATGCTTCAAGAAATATGCGTGTAATAATGAATATTGGGAAACGTTTAAAGAAAACCCCATCAGGCT
 TACTGAAGAAAACGTGGTAAATTACATTACCAAAATTAAGGACCAAAAGCTGCTGCTCTTTTTCGGAAGA
 CACATAATTTGAATATGTTGCAGGACATACCAATGGACAGGTTTGTAAATGGACTTAAAGAGAGACGCTGAAA
 GTGACTTCCAGGAACAAACATACTGAAGAACGGCCCAAGGTACAGGTGATCCAGGCTCCGATCCGCTAGC
 AACAGCGTATCTGTGCGGAATCCACCGAGAGCTGGTTAGGAGATTAAATGCGGTCCTGCTTCCGAACATTC
 ATACACTGTTTGTATATGTGCGCTGAAGACTTTGACGCTATTATAGCCGAGCACTTCCAGCCTGGGGATTGT
 GTTCTGGAAGACTGACATCGCGTCTGTTTGTATAAAAGTGAGGACGACGCCATGGCTCTGACCGCGTTAATGAT
 TCTGGAAGACTTAGGTGTGGACGCGAGAGCTGTTGACGCTGATTGAGGCGGCTTTTCGGCGAAATTTTCATCAA
 TACATTTGCCCACTAAAATAAATTTAAATTCGGAGCCATGATGAAATCTGGAATGTTTCTCAGACTGTTT
 GTGAACACAGTCATTAACATTGTAATCGCAAGCAGAGTGTGAGAGAACGGCTAACCGGATCACCATGTGC
 AGCATTTCATTGGAGATGACAATATCGTGAAAGGAGTCAAATCGGACAAATTAATGGCAGACAGGTGCGCCA
 CCTGGTTGAATATGGAAGTCAAGATTATAGATGCTGTGGTGGGCGAGAAAGCGCCTTATTTCTGTGGAGGG

TTTATTTTGTGTGACTCCGTGACCGGCACAGCGTGCCGTGTGGCAGACCCCCTAAAAAGGCTGTTTAAGCT
 TGGCAAACCTCTGGCAGCAGACGATGAACATGATGATGACAGGAGAAGGGCATTGCATGAAGAGTCAACAC
 GCTGGAACCGAGTGGGTATTCTTTTTCAGAGCTGTGCAAGGCAGTAGAATCAAGGTATGAAACCGTAGGAACT
 TCCATCATAGTTATGGCCATGACTACTCTAGCTAGCAGTGTTAAATCATTTCAGCTACCTGAGAGGGGGCCCC
 TATAACTCTCTACGGCTAACCTGAATGGACTACGACTATCACGCCCCAACATTTACAGCCGCGGTGTCAAA
 AACCGCGTGGACGTGGTTAACATCCCTGCTGGGAGGATCAGCCGTAATTATTATAATTGGCTTGGTGCTGG
 CTACTATTGTGGCCATGTACGTGCTGACCAACCAGAAACATAATTGAATACAGCAGCAATTGGCAAGCTGC
 TTACATAGAACTCGCGGCGATTGGCATGCCGCTTAAATTTTTTATTTTATTTTCTTTTCTTTTCCGAAT
 CGGATTTTGTTTTTTAATATTTTCAA
 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAacgtagtttaaac

VEE-UbAAY (SEQ ID NO:14); vector de suministro VEE con epítomos
 tumorales de ratón del MHC clase I SIINFEKL (SEQ ID NO: 57) y AH1-A5
 insertados

ATGggcgcgcatgagagaagcccagaccaattacctacccaaaatggagaaagtacacgttgac
 atcgaggaagacagcccattcctcagagctttgcagcggagcttcccgcagtttgaggtagaagccaagca
 ggtcactgataatgaccatgctaatagccagagcgttttcgcatctggcttcaaaactgatcgaaacgggagg
 tggacccatccgacacgatccttgacattggaagtgcgcccgcgcagaatgtattctaagcacaagtat
 cattgtatctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaa
 aaactgtaaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgcgctcatgagcgcacc
 ctgacctggaaactgagactatgtgcctccacgacgacgagctcgtgtcgctacgaagggaagctcgctggt
 taccaggatgtatacgcggttgacggaccgacaagtctctatcaccaagccaataagggagtttagagtcgc
 ctactggataggctttgacaccaccccttttatgttttaagaacttggctggagcatatccatcatactcta
 ccaactgggcccgcagaaaccggtgttaacggctcgttaacataggcctatgcagctctgacggttatggagcgg
 tcacgtagagggatgtccattccttagaaagaagtatgtgaaccatccaacaatgttctattctctgttg
 ctgcacatctaccacgagaagagggacttactgaggagctggcacctgcgctctgtatttcacttacgtg
 gcaagcaaaattacacatgtcgggtgtgagactatagtttagttgcgacgggtacgtcggttaaagaatagct
 atcagtcacaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctg
 caaagtgcagacacattgaacggggagaggggtctcttttcccgtgtgcacgtatgtgccagctacattgt
 gtgaccaaataactggtcactggtcacaagatgtcagtgcgacgacgcgcaaaaactgctggttgggctc
 aaccagcgtatagtcgtcaacgggtcgcacccagagaaacaccaataaccatgaaaaattaccttttgcggcgt
 agtggcccaggcatttgcctaggtgggcaaaggaatataaggaagatcaagaagatgaaaggccactaggac
 tacgagatagacagtttagtcatgggggtgttgttgggcttttagaaggcacaagataacatctatttataag
 cgccggataccacaaacatcatcaaagtgaacagcgaatttcactcattcgtgctgcccaggataggcag
 taacacattggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagcgcgtcacctc
 tcattaccgcccaggagcgtacaagaagctaagtgcgcagcgcgatgaggctaaggaggtgcgtgaagccgag
 gagttgcgcgcagctctaccacctttggcagctgatgttgaggagccactctggaagccgatgtcgactt
 gatgttacaagaggctggggccggctcagtgaggacacctcgtggcttgataaagggtaccagctacgctg
 gcgaggacaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaaattatcttgc
 atccaccctctcgctgaacaagtcatagtgataacacactctggccgaaaaggcggttatgccgtggaacc
 ataccatggttaaagtagtggtgccagagggacatgcaataaccgctccaggactttcaagctctgagtga
 gtgccaccattgtgtacaacgaacgtgagttcgttaaacaggtacctgcaccatattgccacacatggagga
 gcgctgaacactgatgaagaatattacaaaactgtcaagcccagcgcagcgcagcggcgaataacctgtacga
 catcgacaggaaacagtgctgcaagaagaactagtcactgggctagggctcacaggcgcagctggtggatc
 ctcccttccatgaattcgctacgagagctctgagaacacgaccagccgctccttaccagtaaccaaccata
 ggggtgtatggcgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcaccaaaaaagatctagt
 ggtgagcgcgaagaaagaaaactgtgcagaaattataagggaagcgtcaagaaaatgaaagggtggagctca
 atgccagaactgtggactcagtgctcttgaatggatgcaaacaccccgtagagaccctgtatattgacgaa
 gcttttgcctgtcatgcaggctactctcagagcgcctcatagccattataagacctaaaaaggcagtgctctg
 cggggatcccaaacagtgcggttttttaacatgatgtgcctgaaagtgcatttttaaccacgagatttgca
 cacaagtcttccacaaaagcatctctcgccgttgactaaatctgtgacttcggctcgtctcaaccttggtt
 tacgacaaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaacca
 accaagcaggagcatctcattctcattgtttcagaggggtgggtgaagcagttgcaaatagattacaag
 gcaacgaaataatgacggcagctgcctctcaagggtgaccgtaaaagggtgtgtatgccgttcggtacaag
 gtgaatgaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggaccg
 catcgtgtggaaaacactagccggcgacccatggataaaaacactgactgccaagtacctgggaatttca
 ctgccacgatagaggagtggaagcagagcatgatgccatcatgaggcacatcttggagagaccggaccct
 accgacgtcttccagaataaggcaaacgtgtgttgggccaaggcttttagtgccggtgctgaagaccgctgg

