

MÉTODO PARA SELECCIONAR UN NEOANTÍGENO PARA EL DESARROLLO DE UNA VACUNA CONTRA EL CÁNCER PERSONALIZADA

【Campo técnico】

5 REFERENCIA CRUZADA A SOLICITUDES RELACIONADAS

Esta solicitud reivindica el beneficio de prioridad y la prioridad a la solicitud de patente coreana n.º 10-2023-0110887, presentada el 23 de agosto de 2023, en la Oficina de Propiedad Intelectual de Corea, cuyas divulgaciones se incorporan en el presente documento mediante referencia.

10

La presente divulgación proporciona un método para seleccionar neoantígenos (péptidos inmunogénicos específicos de tumor) y el uso de los neoantígenos asociados a tumor seleccionados para el desarrollo de vacunas contra el cáncer personalizadas.

15 **【Antecedentes】**

El término “neoantígeno”, tal como se usa en el presente documento, se refiere a un antígeno generado mediante mutaciones específicas de cáncer y, más específicamente, a un péptido mutado expresado en la superficie celular de manera específica de cáncer que puede reconocerse como antígeno, incluyendo epítomos (proteínas de superficie). Dado que los neoantígenos se expresan exclusivamente en células cancerosas, pueden aplicarse al desarrollo de terapias dirigidas contra el cáncer. En la presente divulgación, los neoantígenos son un tipo de péptidos inmunogénicos y también pueden denominarse “neoantígenos específicos de tumor”, “antígenos específicos de tumor”, o “péptidos inmunogénicos específicos de tumor”.

25 El término “vacuna contra el cáncer personalizada” se refiere a un agente terapéutico o a un método de tratamiento en el que se seleccionan neoantígenos con capacidades de amplificación inmunitaria a partir de antígenos específicos de cáncer expresados en pacientes individuales, se producen en forma de proteínas o ARNm, y se

administran al paciente usando nanopartículas a base de lípidos tales como nanopartículas lipídicas (LNP, nanopartículas lipídicas sólidas (SLN), portadores lipídicos nanoestructurados (NLC), etc.) o complejos liposoma-ácido nucleico (por ejemplo, lipoplejo, liposomas), activando el sistema inmunitario del paciente para eliminar las
5 células cancerosas.

Con el fin de mediar en la eficacia de las vacunas contra el cáncer personalizadas, el factor clave es descubrir el conjunto de neoantígenos específicos para cada paciente individual y seleccionar y priorizar los neoantígenos inmunogénicos. Los neoantígenos inmunogénicos deben tener una afinidad de unión por el antígeno leucocitario humano
10 (HLA), que regula respuestas inmunitarias tales como la inmunidad celular y humoral en respuesta a la estimulación con antígeno. Cuando los receptores de células T reconocen las secuencias de neoantígeno específicas, que están unidas a HLA, en el cuerpo, se produce la activación inmunitaria frente a los neoantígenos.

Sin embargo, existen desafíos a la hora de descubrir neoantígenos
15 inmunogénicos aplicables al desarrollo de vacunas contra el cáncer personalizadas, ya que las mutaciones específicas de cáncer varían para cada paciente, y el tipo de HLA expresado difiere entre los individuos. Además, son escasos la recopilación y selección de datos para entrenar los modelos de predicción de neoantígenos por IA que puedan predecir neoantígenos inmunogénicos, así como el establecimiento de métodos para
20 seleccionar neoantígenos aplicables a pacientes individuales en el desarrollo de vacunas contra el cáncer personalizadas. Esto dificulta utilizar de manera eficiente los modelos de predicción de neoantígenos. Convencionalmente, se han desarrollado modelos de predicción por IA para medir la afinidad de unión entre péptidos inmunogénicos, que incluyen mutaciones en células cancerosas de pacientes, y HLA para seleccionar
25 neoantígenos que se espera que tengan eficacia antineoplásica cuando se aplican a las vacunas contra el cáncer. Sin embargo, ha sido difícil usar los péptidos inmunogénicos derivados como vacunas, ya que una alta afinidad de unión entre epítomos del paciente y HLA no conduce necesariamente a una fuerte respuesta inmunitaria de células T.

Adicionalmente, se han desarrollado tecnologías para predecir la inmunogenicidad de células T comparando las propiedades fisicoquímicas de los aminoácidos entre péptidos que contienen mutaciones en células cancerosas y péptidos no mutados en células no cancerosas. Sin embargo, este método, que se basa únicamente en información de
5 secuencia de aminoácidos, no logra reflejar la estructura proteica tridimensional. Como resultado, la tasa de éxito de predicción de respuestas inmunitarias no es alta porque el complejo HLA-péptido, que incluye los péptidos neoantígenos con mutaciones en células cancerosas, debe reconocerse exhaustivamente por el receptor de células T para desencadenar una respuesta inmunitaria, y la afinidad de unión sola no puede abarcar
10 las diversas respuestas inmunitarias.

【Divulgación】

【Problema técnico】

Un aspecto de la presente solicitud proporciona un método para seleccionar un
15 péptido inmunogénico específico de tumor (neoantígeno, antígeno específico de tumor) (incluyendo un método de predicción, un método de selección, un método de identificación, o un método para proporcionar información para la selección).

El método para seleccionar un péptido inmunogénico específico de tumor puede incluir las etapas de: (I) construir un modelo de predicción de péptido inmunogénico; y (II)
20 predecir y/o seleccionar un péptido inmunogénico específico de tumor.