catagacatgaccactgaacaatggaacactgtggattatTTTTgaaacggacaaagctcactcagcagaga
 tagtattgaaccaactatgcgtgaggttctttggactcgatctggactccgggtctatTTTTctgacccact
 gttccgttatccattaggaataatcactgggataactccccgtcgcctaacatgtacgggctgaataaaga
 agtgggtccgtcagctctctcgcaggtaccacaaactgcctcgggcagttgccactggaagagtctatgaca
 tgaacactgggtacactgcgcaattatgatccgcgcataaacctagtagcctgtaaacagaagactgcctcat
 gctttagtcctccaccataatgaacacccacagagtgaacttttcttcattcgtcagcaaattgaagggcag
 aactgtcctgggtgggtcggggaaaagtgtgcgtcccaggcaaaatgggtgactgggtgtcagaccggcctg
 aggctaccttcagagctcgggtggatttaggcatcccagggtgatgtgccaaatatgacataatatTTgtt
 aatgtgaggaccccatataaataccatcactatcagcagtgatgaagaccatgccattaagcttagcatggt
 gaccaagaaaagcttgtctgcatctgaatccccggcggaacctgtgtcagcatagggttatgggtacgtgaca
 gggccagcgaaaagcatcattgggtgctatagcggcgagttcaagttttccggggtatgcaaaccgaaatcc
 tcaattgaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtagcacaatcctta
 caagctttcatcaaccttgaccaacatttatacaggttccagactccacgaagccggatgtgcacctcat
 atcatgtggtgagggggatattgccacggccacgaaggagtattataaatgctgctaacagcaaaagga
 caacctggcgaggggtgtgcggagcgtgtataagaaattcccgaaagcttcgatttacagccgatcga
 agtaggaaaagcgcgactgggtcaaagggtgcagctaaacatatcattcatgccgtaggaccaaacttcaaca
 aagtttcggaggttgaagggtgacaaacagttggcagaggcttatgagtccatcgctaagattgtcaacgat
 aacaattacaagtcagtagcgattccactgttgtccacccggcatcttttccgggaacaaagatcgactaac
 ccaatcattgaaccatttgcgtgacagcttttagacacccactgatgcagatgtagccatatactgcagggaca
 agaaatgggaaatgactctcaaggaagcagtggttaggagagaagcagtgaggagatatgcataatccgac
 gactcttcagtgacagaacctgatgcagagctgggtgaggggtgcacccgaagagttctttgggtggaaggaa
 gggctacagcacaagcgatggcaaaactttctcatatttggaaaggaccaaagtttcaccaggcgcccaagg
 atatagcagaaattaatgccatgtggcccggttgcaacggaggccaatgagcaggtatgcatttatctc
 ggagaaagcatgagcagatttaggtcgaaatgccccgtcgaagagtcggaagcctccacaccacctagcac
 gctgccttgcttgcacatccatgcatgactccagaaagagtacagcgcctaaaagcctcacgtccagaac
 aaattactgtgtgctcatcctttccattgccgaagtatagaatcactgggtgtgcagaagatccaatgctcc
 cagcctatatTTgttctcaccgaaagtgcctgcgtatatcattcatccaagggaagtatctcgtggaaacaccacc
 ggtagacgagactccggagccatcggcagagaaccaatccacagaggggacacctgaacaaccaccactta
 taaccgaggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcata
 agtttgcgtgacagatggcccgacccaccagggtgctgcaagtcgaggcagacattcacgggcccgcctctgt
 atctagctcatcctgggtccattcctcatgcatccgactttgatgtggacagtttatccataacttgacaccc
 tggagggagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaagagttatggag
 tttctggcgagccgggtgcctgcgcctcgaacagtatccaggaaccctccacatcccgtcccgcgcaaaag
 aacacggtcacttgcacccagcagggcctgctcgagaaccagcctagtttccaccccgcagggcgtgaata
 ggggtgatcactagagaggagctcgaggcgcttaccgcgtcacgcactcctagcaggtcgggtctcgagaacc
 agcctgggtctccaaccgcgcagggcgtaaatagggtgattacaagagaggagtttgaggcggttcgtagcaca
 acaacaatgacgggtttgatgcgggtgcatacatcttttcccgacacccggtcaagggcattttacaacaaa
 aatcagtaaggcacaacgggtgctatccgaagtggtgttggagaggaccgaattggagatttctgtatgccccg
 cgcctcgaccaagaaaaagaagaattactacgcaagaaattacagttaaatccacacactgctaacagaag
 cagataccagtcaggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggccctagggc
 attatttgaaggcagaaggaaaagtggagtgtaccgaaccctgcatcctgttccctttgtattcatctagt
 gtgaaccgtgccttttcaagccccagggtcgagtggaagcctgtaacgccatggtgaaagagaactttcc
 gactgtggcttcttactgtattattccagagtacgatgcctatttggacatgggttgacggagcttcatgct
 gcttagacactgccagtttttgcctgcaaagctgcgcagctttccaaagaaacactcctatttggaaacc
 acaatacgatcggcagtgccctcagcgatccagaacacgctccagaacgctcctggcagctgccacaaaaag
 aaattgcaatgtcacgcaaatgagagaattgcccgtattggattcggcggcctttaatgtggaatgcttca
 agaaatatgcgtgtaataatgaatatTTgggaaacgtttaaagaaaaccccatcaggcttactgaagaaaac
 gtggtaaatattacattaccaaattaaaaggaccaaaagctgctgctctttttgcgaagacacataatttgaa
 tatgttcaggacataccaatggacaggtttgtaatggacttaagagagacgtgaaagtgactccaggaa
 caaacatactgaagaacggcccaagggtacaggtgatccaggctgccgatccgtagcaacagcgtatctg
 tgcggaatccaccgagagctgggttaggagattaaatgcgggtcctgcttccgaacattcatacactggttga
 tatgtcggctgaagactttgacgctattatagccgagcactccagcctggggattgtgttctggaaactg
 acatcgcgtcgtttgataaaagtggagacgagccatggctctgaccgcgttaatgattctggaagactta
 ggtgtggagcagagctgttgacgctgattgaggcggtttcggcgaaatttcatcaatacatttgcacac
 taaaactaaattttaaattcggagccatgatgaaatctggaatgttccctcacactgtttgtgaacacagtc
 ttaacattgtaatcgcaagcagagtggttagagagaacggctaaccggatcaccatgtgcagcattcattgga
 gatgacaatatcgtgaaaggagtc aaatcggacaaattaatggcagacaggtgcgccacctgggttgaatat

ggaagtcaagattatagatgctgtggtgggcgagaaagcgcccttatttctgtggaggggtttattttgtgtg
 actccgtgaccggcacagcgtgccgtgtggcagacccccctaaaaaggctgtttaagcttggcaaacctctg
 gcagcagacgatgaacatgatgatgacaggagaagggcattgcatgaagagtcaacacgctggaaccgagt
 gggatttctttcagagctgtgcaaggcagtagaatcaaggatgaaaccgtaggaacttccatcatagtta
 tggccatgactactctagctagcagtggttaaatcattcagctacctgagagggggcccctataactctctac
 ggctaacctgaatggactacgactctagaatagtctttaattaaagtcggccatatgaggccaccatgCAG
 ATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCTAGAGGTGGAGCCCAGTGACACCATCGAGAACGT
 GAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAAGCAGC
 TGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCTG
 AGAGGTGGCGCTGCTTACAGTATAATCAACTTTGAAAACTGGCTGCTTACGGCATCCTGGGCTTTGTGTT
 TACACTGGCTGCCTACCTGCTGTTTGGCTATCCTGTGTACGTGGCCGCTTATGGACTGTGTACCCTGGTGG
 CCATGCTGGCTGCTTACAATCTGGTGCCTATGGTGGCCACAGTGGCCGCCTATTGTCTTGGCGGACTGCTG
 CCAATGGTGGCAGCCTACAgcccagcctatgcgtatcatcagtttGCAGCCTACGGCCCAGACCAGGCgC
 TAAATTTGTGGCTGCCTGGACACTGAAAGCCCGCTGGACCAGGTCCTGGACAGTACATCAAGGCCAACA
 GCAAGTTCATCGGCATCACCGAACTCGGCCAGGACCAGGCTATCCCTACGATGTGCCTGATTACGCCTGA
 TagTGATGATTGCAACGGCCGtatcacgccccaaacattttacagccgcggtgtcaaaaaaccgcgtggacgtg
 gttaacatccctgctgggaggatcagccgtaattattataaattggcttggctgctggctactattgtggcca
 tgtacgtgctgaccaaccagaaacataattgaatacagcagcaattggcaagctgcttacatagaactcgc
 ggcgattggcatgccgccttaaaattttttatt
 aatattttAA
 AAAAAAAAAAAAAAAAAA

VEE-Luciferasa (SEQ ID NO:15); vector de suministro VEE con el gen de
 luciferasa insertado en 7.545

ATGggcgggcgcagatgagagaagccagaccaattacctacccaaaATGgagaaagttcacggtgacatcgag
 gaagacagcccatttcctcagagctttgcagcggagcttccgcagtttgaggtagaagccaagcaggtcac
 tgataatgaccatgctaattgccagagcgttttcgcatctggcttcaaaactgatcgaaacggaggtggacc
 catccgacacgatcccttgacattggaagtgcgcccgcgcgcagaatgtattctaagcacaagtatcattgt
 atctgtccgatgagatgtgcggaagatccggacagattgtataagtatgcaactaagctgaagaaaaactg
 taaggaaataactgataaggaattggacaagaaaatgaaggagctcgccgcgctcatgagcgcacctgacc
 tggaaactgagactatgtgcctccacgacgacgagtcgtgtcgctacgaagggcaagtcgctgtttaccag
 gatgtatacgcggttgacggaccgacaagtctctatcaccaagccaataagggagtttagagtcgcctactg
 gataggctttgacaccaccccttttatgtttaagaacttggctggagcatatccatcatactctaccaact
 gggccgacgaacccgtgttaacggctcgtaacataggcctatgcagctctgacgttatggagcggtcacgt
 agagggatgtccattcttagaagaagtatattgaaacatccaacaatgttctattctctgttggctcgac
 catctaccacgagaagagggacttactgaggagctggcacctgccgtctgtattttcacttacgtggcaagc
 aaaattacacatgtcggtgtgagactatagtttagttgcgacgggtacgtcgttaaaagaatagctatcagt
 ccaggcctgtatgggaagccttcaggctatgctgctacgatgcaccgcgagggattcttgtgctgcaaggt
 gacagacacattgaacggggagaggggtctcttttcccgctgtgcagctatgtgccagctacattgtgtgacc
 aaatgactggcatactggcaacagatgtcagtgccgacgacgcgcaaaaactgctgggttgggctcaaccag
 cgtatagtcgtcaacggctgcacccagagaaacaccaataccatgaaaaattaccttttggccgtagtggc
 ccaggcatttgcctaggtgggcaaggaatataaggaagatcaagaagatgaaaggccactaggactacgag
 atagacagtttagtcatgggggtgttgggtgggcttttagaaggcacaagataacatctatttataagcgccg
 gatacccaaaccatcatcaaagtgaacagcgatttccactcattcgtgctgccaggataggcagtaacac
 attggagatcgggctgagaacaagaatcaggaaaatgttagaggagcacaaggagccgtcacctctcatta
 ccgcccaggacgtacaagaagctaagtgcgcagccgatgaggctaaggaggtgcgtgaagccgaggaggtg
 cgcgacgctctaccacctttggcagctgatgttgaggagccactctggaagccgatgtcgacttgatgtt
 acaagaggctggggccggctcagtgaggacacctcgtggcttgataaagggttaccagctacgctggcgagg
 acaagatcggctcttacgctgtgctttctccgcaggctgtactcaagagtgaaaaaattatcttgcacccac
 cctctcgctgaacaagtcatagtgataacacactctggccgaaaagggcggttatgccgtggaaccatacca
 ttggtaaagtagtggtgccagagggacatgcaatacccgctccaggactttcaagctctgagtgaagtgcca
 ccattgtgtacaacgaacgtgagttcgtaaacaggtacctgcacatattggcacacatggaggagcgctg
 aacactgatgaagaatattacaaaactgtcaagccagcgcagcagcgcggaataacctgtacgacatcga
 caggaaacagtgctcaagaaagaactagtcactgggctagggctcacaggcgagctggtggatccctccct
 tccatgaattcgcctacgagagctctgagaacacgaccagccgctccttaccagtaaccaaccataggggtg