La etapa (I) de construir un modelo de predicción de péptido inmunogénico puede incluir las siguientes etapas (a) a (d):

(a) obtener información de péptido a partir de una base de datos de péptidos, en el que la información de péptido puede incluir uno o más elementos de información
25 seleccionados de información de secuencia de péptidos derivados de células tumorales o tejidos tumorales, información de secuencia de aminoácidos de péptidos derivados de células normales o tejidos normales, información sobre el tipo de HLA de clase I que se une a los péptidos, información de afinidad de unión entre los péptidos y HLA de clase I,

y/o información de inmunogenicidad de los péptidos;

(b) obtener información de secuencia de proteína del HLA de clase I a partir de una base de datos de secuencias biológicas;

(c) entrenar la información obtenida en las etapas (a) y (b) en un modelo que puede predecir la puntuación de afinidad de unión (BA) para HLA para construir un modelo de predicción de afinidad de unión y obtener una puntuación de afinidad de unión; y

(d) entrenar la información obtenida en las etapas (a) y (b) en un modelo que puede predecir la puntuación de inmunogenicidad (IMM) para construir un modelo de predicción de inmunogenicidad y obtener una puntuación de inmunogenicidad.

10 Las bases de datos de péptidos disponibles para su uso en la etapa (I)-(a) pueden incluir una o más bases de datos de péptidos seleccionadas de un grupo que consiste en recurso de análisis y base de datos de epítomos inmunes (IEDB), base de datos exhaustiva de péptidos de unión y no unión a MHC (MHCBN), pero no se limitan a las mismas. Puede usarse sin particular limitación cualquier base de datos que contenga
15 información de péptido, tal como información de secuencia de aminoácidos de péptidos de origen humano (células normales/tejidos normales y células tumorales/tejidos tumorales), e información de unión para HLA (por ejemplo, CI50, valores de KD, etc.) y/o información relacionada con inmunogenicidad.

Las bases de datos de secuencias biológicas disponibles para su uso en la etapa
20 (I)-(b) pueden incluir una o más bases de datos de proteínas seleccionadas de un grupo que consiste en Centro Nacional para la Información Biotecnológica (NCBI), base de datos de proteínas, y atlas de motivos de MHC, pero no se limitan a las mismas. Siempre que proporcione secuencias de aminoácidos de proteínas HLA, para el entrenamiento puede usarse sin particular limitación cualquier base de datos.

25 La información obtenida a partir de la base de datos de péptidos y/o base de datos de secuencias biológicas puede usarse para el entrenamiento de los modelos de predicción en las etapas (c) y (d).

La etapa (II) de predecir y/o seleccionar péptidos inmunogénicos específicos de

tumor puede incluir las siguientes etapas (1) a (7):

(1) obtener información de secuencia de NGS a partir de células tumorales o tejidos tumorales y células normales o tejidos normales derivados de un paciente;

(2) obtener información de secuencia de péptidos específicos de tumor de al menos 7 aminoácidos de longitud que contienen una o más mutaciones de aminoácidos en la secuencia de NGS de células tumorales o tejidos tumorales derivados del paciente en comparación con la secuencia de NGS de células normales o tejidos normales;

(3) medir el nivel de expresión (TPM; transcritos por millón) de los péptidos específicos de tumor o los genes codificantes de los péptidos específicos de tumor;

(4) aplicar la información de secuencia de los péptidos específicos de tumor al modelo de predicción de afinidad de unión construido en la etapa (c) para medir (confirmar, predecir) la puntuación de afinidad de unión (BA) a la molécula de HLA;

(5) aplicar la información de secuencia de los péptidos específicos de tumor al modelo de predicción de inmunogenicidad construido en la etapa (d) para medir (confirmar, predecir) la puntuación de inmunogenicidad (IMM);

(6) determinar la clasificación de prioridad de los péptidos específicos de tumor basándose en la puntuación calculada usando la siguiente fórmula:

$$\text{Puntuación de priorización de neoantígeno} = \min(\text{TPM}, 1) * (W1 * BA + W2 * IMM) / (W1 + W2)$$

[en la que se establecen independientemente $0 < W1 < 1$ (o $0,1 \leq W1 \leq 0,9$) y $0 < W2 < 1$ (o $0,1 \leq W2 \leq 0,9$); “ $\min(\text{TPM}, 1)$ ” significa que el valor se usa como tal cuando $0 < \text{TPM} \leq 1$, y TPM se establece a 1 para el cálculo cuando $1 \leq \text{TPM}$ (esto es para reducir el sesgo en la interpretación de puntuación provocado por valores de TPM mayores de 1, tales como 100 ó 1000)]; y

(7) seleccionar dos o más péptidos inmunogénicos específicos de tumor basándose en la clasificación de prioridad determinada de los péptidos específicos de tumor.

La información de secuencia de secuenciación de nueva generación (NGS) en las

etapas (1) y (2) puede ser información de secuencia de ácido nucleico (gen) y/o información de secuencia convertida en secuencias de aminoácido (proteína). En una realización, la etapa (1) puede incluir además una etapa de convertir la secuencia de NGS obtenida en una secuencia de aminoácidos después de obtener la información de
5 secuencia de NGS.

Las células tumorales o los tejidos tumorales y las células normales o los tejidos normales en las etapas (1) y/o (2) pueden derivarse del mismo paciente, pero sin limitaciones a ello.

La etapa de medición de TPM en la etapa (3) puede realizarse sobre células
10 tumorales y/o tejidos tumorales, y las células tumorales y/o los tejidos tumorales pueden obtenerse de muestras derivadas de paciente o líneas celulares tumorales (cancerosas) disponibles. Si la medición de TPM en la etapa (3) se realiza sobre células tumorales o tejidos tumorales derivados de paciente, las células tumorales o los tejidos tumorales y las células normales o los tejidos normales en las etapas (1) y/o (2) pueden derivarse del
15 mismo paciente, pero sin limitaciones a ello.