tatggcgtgccaggatcaggcaagtctggcatcattaaaagcgcagtcacccaaaaagatctagtggtag
 cgccaagaaagaaaactgtgcagaaattataagggacgtcaagaaaatgaaagggctggacgtcaatgcc
 gaactgtggactcagtgtctctgaatggatgcaaacaccccgtagagacctgtatattgacgaagctttt
 gcttgtcatgcaggactctcagagcgctcatagccattataagacctaaaaaggcagtgctctgcgggga
 tcccaaacagtgccgttttttaacatgatgtgcctgaaagtgcattttaaccacgagatttgcacacaag
 tcttccacaaaagcatctctcgccgttgactaaatctgtgacttcgggtcgtctcaacctgttttacgac
 aaaaaatgagaacgacgaatccgaaagagactaagattgtgattgacactaccggcagtaccaaacctaa
 gcaggacgatctcattctcacttgtttcagaggggtgggtgaagcagttgcaaatagattacaaaggcaacg
 aaataatgacggcagctgcctctcaagggctgacccgtaaaggtgtgtatgccgttcgggtacaaggtgaat
 gaaaatcctctgtacgcacccacctcagaacatgtgaacgtcctactgacccgcacggaggacccgatcgt
 gtggaaaacactagccggcgacccatggataaaaaacactgactgccaagtacctgtgggaatttctactgcca
 cgatagaggagtggaagcagagcatgatgccatcatgaggcacatcttggagagaccggacctaccgac
 gtcttccagaataaggcaaacgtgtgttgggccaaggctttagtgccgggtgctgaagaccgctggcataga
 catgaccactgaacaatggaacactgtggattattttgaaacggacaagctcactcagcagagatagtat
 tgaaccaactatgcgtgaggttcttggactcgatctggactccgggtctattttctgcacccactgttccg
 ttatccattaggaataatcactgggataactcccgctgcctaactgtacgggctgaataaagaagtggt
 ccgtcagctctctcgaggtaccacaaactgcctcgggcagttgccactggaagagtctatgacatgaaca
 ctggtacactgcgcaattatgatccgcgcataaacctagtacctgtaaacagaagactgcctcatgcttta
 gtctccaccataatgaacacccacagagtgacttttcttctcgtcagcaaatgaaagggcagaactgt
 cctggtggtcggggaaaagtgttgcgtcccaggcaaaatgggtgactgggtgtcagacccggcctgaggcta
 ccttcagagctcggttggtatttaggcattcccagggtgatgtgcccaaataatgacataatatttgttaagt
 aggaccccatataaataccatcactatcagcagtgatgaagaccatgccattaagcttagcatgttgacca
 gaaagcttgtctgcattctgaatcccgccggaacctgtgtcagcataggttatggttacgctgacagggcca
 gcgaaagcatcattggtgctatagcgcggcagttcaagttttcccggtatgcaaaccgaaatcctcactt
 gaagagacggaagttctgtttgtattcattgggtacgatcgcaaggcccgtagcacaatccttacaagct
 ttcatcaaccttgaccaacattttatacaggttccagactccacgaagccggatgtgcacctcatatcatg
 tgggtgcgaggggatatttgccacggccacccgaaggagtgtattataaatgctgctaacagcaaaaggacaac
 ggccgaggggtgtgcggagcgctgtataagaaattcccggaagcttcgatttacagccgatcgaagtagg
 aaaagcgcgactggtcaaaggtgcagctaaacatatcattcatgcccgtaggaccaaacttcaacaaagttt
 cggaggttgaaggtgacaaacagttggcagaggttatgagtcctcgttaagattgtcaacgataacaat
 tacaagtcagtagcgattccactgttgtccaccggcactcttttccgggaacaaagatcgactaaccacatc
 attgaaccatttgcctgacagctttagacaccactgatgcagatgtagccatatactgcaggggacaagaaat
 gggaaatgactctcaaggaagcagtggttaggagagaagcagtgaggagatagcatatccgacgactct
 tcagtgacagaacctgatgcagagctgggtgaggggtgcacccaagagttcttgggtggaaggaagggtc
 cagcacaagcgatggcaaaactttctcatatttgggaagggaaccaagtttaccaggcgcccaaggatatag
 cagaaattaatgccatgtggcccggttgcaacggaggccaatgagcaggtatgcatgtatatcctcgagaa
 agcatgagcagttattaggtcgaaatgccccgtcgaaagagtcggaagcctccacaccacctagcagctgcc
 ttgcttgtgcattccatgccatgactccagaaagagtacagcgcctaaaagcctcacgtccagaacaaatta
 ctgtgtgctcatcctttccattgccgaagtatagaatcactgggtgtgcagaagatccaatgctcccagcct
 atattgttctcaccgaaagtgcctgcgtatattcatccaaggaagtatctcgtggaacaccaccggtaga
 cgagactccggagccatccggcagagaaccaatccacagaggggacacctgaacaaccaccacttataaccg
 aggatgagaccaggactagaacgcctgagccgatcatcatcgaagaggaagaagaggatagcataagtttg
 ctgtcagatggcccgaccaccagggtgctgcaagtcgaggcagacattcacgggcccgcctctgtatctag
 ctcatcctggtccattcctcatgcacccgactttgatgtggacagtttatccatacttgacacctggagg
 gagctagcgtgaccagcggggcaacgtcagccgagactaactcttacttcgcaaagagtatggagtttctg
 gcgcgacccgtgcctgcgcctcgaaacagtattcagggaacctccacatcccgctccgcgcacaagaacacc
 gtacttgcacccagcagggcctgctcgagaaccagcctagtttccaccccgccaggcggtgaataggggtga
 tcaatagagaggagctcgaggcgcttaccgccgtcacgcactcctagcaggtcggtctcgagaaccagcctg
 gtctccaaccgcccaggcgtaaataggggtgattacaagagaggagtttgaggcggttcgtagcacaaca
 atgacgggttgcgtgggtgcatacatcttctcccgacacgggtcaagggcatttacaacaaaaatcag
 taaggcaaacgggtgctatccgaagtgtgttggagagagaccgaattggagatttctgattccccgcgcctc
 gaccaagaaaaagaagaattactacgcaagaaattacagttaaatccacacctgctaacagaagcagata
 ccagtcacaggaaggtggagaacatgaaagccataacagctagacgtattctgcaaggcctagggcattatt
 tgaaggcagaaggaaggtggagtgctaccgaacctgcatcctgttcttctgtattcatctagtgtgaac
 cgtgccttttcaagcccccaaggtcgagtggaagcctgttaacgccatgttgaaagagaactttccgactgt
 ggcttcttactgtattattccagagtacgatgcctatttggacatgggttgacggagcttcatgctgcttag
 aactgccagtttttgcctgcaagctgcgcagctttccaaagaaacactcctatttggaaaccacaata

Ubiquitina (SEQ ID NO:38)