La etapa de medición de TPM en la etapa (3), la etapa de medición de puntuación de afinidad de unión en la etapa (4), y la etapa de medición de puntuación de inmunogenicidad en la etapa (5) no están particularmente restringidas en cuanto a su orden de ejecución. Por ejemplo, pueden realizarse de manera secuencial
20 independientemente del orden, o pueden realizarse dos o más etapas de manera simultánea, pero sin limitaciones a ello.

Los dos o más péptidos inmunogénicos específicos de tumor seleccionados pueden usarse como sustancias candidatas finales para los péptidos inmunogénicos específicos de tumor (neoantígenos).

25

En una realización, el método para seleccionar péptidos inmunogénicos específicos de tumor proporcionado en esta descripción puede incluir la etapa (II) descrita previamente de predecir y/o seleccionar péptidos inmunogénicos específicos de tumor

(es decir, las etapas (1) a (7)).

En otra realización, el método para seleccionar péptidos inmunogénicos específicos de tumor puede incluir adicionalmente la etapa (I) de construir un modelo de predicción de péptido inmunogénico (es decir, las etapas (a) a (d)) además de la etapa (II) de predecir y/o seleccionar péptidos inmunogénicos específicos de tumor. En este caso, la etapa (I) adicional de construir un modelo de predicción de péptido inmunogénico puede incluirse antes de la etapa (II) de predecir y/o seleccionar péptidos inmunogénicos específicos de tumor.

Para la base de datos de péptidos usada para construir el modelo de predicción en la etapa (I) del método de selección, pueden usarse datos obtenidos a través de uno o más procedimientos de selección elegidos de los siguientes i) a iv):

- i) eliminar datos que no corresponden a HLA-A, HLA-B, o HLA-C;
- ii) eliminar datos con mutaciones en moléculas de HLA;
- iii) eliminar datos para péptidos con longitudes más cortas de 7 o más largas de 15; y
- iv) retirar secuencias de aminoácidos que contienen residuos de aminoácido distintos de los 20 aminoácidos de origen humano.

Una o más etapas de la etapa (I) (etapas (a) a (d)) o la etapa (II) (etapas (1) a (7)) pueden realizarse mediante un dispositivo de procesamiento computacional, tal como un ordenador, o un sistema de procesamiento de datos.

Las bases de datos de péptidos disponibles para el entrenamiento del modelo de predicción de afinidad de unión en la etapa (I) pueden incluir una o más bases de datos de péptidos seleccionadas de un grupo que consiste en recurso de análisis y base de datos de epítomos inmunes (IEDB), base de datos exhaustiva de péptidos de unión y no unión a MHC (MHCBN), pero no se limitan a las mismas. Para el entrenamiento puede usarse sin particular limitación cualquier base de datos que contenga información de péptido, tal como información de secuencia de aminoácidos de péptidos de origen

humano e información de unión para HLA (por ejemplo, CI50, valores de KD, etc.) y/o información relacionada con inmunogenicidad.

En una realización, la etapa de confirmar (medir, determinar) la puntuación de inmunogenicidad (IMM) puede incluir aprender datos a partir de dos o más modelos de inmunogenicidad relacionados con respuestas inmunitarias cuando un complejo péptido-MHC (HLA) se une a un receptor de células T (TCR).

Los datos de entrenamiento de modelo de inmunogenicidad pueden incluir dos o más tipos de información relacionada con inmunogenicidad para secuencias candidatas de péptidos obtenidas a partir de bases de datos de péptidos o bibliografía relacionada, por ejemplo, seleccionadas del grupo que consiste en niveles de secreción de citocinas derivadas de células T incluyendo INF-g, TNF-a, IL-2, IL-1b, etc., proliferación de células T debida a activación inmunogénica, niveles de quimiocinas incluyendo CCL4 y CXCL9, y valores de citotoxicidad de células T activadas tales como granzima B. Específicamente, las bases de datos de péptidos pueden incluir una o más bases de datos de péptidos seleccionadas de un grupo que consiste en recurso de análisis y base de datos de epítomos inmunes (IEDB) y base de datos exhaustiva de péptidos de unión y no unión a MHC (MHCBN), pero no se limitan a las mismas. Siempre que contenga información de péptido, tal como información de secuencia de aminoácidos de péptidos de origen humano y valores de medición de inmunogenicidad, para el entrenamiento puede usarse sin particular limitación cualquier base de datos.

Los valores de medición de inmunogenicidad pueden medirse mediante métodos incluyendo ELISPOT, ELISA, citometría de flujo, tinción intracelular, ensayos de unión, ensayos de proliferación (por ejemplo, ensayos de proliferación de células T), ensayos de citotoxicidad (por ejemplo, ensayos de citotoxicidad *in vitro/ex vivo/in vivo*), y similares.

Los datos de entrenamiento de modelo de inmunogenicidad pueden incluir valores medidos tanto cualitativamente como cuantitativamente, convirtiéndose los valores cualitativos estadísticamente en datos numéricos, por ejemplo, mediante transformación de distribución beta, en comparación con valores de control negativo.

La etapa (II) puede realizarse usando información de secuencia de NGS obtenida a partir de una muestra (célula/tejido tumoral y célula/tejido normal) derivada de un sujeto personalizado (mamífero, por ejemplo, ser humano) y/o información convertida en información de péptido.

5 El péptido inmunogénico específico de tumor puede ser un péptido derivado de una mutación de célula cancerosa. En una realización, el péptido inmunogénico específico de tumor puede derivarse de una secuencia (por ejemplo, secuencia de ácido nucleico obtenida a partir de NGS o una secuencia de aminoácidos convertida a partir de la misma) obtenida a partir de análisis de secuenciación genómica (por ejemplo, 10 secuenciación de nueva generación (NGS), etc.) de una muestra derivada de paciente (por ejemplo, célula cancerosa). Más específicamente, el péptido inmunogénico específico de tumor puede derivarse de una secuencia (secuencia de ácido nucleico o una secuencia de aminoácidos convertida a partir de la misma) obtenida mediante análisis de secuenciación genómica (por ejemplo, secuenciación de nueva generación (NGS), etc.) 15 de una muestra derivada de paciente (por ejemplo, célula cancerosa), donde la secuencia difiere de la de células normales (por ejemplo, una secuencia con una mutación en comparación con células normales).

La molécula de HLA puede seleccionarse de un grupo que consiste en moléculas de complejo mayor de histocompatibilidad (MHC) derivadas de seres humanos. La 20 molécula de HLA puede incluir diferentes tipos de molécula de HLA y/o diferentes alelos de HLA. En una realización, la molécula de HLA puede ser una molécula de HLA de tipo I.

En la etapa de construir el modelo de predicción de péptido inmunogénico (etapa (I)), la puntuación de afinidad de unión en la etapa (c) puede obtenerse combinando 25 (agrupando) valores predichos entrenados usando dos o más métodos basándose en la información de secuencia de aminoácidos del péptido inmunogénico y la molécula de HLA. La información de secuencia de aminoácidos de la molécula de HLA puede obtenerse a partir de la base de datos de péptidos descrita anteriormente, pero sin

limitaciones a ello.

La puntuación de inmunogenicidad en la etapa (d) puede obtenerse basándose en la cantidad de citocina secretada cuando el péptido inmunogénico se trata con una muestra que contiene células inmunitarias. La muestra que contiene células inmunitarias
5 puede seleccionarse del grupo que consiste en sangre, glóbulos blancos, y células mononucleares de sangre periférica (PBMC), y la citocina puede seleccionarse de uno o más del grupo que consiste en IFN-g, IL-2, TNF-a, y granzima B, pero sin limitaciones a ello. En una realización, la puntuación de inmunogenicidad en la etapa (2) puede determinarse basándose en información de secreción de citocinas cuando el péptido
10 candidato se trata con células inmunitarias (por ejemplo, PBMC, etc.) o una muestra que contiene las mismas. La información de secreción de citocinas puede obtenerse a partir de bases de datos seleccionadas de un grupo que consiste en IEDB, MHCBN, etc., pero sin limitaciones a ello.

La puntuación de inmunogenicidad puede medirse indirectamente como la razón
15 de experimentos positivos con respecto al número total de experimentos. En una realización, la base de datos IEDB proporciona valores de medición cualitativos, recuentos de experimento (sometidos a prueba), recuentos de respuesta (que responden), y frecuencia como datos en lugar de valores de medición cuantitativos para la secreción de citocinas, y estos pueden usarse como datos de entrenamiento para
20 cuantificar la puntuación de inmunogenicidad. Los valores experimentales reales pueden convertirse en datos numéricos reflejando la razón de recuentos de respuesta con respecto a recuentos de experimento y pueden usarse para el reentrenamiento del modelo de predicción de inmunogenicidad. Esta cuantificación puede realizarse usando uno o más métodos seleccionados del grupo que consiste en media, mediana, y
25 transformación de distribución beta.

La puntuación de inmunogenicidad en la etapa (d) puede obtenerse combinando (agrupando) valores predichos entrenados usando dos o más métodos basándose en la información de secuencia de aminoácidos de péptido a partir de los datos de

entrenamiento de inmunogenicidad, la secuencia de aminoácidos correspondiente a la molécula de HLA de las células inmunitarias, y la información de secreción de citocinas cuando el péptido se trata con células inmunitarias.

La combinación (agrupación) de etapas (c) y (d) puede implicar seleccionar dos o más modelos de predicción, entrenados con métodos diferentes, que tienen el error de predicción más pequeño para conjuntos de datos de validación diferentes de los datos de entrenamiento, y combinando sus valores predichos. Específicamente, el conjunto de datos de validación puede formar parte de los datos de entrenamiento de modelo apartados de antemano.

En la etapa (c) (modelado de predicción de unión y medición de puntuación) y/o la etapa (d) (modelado de predicción de inmunogenicidad y medición de puntuación):

- el entrenamiento puede realizarse usando uno o más modelos seleccionados de modelos de aprendizaje por inteligencia artificial (IA) usados comúnmente, tales como unidad recurrente cerrada (GRU), transformador, red neuronal de prealimentación (FFN), red neuronal gráfica (GNN), etc.;

- el entrenamiento de la etapa (c) (medición de puntuación de afinidad de unión) y la etapa (d) (medición de puntuación de inmunogenicidad) puede realizarse combinando un método de aprendizaje seleccionado de los usados comúnmente en el campo, o bien aplicado o bien no aplicado, con el modelo de IA; y

- la puntuación de afinidad de unión y/o la puntuación de inmunogenicidad pueden ser el valor predicho final obtenido combinando (agrupando) los valores predichos a partir de cada modelo de predicción entrenado usando uno o más modelos de IA y uno o más métodos de aprendizaje con un modelo de agrupación convencional, y en una realización, el valor predicho final puede obtenerse como [(media de valores predichos) - (varianza de valores predichos)].