UbG76 0-228

ATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCTAGAGGTGGAGCCCAG
 TGACACCATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGGCATCCCCCCTGACC
 5 AGCAGAGGCTGATCTTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTAC
 AACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCTGAGAGGTGGC

Ubiquitina A76 (SEQ ID NO:39)

UbA76 0-228

10 ATGCAGATCTTCGTGAAGACCCTGACCGGCAAGACCATCACCTAGAGGTGGAGCCCAG
 TGACACCATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGGCATCCCCCCTGACC
 AGCAGAGGCTGATCTTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTAC
 AACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCTGAGAGGTGGC

15 Péptido señal de HLA-A2 (MHC clase I) (SEQ ID NO:40)

>MHC PepSeñal 0-78

atggccgcatggcgccccgaaccctcgtcctgctactctcgggggctctggccctgacccagacctggcgggctct

Dominio transmembrana de HLA-A2 (MHC clase I) (SEQ ID NO:41)

20 >Dominio TM de HLA A2 0-201

CCGtcttcccagcccaccatccCCATCGTGGGCAtcattgctggcctggttctcttggagctgtgatcactggagctgt
 ggtcgctgctgtgatgtggaggaggaagagctcagatagaaaaggaggaggactactctcaggctgcaagcagtgacagtg
 ccagggtctgatgtgtctctcacagcttgtaaagtgtga

25 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

```

ggcaccgattcggggcctgccggacttcgccgcacgctgcagaacctcgcccagcgcccaccatgccccggcagctcag
cgcgggcgccgcgctcttcgctccctggccgtaattttgcacgatggcagtc aaatgagagcaaaagcatttcagaaacc
5 agagattattctcaacctactgcagcagcaacagtacaggacataaaaaaacctgtccagcaaccagctaagcaagcacct
caccaaacttttagcagcaagattcatggatggatcatatcacctttcaaacagcgggccacagtaaaaattccaacaactacccc
agcaactacaaaaaactgcaaccaccagcccaattacctacaccctggtcacaaccaggccacaccaacaactca
cacacagctcctccagttactgaagttacagtcggccctagcttagccccttattcactgccaccaccatcacccaccagct
catacagctggaaccagttcatcaaccgtcagccacacaactgggaacaccactcaaccagtaaccagaccaccttcc
10 agcaactttatcgatagcactgcacaaaagcacaaccggtcagaagcctgatcaaccacccatgccccaggaacaacg
gcagctgcccacaataccacccgcacagctgcacctgcctccacggttcctgggcccaccctgcacctcagccatcgctcagt
caagactggaatttatcaggttctaaacggaagcagactctgtataaaagcagagatggggatacagctgattgttcaagaca
aggagtcggtttttcacctcgagatacttcaacatcgacccaacgcaacgcaagcctctgggaactgtggcaccggaaa
atccaaccttctgtgaattttcagggcggtttgtgaatctcacatttaccaggatgaagaatcatattatcagtgaaagtg
15 agcctatttgaccgtctcagatccagagacagttaccaaggaatcaaacatgcggtgggtgatgtccagacagcagtcgggc
attccttcaagtcgctgagtgaggaagcctccagttgtcagcccacctgcaggtgaaaacaaccgatgtccaacttcaagcc
tttgattttgaagtgaccactttggaaatgtggatgagtgctcgtctgactacacaattgtgttctctgtgattggggccatcggtt
tggtctctgccttatgggtatgggtgtctataaaatccgcctaagggtgtcaatcatctggataccagagaatctaattgttggccgg
ggggaatgaaaataatggaatttagagaactctttcatccctccaggatggatgttgggaaattccctcagagtgtgggtccttc
20 aaacaatgtaaaccaccatcttctattcaaatgaagtgaatcatgtgtgatttaagttcaggcagcacatcaatttctaaatactttt
tgtttttttatgaaagatatagtgagctgttttttctagtttccctttagaataatttttagccactcaaagtcaacattgagatatgttga
attaacataatatatgtaaagtagaataagccttcaaattataaaccaagggtcaattgtaactaataactactgtgtgtgcattga
agattttttttacccttgatcttaacaaagcctttgctttgttatcaaatggactttcagtgcttttactatctgtgtttatggttcatgtaa
catacatattcctgggtgtagcacttaactcctttccactttaaattgttttggttttgagacggagttcactctgtcaccaggctg
25 gagtacagtggcacgatctcggcttatggcaacctccgcctcccgggttcaagtattctcctgcttcagcttcccagtagctg
ggattacaggcacacactaccacgcctggctaattttgtattttattatagacgggttcaccatgttggccagactgggtcttgaa
ctcttgacctcaggtgatcccccacctcagcctcccaaagtgtctgggattacaggcatgagccattgcgcccggccttaaatg
tttttttaatatcaaaaaagaacaacatatctcaggttgtctaagtggtttttatgtaaaaccaaaaaagaacaatacagcttata
tttttatcttgatgactcctgctccagaattgctagactaagaattagggtggctacagatggtagaactaaacaataagcaagag
30 acaataataatggcccttaattattaacaaagtgcagagctaggctaagcactttatctatatctcatttcatttcacaacttata
agtgaatgagtaaactgagacttaagggaactgaatcacttaaatgtcacctggctaactgatggcagagccagagcttgaat
tcatgttggtctgacatcaaggctttgtcttctccctacaccaagttacctacaagaacaatgacaccacactctgcctgaagg
ctcacacctcataccagcatagctcaccttacagggaatgggtttatccaggatcatgagacattagggtagatgaaagga

```

gagctttgcagataacaaaatagcctatccttaataaatcctccactctctggaaggagactgaggggctttgtaaacattagt
 cagttgctcatTTTTatgggattgcttagctgggctgtaaagatgaaggcatcaaataaactcaaagtatttttaatttttgataat
 agagaaacttcgctaaccaactgttcttcttgagtgtatagccccatctgttgtaactgctgcttctgcacttcatatccatattcc
 tattgttactttattctgtagagcagcctgccaagaattttatttctgctgttttttctgctgctaaagaaaggaactaagtcaggatgtt
 5 aacagaaaagtccacataaccctagaattcttagtcaaggaataattcaagtcagcctagagacatgttgacttctcatgtg
 tttccttatgactcagtaagttggcaaggtcctgacttttagtcttaataaaacattgaattgtagtaaaggttttgcaataaaaactt
 actttgg

LAMP1 de ratón (SEQ ID NO:44)