En la etapa (II) de predecir y/o seleccionar péptidos inmunogénicos específicos de tumor, la etapa (1) puede realizarse a través de un procedimiento de análisis

bioinformático que compara las secuencias de NGS de células tumorales o tejidos tumorales derivados de un paciente con las secuencias de NGS de células normales o tejidos normales derivados del mismo paciente. Las células tumorales o los tejidos tumorales y las células normales o los tejidos normales pueden originarse a partir del mismo paciente. El procedimiento de análisis bioinformático puede incluir, pero no se limita a, los siguientes procedimientos (i) a (vii), y pueden añadirse o retirarse etapas adicionales según sea necesario:

- (i) verificar la integridad de datos de NGS; por ejemplo, puede usarse software tal como FastQC, pero sin limitaciones a ello;
- 10 (ii) retirar datos de baja calidad; por ejemplo, puede usarse software tal como Cutadapt, pero sin limitaciones a ello;
- (iii) alineación de secuencias con respecto al genoma de referencia; por ejemplo, pueden usarse el genoma de referencia de hg19 y el software bwa, pero sin limitaciones a ello;
- 15 (iv) verificar y corregir la calidad de los resultados de alineación de secuencias; por ejemplo, puede usarse software tal como GATK para aplicar realineación de indel y recalibración de puntuación de calidad de base (BQSR), pero sin limitaciones a ello;
- (v) identificar mutaciones específicas de tumor; por ejemplo, puede usarse software tal como GATK, Strelka, Mutect2, etc., pero sin limitaciones a ello;
- 20 (vi) clasificar mutaciones específicas de tumor; por ejemplo, las mutaciones pueden clasificarse usando software tal como SnpEff, y pueden seleccionarse únicamente mutaciones de aminoácido, de inserción, y de delección que afecten a la secuencia de péptido, pero sin limitaciones a ello;
- (vii) extraer secuencias de péptido que tienen 7 o más aminoácidos de longitud
- 25 incluyendo las mutaciones específicas de tumor y pueden generarse.

TPM (transcritos por millón) en la etapa (3) es una medida de expresión génica durante la transcripción, el procedimiento donde se transcribe una secuencia génica

compuesta por nucleótidos para formar transcritos. Los métodos para medir valores de expresión incluyen medir la concentración de ARNm y medir la concentración de transcrito, pero no se limitan a los mismos, y los valores de expresión pueden medirse de diversas maneras. En una realización de la presente divulgación, el valor de expresión es el número de lecturas obtenidas a partir de los resultados de NGS mapeadas al transcriptoma, y puede calcularse en unidades de TPM, pero sin limitaciones a ello, y puede expresarse en otras unidades para indicar el valor de expresión. El método puede incluir además una etapa de filtrar secuencias de ARN con TPM (transcritos por millón) de 1 o superior de entre las secuencias de nucleótidos mutadas o toda la secuencia de ARN.

El nivel de expresión del péptido específico de tumor o un gen codificante para el mismo en la etapa (3) (por ejemplo, TPM, transcritos por millón; Li *et al.* (2010); se normaliza la proporción de cada transcrito (recuento de lecturas de NGS mapeadas) dentro de la biblioteca, y la suma de los valores de TPM en una biblioteca equivale a 1.000.000, lo que permite la comparación entre muestras diferentes) puede indicar la probabilidad de que el péptido específico de tumor esté presente en la muestra de paciente real (por ejemplo, probabilidad, frecuencia, o grado de que la mutación génica de célula cancerosa se transcriba y traduzca en una proteína). El nivel de expresión del péptido específico de tumor o su gen codificante en la etapa (3) puede obtenerse a través de análisis de secuencia de datos de muestra a partir del paciente sometido a tratamiento personalizado, y más específicamente, puede calcularse procesando datos de secuenciación de ARN o micromatriz. En una realización, TPM del péptido específico de tumor o un gen codificante para el mismo puede calcularse analizando los datos de secuenciación de ARN del paciente, y puede darse prioridad a los péptidos específicos de tumor con un valor de TPM de 1 o superior. Por ejemplo, dado que TPM representa el nivel de expresión por millón de unidades de datos de secuenciación de ARN normalizados por longitud del gen y salida de datos, usar $TPM \geq 1$ para la priorización da preferencia a péptidos candidatos que ocupan al menos la millonésima parte de los datos

de secuenciación de ARN totales, priorizando de ese modo aquellos que es más probable que estén presentes en la muestra de paciente real.

La etapa de priorización en la etapa (6) se realiza usando la siguiente fórmula:

$$\text{Puntuación de priorización de neoantígeno} = \min(\text{TPM}, 1) * (W1 * \text{BA} + W2 * \text{IMM}) / (W1 + W2).$$

en la que, W1 y W2 representan las ponderaciones de la puntuación de afinidad de unión (BA) y la puntuación de inmunogenicidad (IMM), respectivamente, y pueden establecerse independientemente dentro del intervalo de $0 < W1 < 1$ (o $0,1 \leq W1 \leq 0,9$) y $0 < W2 < 1$ (o $0,1 \leq W2 \leq 0,9$), respectivamente. En una realización, W1+W2 puede equivaler a 1, pero sin limitaciones a ello.

En una realización, la fórmula puede aplicarse a péptidos candidatos que satisfacen las condiciones de $\text{TPM} > 1$, $0,6 > \text{puntuación de afinidad de unión (BA)} \geq 0,1$, y puntuación de inmunogenicidad (IMM) $\geq 0,25$, pero sin limitaciones a ello.

La etapa de selección de péptidos inmunogénicos específicos de tumor (neoantígenos) en la etapa (7) puede incluir además seleccionar uno o más péptidos inmunogénicos específicos de tumor a partir de los 2 a 100, 2 a 75, 2 a 50, 2 a 40, 2 a 30, 2 a 20, 2 a 10, 5 a 100, 5 a 75, 5 a 50, 5 a 40, 5 a 30, 5 a 20, 5 a 10, 10 a 100, 10 a 75, 10 a 50, 10 a 40, 10 a 30, 10 a 20, 20 a 100, 20 a 75, 20 a 50, 20 a 40, o 20 a 30 mejores péptidos, dispuestos en orden descendente basándose en los resultados de la etapa (6).