10 >LampRatón1 0-1858

attccggagggtgaaaaacaatggcacaacgtgtataatggccagcttctctgcctccttctgaccacctacgagactgcgaat
 ggttctcagatcgtgaacatttccctgccagcctctgcagaagtactgaaaaatggcagttctgtggtaaagaaaatgtttctga
 cccagcctcacaattacttttgaagaggatatttactgacactcaacttcacaaaaatacaacacgttacagtggtccagcat
 atgtattttacataaactgtcagatacagaacattttcccaatgccatcagcaaagagatctacacatggattccacaactga
 15 catcaaggcagacatcaacaaagcataccgggtgtgtcagtgatatccgggtctacatgaagaatgtgaccgttgtgtccggg
 atgccactatccaggcctacgtgtcagtggtgcaacttcagcaaggaagagacacactgcacacaggatggaccttcccaa
 ccactgggccaccagccctcaccaccactgtgccacaaacccactgtatccaagtacaatgttactggtaacaacgg
 aacctgcctgtgtgcctctatggcactgcaactgaatatcacctacctgaaaaaggacaacaagacgggtgaccagagcgttc
 aacatcagcccaaatacacaatctagtgggagttgcggtatcaacttgggtgacctgaaagtggagaacaagaacagagcc
 20 ctggaattgcagtttgggatgaatgccagctctagcctgttttctgcaaggagtgcgcttgaatatgactcttctgatgccctagt
 gccacattcagcatctcaaccattcactgaaagctctcaggccactgtgggaaactcatacaagtgaacactgaggaa
 cacatctttgtcagaagatgctctccctcaatgtcttcagtggtgcagggtccaggcttcaaggtggacagtgcagggttgggtct
 gtggaagagtgtgttcaggatggttaacaacatgttgatccccattgtgtgtggcggtgcctggcagggtgatcctcatcgtcc
 tcattgcctacctcattggcaggaagaggagtacgcccgtatcagaccatctagcctgggtgggcagggtgcaccagagatg
 25 cacaggggctgttctcacatccccaagcttagataggtgtggaagggaggcacactttctggcaaactgttttaaatctgcttt
 atcaaatgtgaagttcatcttgcaacatttactatgcacaaaggaataactattgaaatgacgggtgtaattttgctaactgggtta
 aatattgatgagaaggctccactgatttacttttaagacttgggtttggttcttattcttttactcagatttaagcctatcaaagga
 tactctgttccagaccttggcctggcaaggggtggctgatggttaggctgcacacacttaagaagcaacgggagcaggggaag
 gcttgcacacaggcacgcacaggggtcaacctctggacacttggcttgggctacctggccttgggggggctgaactctggcatc
 30 tggctgggtacacaccccccaatttctgtgctctgccaccgtgagctgccactttcctaaatagaaaatggcattatttttatta
 ctttttgtaaagtgattccagttgtgttggcggtcaggggtggccctgtctctgcactgtgtacaataatagattcacactgctgac
 gtgtcttcagcgtagggtgggtgtacactgggcatcagctcacgtaatgcattgcctgtaacgatgctaataaaaa

ADNc de Lamp1 humana (SEQ ID NO:45)

>Lamp1 humana 0-2339

```

ggcccaaccgcccgcgccccgctctccgcaccgtacccggccgctcgcgccatggcggccccggcagcgcccg
5  gcgaccctgctgctgctactgctgttgctgctcggcctcatgcattgtgctcagcagcaatgtttatggtgaaaaatggca
   acgggaccgctgcataatggccaacttctgctgccttctcagtgaactacgacaccaagagtggccctaagaacatgac
   cttgacctgccatcagatgccacagtgggtgctcaaccgcagctcctgtggaaaagagaacacttctgaccccagtctcgtgat
   tgcttttgaagaggacatacactcactctcaatttcacgagaaatgcaacacgttacagcgtccagctcatgagttttgtttataa
   ctgtcagacacacaccttttcccaatgcgagctccaaagaaatcaagactgtggaatctataactgacatcagggcagata
10  tagataaaaaatacagatgtgttagtggcaccagggtccacatgaacaacgtgaccgtaacgctccatgatgccaccatcca
   ggcgtacctttcaacagcagcttcagcaggggagagacacgctgtgaacaagacaggccttcccaaccacagcgcccc
   ctgcgccaccagcccctcgccctcaccgctgcccaagagcccctctgtggacaagtacaacgtgagcggcaccaacggg
   acctgcctgctggccagcatggggctgcagctgaacctcacctatgagaggaaggacaacacgacggtgacaaggcttctc
   aacatcaacccaacaagacctcggccagcgggagctgcggcgcccacctgggtgactctggagctgcacagcaggggca
15  ccaccgtcctgctctccagttcgggatgaatgcaagttctagccggttttctacaaggaatccagttgaatacaattctctga
   cgccagagacctgcctttaaagctgccaacggctccctgcgagcgctgcaggccacagtcggcaattcctacaagtga
   cgcgaggagcacgtccgtgtcacgaaggcggtttcagtcataatcaaaagtgtgggtccaggcttcaagggtgaagggtg
   gccagtttgctctgtggaggagtgtctgctggacgagaacagcatgctgatcccatcgctgtgggtgggtgcctggcgggg
   ctggtcctcatgctcctcatgcctacctcgtcggcaggaagaggagtcacgcaggctaccagactatctagcctggtgcacg
20  caggcacagcagctgcaggggacctgttcttctctgggcttagggctcgtcgaaggggaggcacactttctggcaaacgtt
   tctcaaactgcttcatccaatgtgaagttcatcttcagcatttactatgcacaacagagtaactatcgaaatgacgggtgtaattt
   gctaactgggttaaataatttgctaactgggttaacattaatattaccaaagtaggattttgaggggtgggggtgctctctgaggg
   ggtgggggtgccgctgtctctgaggggtgggggtgccgctgtctctgaggggtgggggtgccgctctctgaggggtgggg
   gtgccgcttctctgaggggtgggggtgccgctctctgaggggtgggggtgctgctctccgaggggtggaatgccgct
25  gtctctgaggggtgggggtgccgctctaaattggctccatatcatttgagtttaggggtctggtttggtttcttattcttactgcact
   cagatttaagccttacaagggaagcctctgcccgtcacacgtaggacgcatgaaggctactcgtggtgaggtgacatgc
   tcacacattacaacagtagagagggaaaatcctaagacagaggaactccagagatgagtgctggagcgttcagttcagc
   tttaaaggccaggacgggccacacgtggctggcgccctcgttcagtgggcgacgctccttggcgctcttaatgtctgcagct
   caagggtggcactttttaaataaaaaatgggtgtattttttttttgtaaagtatttttggcttctgttgacattcggggtgatc
30  ctgttctgcgctgtgtacaatgtgagatcggtgcgttctcctgatgtttgccgtggcttggggattgtacacgggaccagctcacgt
   aatgcattgcctgtaacaatgtaataaaaagcctcttcttttaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

```

Secuencia de ácidos nucleicos de toxoide tetánico (SEQ ID NO:46)

CAGTACATCAAGGCCAACAGCAAGTTCATCGGCATCACCGAACTC

Secuencia de aminoácidos de toxoide tetánico (SEQ ID NO:47)

5 QYIKANSKFIGITEL

Secuencia de nucleótidos de PADRE (SEQ ID NO:48)