En casos donde la predicción de neoantígeno (péptido inmunogénico específico de tumor) y la vacuna contra el cáncer personalizada descritas en la presente divulgación se producen como ARNm, puede que no sea necesaria ninguna consideración específica de la hidrofobia. Sin embargo, cuando la vacuna contra el cáncer se produce en forma de péptido o se realizan pruebas de inmunogenicidad, si la hidrofobia del péptido es alta (por ejemplo, un alto contenido de aminoácidos hidrófobos), puede ser imposible producir el péptido, por lo que es necesario medir la hidrofobia del péptido. En una realización, tal

hidrofobia puede medirse calculando la puntuación de Gravy (promedio global de hidropatía), pero sin limitaciones a ello.

Los péptidos inmunogénicos específicos de tumor seleccionados a través del
5 método para seleccionar un péptido inmunogénico específico de tumor, proporcionado
en la presente divulgación, son péptidos (por ejemplo, péptidos mutantes) que están
presentes específicamente en muestras cancerosas (tumores) y ausentes en muestras
normales sin cáncer (tumor). Por tanto, los péptidos inmunogénicos específicos de tumor
pueden usarse eficazmente en composiciones farmacéuticas para la prevención y/o el
10 tratamiento de cáncer, tal como en la preparación de vacunas contra el cáncer,
particularmente en la producción de vacunas contra el cáncer personalizadas para
pacientes basándose en el método de selección.

Otro aspecto proporciona una composición de vacuna antineoplásica que incluye
péptidos inmunogénicos específicos de tumor (neoantígenos) seleccionados mediante el
15 método de selección de péptido inmunogénico asociado a tumor, y/o genes codificantes
para los mismos (ARNm, ADN, etc.). La composición de vacuna antineoplásica puede
incluir además un portador y/o excipiente farmacéuticamente aceptables y/o uno o más
adyuvantes, estabilizadores, y similares.

Otro aspecto proporciona un método para preparar una vacuna contra el cáncer,
20 incluyendo el método una etapa de mezclar el péptido inmunogénico específico de tumor
seleccionado mediante el método de selección de péptido inmunogénico específico de
tumor y/o el gen codificante para el mismo (ARNm, ADN, etc.) con un portador
farmacéuticamente aceptable y/o uno o más adyuvantes, estabilizadores, y similares.

El portador y/o excipiente y/o uno o más adyuvantes, estabilizadores, y similares
25 pueden seleccionarse de cualquier sustancia farmacéuticamente aceptable usada
convencionalmente en la formulación de vacunas.

La vacuna puede estar en forma de una vacuna terapéutica o profiláctica.

Otro aspecto proporciona un método para prevenir y/o tratar un cáncer, incluyendo

el método una etapa de administrar el péptido inmunogénico específico de tumor seleccionado mediante el método de selección de péptido inmunogénico específico de tumor o la vacuna a un paciente que necesita prevención y/o tratamiento de cáncer.

El tratamiento de cáncer puede combinarse con otros agentes antineoplásicos y/o resección quirúrgica y/o radioterapia y/o quimioterapia tradicional. Los otros agentes antineoplásicos pueden seleccionarse de cualquier agente antineoplásico conocido (tal como fármacos químicos, anticuerpos, etc.), por ejemplo, inhibidores de punto de control inmunitario (tales como inhibidores de PD-1 y/o inhibidores de PD-L1), pero sin limitaciones a ello.

10

【Solución técnica】

Para descubrir neoantígenos con fuerte inmunogenicidad entre las mutaciones de secuencia específicas de cáncer, se desarrolló un modelo de predicción de neoantígeno impulsado por IA combinando un modelo de afinidad de unión, que predice la afinidad de unión entre HLA y la secuencia de neoantígeno, con un modelo de inmunogenicidad que predice la inmunogenicidad de la secuencia de neoantígeno. El modelo de predicción de neoantígeno se entrenó usando recurso de análisis y base de datos de epítomos inmunes (IEDB) y se mejoró a través de actualizaciones *in silico* con datos a partir de la base de datos.

20

A través del entrenamiento del modelo de predicción de neoantígeno por IA, fue posible predecir neoantígenos candidatos con potencial inmunogénico. Aplicando este modelo de predicción a datos de NGS de pacientes con cáncer, se identificaron neoantígenos candidatos. Se estableció un umbral para los candidatos a neoantígeno predichos por los modelos de afinidad de unión e inmunogenicidad durante la aplicación del modelo de predicción. Luego se sintetizaron péptidos correspondientes a los neoantígenos predichos para los ensayos de afinidad de unión y los ensayos de inmunogenicidad (por ejemplo, ELISPOT de IFN γ) para verificar la precisión de predicción

de neoantígeno del modelo y seleccionar neoantígenos aplicables a vacunas contra el cáncer personalizadas.

La estructura de un aspecto de la presente solicitud puede ejemplificarse tal como sigue:

Modelo de predicción de neoantígeno (modelo de afinidad de unión + modelo de inmunogenicidad basados en transformador);

Selección de datos de entrada para el modelo de predicción de neoantígeno;

Método de entrenamiento para el modelo de predicción de neoantígeno;

Método de procesamiento para los conjuntos de datos de paciente, incluyendo datos de WES (secuenciación de exosoma completo; selección de mutación) y de secuenciación de ARN (nivel de expresión);

Priorización de neoantígeno (usando TPM, puntuación de BA, y puntuación de IMM para confirmar la puntuación de clasificación de neoantígeno);

Verificación de la aplicabilidad de los neoantígenos predichos a vacunas contra el cáncer personalizadas a través de ensayos de afinidad de unión e inmunogenicidad.

Definición de términos

Tal como se usa en el presente documento, el término “péptido” se refiere a una sustancia que contiene dos o más, tres o más, cuatro o más, seis o más, ocho o más, diez o más, 13 o más, 16 o más, 21 o más, y hasta 8, 10, 20, 30, 40, 50, ó 100 aminoácidos unidos de manera covalente por enlaces peptídicos. Los términos “polipéptido” o “proteína” pueden referirse a sustancias con más residuos de aminoácido unidos de manera covalente por enlaces peptídicos que en los péptidos. Sin embargo, los términos “péptido”, “polipéptido”, y “proteína”, tal como se usan en el presente documento, pueden usarse de manera intercambiable.

Tal como se usa en el presente documento, el término “modificación” o “mutación” en relación con un péptido, un polipéptido, o una proteína se refiere a una secuencia que

se ha modificado a partir de su secuencia original (parental), implicando una inserción, adición, deleción, o sustitución de aminoácido. Por ejemplo, la modificación puede ser una sustitución de aminoácido. En esta memoria descriptiva, cualquier cambio de secuencia puede crear potencialmente un nuevo epítopo.

5 Tal como se usa en el presente documento, el término “péptido inmunogénico” se refiere a un péptido que puede inducir una respuesta inmunitaria, incluyendo un epítopo.

 El término “neoantígeno”, tal como se usa en el presente documento, se refiere a un péptido o una proteína que incluye una o más modificaciones de aminoácidos en comparación con el péptido o la proteína parental del mismo. Estas modificaciones de aminoácidos pueden ser el resultado de mutaciones específicas de tumor. En esta memoria descriptiva, el neoantígeno es un tipo de péptido inmunogénico que incluye mutaciones específicas de cáncer y puede no estar presente ni ser detectable en células normales. El neoantígeno también puede denominarse neoantígeno específico de tumor, antígeno específico de tumor, o péptido inmunogénico específico de tumor.

15 El término “respuesta inmunitaria” se refiere a la respuesta integrada en el cuerpo a una diana, tal como un antígeno, y puede implicar inmunidad celular o respuestas inmunitarias tanto celulares como humorales. La respuesta inmunitaria puede ser protectora, preventiva, o terapéutica.

 El término “inducir una respuesta inmunitaria” puede referirse a la ausencia de una respuesta inmunitaria antes de la inducción o puede significar que existía una respuesta inmunitaria inicial pero que se potencia tras la inducción. Por tanto, “inducir una respuesta inmunitaria” también incluye “potenciar una respuesta inmunitaria”. Por ejemplo, después de la inducción de respuesta inmunitaria, el sujeto puede protegerse de enfermedades tales como cáncer o el estado patológico puede mejorarse, aliviarse, o mitigarse por la respuesta inmunitaria inducida.

25 Los términos “respuesta inmunitaria celular” y “respuesta celular” se refieren a respuestas inmunitarias dirigidas por células que se dirigen por células caracterizadas por presentación de antígeno a través de moléculas de MHC de clase I o clase II junto con

células T o linfocitos T que actúan como “cooperadores” o “supresores”. Las células T cooperadoras (también denominadas células T CD4+) desempeñan un papel central en la regulación de respuestas inmunitarias, mientras que las células supresoras (células T CD8+, linfocitos T citotóxicos, o CTL) destruyen las células enfermas, tales como células cancerosas, previniendo la formación de poblaciones de células enfermas. En una realización preferida, la presente divulgación puede implicar estimular una respuesta de CTL antitumoral frente a células tumorales que expresan uno o más antígenos expresados en tumor, y/o que presentan estos antígenos expresados en tumor a través de HLA de tipo I o MHC de clase I.

El término “antígeno” abarca cualquier sustancia que provoca una respuesta inmunitaria, tal como un péptido o una proteína, al interaccionar con un anticuerpo o linfocitos T (células T). Por ejemplo, un antígeno incluye al menos un epítipo, tal como un epítipo de células T. En el presente documento, los antígenos pueden ser moléculas que inducen una respuesta inmunitaria específica tras el procesamiento (incluyendo las células que expresan el antígeno). Un antígeno o su epítipo de células T puede presentarse por células presentadoras de antígeno, incluyendo células enfermas tales como células cancerosas, a través de moléculas de MHC, lo que conduce a una respuesta inmunitaria frente al antígeno (incluyendo las células que expresan el antígeno).

Tal como se usa en el presente documento, los términos “antígeno específico de tumor”, “antígeno expresado específicamente en tumor”, “antígeno específico de cáncer”, y “antígeno expresado específicamente en cáncer” son equivalentes en cuanto a significado y pueden usarse de manera intercambiable con neoantígeno.

El término “inmunogenicidad” se refiere a la capacidad de inducir una respuesta inmunitaria, tal como en el tratamiento contra el cáncer (por ejemplo, formación de anticuerpos).

El término “complejo mayor de histocompatibilidad” (MHC) se refiere a moléculas de MHC de clase I y clase II y está asociado con el complejo génico que se produce en todos los vertebrados. Las proteínas o moléculas de MHC son críticas para la señalización

entre linfocitos y células presentadoras de antígeno o células enfermas durante las respuestas inmunitarias. Las proteínas o moléculas de MHC se unen a péptidos y los presentan para su reconocimiento por receptores de células T. Las proteínas codificadas por MHC se expresan en la superficie celular y presentan tanto autoantígenos (fragmentos peptídicos de la propia célula) como no autoantígenos (por ejemplo, fragmentos de microorganismos invasores) a las células T.