GCTAAATTTGTGGCTGCCTGGACACTGAAAGCCGCCGCT

10 Secuencia de aminoácidos de PADRE (SEQ ID NO:49)

AKFVAAWTLKAAA

WPRE (SEQ ID NO:50)

>WPRE 0-593

15 aatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttgctccttttacgctatgtggatacgtgcttta
atgcctttgtatcatgctattgcttcccgtatggctttcattttctcctcctgtataaatcctgggtgctgtctctttatgaggagttgtggc
ccgttgtcaggcaacgtggcgtggtgtgcactgtgtttgctgacgcaacccccactggttggggcattgccaccacctgtcagct
cctttccgggactttcgctttccccctccctattgccacggcggaactcatcgccgctgccttcccgtgctggacaggggctc
ggctgttgggactgacaattccgtggtgtgtcggggaagctgacgtccttccatggctgctgcctgtgttggcacctggattct
20 gcgcgggacgtccttctgctacgtccctcgccctcaatccagcggaccttcttcccgcgcctgctgcgggctctgcggcct
cttccgcgtcttcgccttcgcctcagacgagtcggatctcccttgggcccgcctccccgcctgt

IRES (SEQ ID NO:51)

>eGFP_IRES_SEAP_Inserto 1746-2335

25 tctccccccccctctccctccccccccctaacgttactggccgaagccgcttgaataaggccggtgtgcgtttgtctatat
gttattttccaccatattgccgtcttttggaatgtgagggcccggaaacctggccctgtcttcttgacgagcattcctaggggtcttt

ccctctcgccaaaggaatgcaaggtctgtgaatgtcgtaaggaagcagttcctctggaagcttctgaagacaaacaacg
 tctgtagcgacccttgcaggcagcggaacccccacctggcgacaggtgcctctgcggccaaaagccacgtgtataagat
 acacctgcaaaggcggcacaaccccagtgccacgtgtgagttggatagttgtgaaagagtcaaattggctctcctcaagcg
 tattcaacaaggggtgaaggatgccagaaggtacccattgtatgggatctgatctggggcctcggtgcacatgctttacat
 5 gtgttagtcgaggttaaaaaaacgtctaggccccccgaaccacggggacgtggtttcctttgaaaaacacgatgataatatg

GFP (SEQ ID NO:52)

atggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaa
 gttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagct
 10 gccgtgacctggccaccctcgtgaccacctgacctacggcgtgcagtgcttcagccgtaccccgaccacatgaagca
 gcacgacttctcaagtccgcatgccgaaggctacgtccaggagcgcaccatcttctcaaggacgacggcaactacaag
 acccgcgccgaggtgaagttcgagggcgacacctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacg
 gcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaagaacgg
 catcaaggtgaactcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactaccagcagaacacc
 15 cccatcggcgacggccccgtgctgctgcccgaaccactacctgagcaccagtcggccctgagcaaagaccccaacg
 agaagcgcgatcacatggtcctgctggagttcgtgaccgcccgggatcactctcgcatggacgagctgtacaagtag

SEAP (SEQ ID NO:53)

atgctgctgctgctgctgctgctgggcctgaggctacagctctccctgggcatcatcccagttgaggaggagaacccggacttc
 20 tgaaccgcgaggcagccgagggccctgggtgcccgaagaagctgcagcctgcacagacagccgccaagaacctcatc
 atcttctgggcatgggatgggggtgtctacggtgacagctgccaggatcctaaaagggcagaagaaggacaaactggg
 gcctgagatacccctggccatggaccgcttccatatgtggctctgtccaagacatacaatgtagacaaacatgtgccagaca
 gtggagccacagccacggcctacctgtgcgggtcaagggaacttccagaccattggcttgagtgcagccgccgctttaa
 ccagtgaacacgacacgcggcaacgaggtcatctccgtgatgaatcgggccaagaaagcaggggaagtcagtgggagtg
 25 gtaaccaccacacgagtgacgacgcctcgccagccggcacctacgcccacacggtgaaccgcaactgtactcgagc
 ccgacgtgcctgcctcgcccgccaggaggggtgccaggacatcgctacgcagctcatctccaacatggacattgacgtgat
 cctagggtggaggccgaaagtacatgtttcgatgggaaccccagaccctgagtaccagatgactacagccaagggtggga
 ccaggctggacgggaagaatctggtgcaggaatggctggcgaagcgccaggggtgccgggtatgtgtggaaccgcactgag
 ctcatgcaggcttccctggacccgtctgtgacctatctatgggtctctttgagcctggagacatgaaatacgagatccaccga
 30 gactccacactggacccctccctgatggagatgacagaggctgccctgcgctgctgagcaggaacccccgggcttctcc
 tctcgtggaggggtggtcgcatcgaccatggtcatcatgaaagcagggcttaccgggcactgactgagacgatcatgttcgac

gacgccattgagagggcgggccagctcaccagcgaggaggacacgctgagcctcgtcactgccgaccactcccacgtctt
 ctcttcggaggctacccctgcgagggagctccatcttcgggctggcccctggcaaggcccgggacaggaaggcctacac
 ggtcctctatacggaaacggtccaggctatgtgctcaaggacggcgcccgccggtgttaccgagagcgagagcgggga
 gccccgagtatcggcagcagtcagcagtgccccctggacgaagagacccacgcaggcgaggacgtggcggtgttcgcgcg
 5 cggcccgagggcgacacctggttcacggcgtgcaggagcagacctcatagcgcacgtcatggccttcgccgctgcctgga
 gccctacaccgctgcgacctggcgcccccccgccgaccaccgacgccgcgccccgggttactctagagtcggggcg
 gccggccgcttcgagcagacatgataa

Luciferasa Firefly (SEQ ID NO:54)

10 atggaagatgccaaaaacattaagaagggcccagcgccattctacccactcgaagacgggaccgcccggcgagcagctgc
 acaaagccatgaagcgctacgccctggtgccggcaccatcgccctttaccgacgcacatatcgaggtggacattacctacgc
 cgagtacttcgagatgagcgttcggctggcagaagctatgaagcgctatgggctgaatacaaacccatcggtcgtggtgtgc
 agcgagaatagcttcagttctcatgcccgtgttggtgccctgttcacgtgtggctgtggccccagctaacgacatctacaa
 cgagcgcgagctgctgaacagcatggcatcagccagcccaccgctcgtattcgtgagcaagaaagggctgcaaaagatcc
 15 tcaacgtgcaaaagaagctaccgatcatacaaaagatcatcatcatggatagcaagaccgactaccagggcttcaaagc
 atgtacaccttcgtgacttcccatttgccaccggcttcaacgagtagcacttcgtgcccgagagcttcgaccgggacaaaacc
 atcgccctgatcatgaacagtagtggcagtagccgattgccaagggcgtagccctaccgcaccgcaccgcttgttccgatt
 cagtcattgccgcgacccccatcttcggcaaccagatcatccccgacaccgctatcctcagcgtggtgccatttcaccacggctt
 cggcatgttcaccacgctgggctacttgatctgcggcttcgggtcgtgctcatgtaccgcttcgaggaggagctattcttcgca
 20 gcttgcaagactataagattcaatctgcctgctggtgccacactatttagcttcttcgtaagagcactctcatcgacaagtac
 gacctaaagcaacttcacgagatcgccagcggcgggcgccgctcagcaaggaggtaggtgaggccgtggccaaacgct
 tccacctaccaggcatccgccagggctacggcctgacagaaacaaccagcgccattctgatccccccgaaggggacga
 caagcctggcgcagtaggcaaggtggtgcccttctcagggctaaggtggtggacttgacaccggtaagacactgggtgtg
 aaccagcgcggcgagctgtgcgtccgtggccccatgatcatgagcggctacgttaacaaccccgaggctacaaacgctctc
 25 atcgacaaggacggctggctgcacagcggcgacatcgctactgggacgaggacgagcacttctcatcgtggaccggctg
 aagagcctgatcaaatacaagggctaccaggtagccccagccgaactggagagcatcctgctgcaacaccccaacatctt
 cgacgccgggggtcgccggcctgcccgacgacgatgccggcgagctgcccgccgagtcgctcgtgctggaacacggtaaa
 accatgaccgagaaggagatcgtggactatgtggccagccaggttacaaccgccaagaagctgcgcgggtggtgtgtgttcg
 tggacgaggtgcctaaaggactgaccggcaagttggacgcccgaagatccgcgagattctcattaaggccaagaagggc
 30 ggcaagatcgccgtgtaa

FMDV 2A (SEQ ID NO:55)

GTAAAGCAAACACTGAACTTTGACCTTCTCAAGTTGGCTGGAGACGTTGAGTC
CAATCCTGGGCCC

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
- 5 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).
- 10 4. Rizvi, N. A. et al. 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 al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189-2199 (2014).
6. Carreno, B. M. et al. Cancer immunotherapy. A dendritic cell vaccine
15 increases the breadth and diversity of melanoma neoantigen-specific T cells. *Science* **348**, 803-808 (2015).
7. Tran, E. et al. Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. *Science* **344**, 641-645 (2014).
8. Hachohen, N. y Wu, C. J.-Y. Solciitud de patente estadounidense:
20 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.PG&NR.>>
9. Lundegaard, C., Hoof, I., Lund, O. y Nielsen, M. State of the art and
25 challenges in sequence based T-cell epitope prediction. *Immunome Res.* **6 Supl 2**, S3 (2010).
10. Yadav, M. et al. 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
30 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 al. 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).
- 5 15. Rajasagi, M. et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124**, 453–462 (2014).
16. Downing, S. R. et al. Solicitud de patente de Estados Unidos 0120208706 - OPTIMIZATION OF MULTIGENE ANALYSIS OF TUMOR SAMPLES. (A1). en
10 <<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.PG NR.>>>
17. Target Capture for NextGen Sequencing - IDT. en
<<http://www.idtdna.com/pages/products/nextgen/target-capture>>
18. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic
15 mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152-1158 (2015).
19. Cieslik, M. et al. 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 al. The hidden genomic landscape of acute myeloid leukemia: subclonal structure revealed by undetected mutations. *Blood* **125**, 600-605 (2015).
- 20 21. Saunders, C. T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinforma. Oxf. Engl.* **28**, 1811–1817 (2012).
22. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213-219 (2013).
23. Wilkerson, M. D. et al. Integrated RNA and DNA sequencing improves
25 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
30 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 al. Nucleotide-resolution analysis of structural variants using BreakSeq and a breakpoint library. *Nat. Biotechnol.* **28**, 47-55 (2010).
27. Frampton, G. M. et al. 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 al. HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).

5 29. Liu, C. et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res.* **41**, e142 (2013).

30. Mayor, N. P. et al. HLA Typing for the Next Generation. *PloS One* **10**, e0127153 (2015).

10 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).

15 33. Maretty, L., Sibbesen, J. A. y Krogh, A. Bayesian transcriptome assembly. *Genome Biol.* **15**, 501 (2014).

34. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290-295 (2015).

20 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).

25 37. Rivas, M. A. et al. 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).

30 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 al. 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 al. 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 al. SF3B1 mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235**, 571-580 (2015).
43. Carithers, L. J. et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation Biobanking* **13**, 311–319 (2015).
- 5 44. Xu, G. et al. 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
- 10 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 al. 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).
- 15 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 al. A Quantitative Spatial Proteomics Analysis of Proteome Turnover in Human Cells. *Mol. Cell. Proteomics* **11**, M111.011429–M111.011429 (2012).
- 20 50. Duan, F. et al. 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>>
- 25 52. Calis, J. J. A. et al. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. *PLoS Comput. Biol.* **9**, e1003266 (2013).
53. Zhang, J. et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256–259 (2014)
- 30 54. Walter, M. J. et al. 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, Mikesch 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.
- 5 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,
10 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.
15 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
20 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.
- 25 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
30 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 al. (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 al. (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.
- 10 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 al. (2012). Aurora kinase A-specific T-cell receptor gene transfer redirects T lymphocytes to display effective antileukemia reactivity. *Blood* 119, 368-376.
- 15 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.
- 20 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.
- 25 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.
- 30

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, Ehrenguber 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.
- 5 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.
- 10 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.

DIVERSAS MODALIDADES

1. En la presente, se describe un vector viral que comprende un neoantígeno o
 15 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
 20 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
 25 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
 30 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.

5 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

10 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:

15 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:

20 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.

25 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.

30 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
5 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;

10 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.

15 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.

20 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
25 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.

30 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.

5 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:

10 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;

15 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.

20 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.

25 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.

30 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.

5 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.

10 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.

15 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.

20 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.

25 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.

30 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.

5 41. En determinadas modalidades, la vacuna tumoral comprende además un adyuvante.

42. En determinadas modalidades, la vacuna tumoral comprende además un excipiente.

10 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.

15 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.

20 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.

25 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.

30 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 mediante 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.
- 5 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
 - 10 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).
- 15 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.
- 20

51. En determinadas modalidades, el conjunto de probabilidades numéricas se identifica además mediante al menos características de no interacción MHC-alelo que

- 25 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
 - 30 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-DOBHLA-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-DOBHLA-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:

5 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;

10 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;

15 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:

20 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.

25 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.

30 61. En determinadas modalidades, las muestras comprenden muestras de tumores frescos 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:

5 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:

10 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
15 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
20 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:

25 codificar las secuencias peptídicas de entrenamiento 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

30 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.

5 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.

10 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.

15 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.

20 76. En determinadas modalidades, entrenar el conjunto de parámetros numéricos del modelo de presentación comprende:
codificar las secuencias peptídicas de entrenamiento mediante el uso de un esquema de codificación one-hot con left-pad.

25 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:

30 recibir datos de espectrometría de masas que comprendan datos asociados a múltiples péptidos aislados 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

- 5 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.

- 10 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

- 15 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.