La región de MHC se divide en tres subgrupos: clase I, clase II, y clase III. Las proteínas MHC de clase I contienen una cadena α y microglobulina β_2 (codificada por el cromosoma 15 y que no forma parte de MHC) y presentan fragmentos de antígeno a las células T citotóxicas. Las proteínas MHC de clase II, encontradas principalmente en células inmunitarias, incluyendo células presentadoras de antígeno, contienen cadenas α y β y presentan fragmentos de antígeno a células T cooperadoras. La región de clase III codifica para otros componentes inmunitarios, tales como citocinas y proteínas del complemento.

MHC es poligénico (existen varios genes de MHC de clase I y MHC de clase II) y polimórfico (cada gen tiene múltiples alelos).

Tal como se usa en el presente documento, el término “haplotipo” se refiere a los alelos de HLA encontrados en un cromosoma y proteínas codificados. Un haplotipo también puede referirse a cualquier alelo presente en cualquier locus dentro del MHC.

Cada clase de MHC puede estar representada por varios locus: por ejemplo,
- para la clase I, antígeno leucocitario humano (HLA)-A, HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, HLA-H, HLA-J, HLA-K, HLA-L, HLA-P, HLA-V, etc.; y
- para la clase II, HLA-DR (HLA-DRA, HLA-DRB1-9, etc.), HLA-DQ (HLA-DQA1, HLA-DQB1, etc.), HLA-DP (HLA-DPA1, HLA-DPB1, etc.), HLA-DM (HLA-DMA, HLA-DMB, etc.), HLA-DO (HLA-DOA, HLA-DOB, etc.).

En el presente documento, los términos “HLA o alelo de HLA” y “MHC o alelo de MHC” se usan de manera intercambiable.

Los MHC presentan un fuerte polimorfismo. Existen un gran número de haplotipos

que contienen, cada uno, alelos distintos en cada locus genético dentro de la población humana. Los diferentes alelos de MHC polimórficos tanto de la clase I como de la clase II tienen especificidades de péptido variables. Cada alelo codifica para una proteína que se une a péptidos con patrones de secuencia específicos.

5 En una realización, las moléculas de MHC pueden ser moléculas de HLA. Las moléculas de HLA pueden indicarse por su tipo serológico según “HLA Dictionary 2008” (John Wiley & Sons A/S; Tissue Antigens 73, 95–170, incorporado mediante referencia en el presente documento).

 El término “péptido de unión a MHC” se refiere a un péptido que puede procesarse
10 para producir un péptido de unión a MHC de clase I y/o clase II. Para los complejos MHC de clase I/péptido, el péptido de unión tiene generalmente 7-12, 7-10, 8-12, u 8-10 aminoácidos de longitud, pero sin limitaciones a ello. Para los complejos MHC de clase II/péptido, el péptido de unión tiene generalmente 9-30 ó 10-25 aminoácidos de longitud, tal como 13-18 aminoácidos, pero sin limitaciones a ello.

15 En una realización, cuando un péptido se presenta directamente, sin procesamiento, especialmente sin escisión, la longitud puede ser adecuada para la unión a una molécula de MHC, particularmente una molécula de MHC de clase I, tal como una longitud de péptido de 7-30, 7-20, 7-12, 8-11, 9, ó 10 aminoácidos. En otra realización, cuando el péptido forma parte de una entidad más grande, tal como un polipéptido o una
20 secuencia de vacuna, y se presenta después de procesamiento, especialmente escisión, la longitud del péptido producido por procesamiento puede ser adecuada para la unión a una molécula de MHC, particularmente una molécula de MHC de clase I, tal como 7-30, 7-20, 7-12, 8-11, 9, ó 10 aminoácidos.

 Los péptidos inmunogénicos específicos de tumor seleccionados mediante los
25 métodos proporcionados en el presente documento pueden unirse a moléculas de MHC de clase I (HLA).

 El término “epítipo” se refiere a un determinante antigénico dentro de una molécula tal como un antígeno, es decir, una parte o un fragmento de la molécula

reconocida por el sistema inmunitario. El epítipo de una proteína, tal como un antígeno tumoral, puede incluir una porción continua o discontinua de la proteína, y su longitud puede ser de 5 a 100, de 5 a 50, de 8 a 30, o de 10 a 25 aminoácidos. En el presente documento, el epítipo puede ser un epítipo de células T.

5 En el presente documento, el epítipo puede ser un “péptido de unión a MHC” que se une a moléculas de MHC en la superficie celular.

 El término “neoepítipo” se refiere a un epítipo encontrado en células cancerosas pero que no está presente en células de referencia, tales como células no cancerosas normales o células germinales. Esto incluye casos donde existe un epítipo
10 correspondiente en células normales, pero las mutaciones en las células cancerosas alteran la secuencia, creando un neoepítipo.

 El término “epítipo de células T”, tal como se usa en el presente documento, se refiere a un péptido que se une a moléculas de MHC en una configuración reconocida por receptores de células T. Los epítipos de células T residen generalmente en la superficie
15 de células presentadoras de antígeno.

 Un epítipo de células T puede incluir una secuencia de aminoácidos que corresponde sustancialmente a la secuencia de aminoácidos de un fragmento de antígeno. Por ejemplo, el fragmento de antígeno puede ser un péptido presentado por moléculas de MHC de clase I y/o clase II.

20 El epítipo de células T según la presente divulgación puede estar relacionado con una porción o un fragmento de un antígeno que puede estimular una respuesta celular, tal como una respuesta inmunitaria a un antígeno o una célula caracterizada por la presentación del antígeno, preferiblemente una célula relacionada con enfermedad, particularmente una célula cancerosa. Por ejemplo, el epítipo de células T puede
25 estimular una respuesta celular a células caracterizada por la presentación de antígeno mediante MHC de clase I, y puede estimular linfocitos T citotóxicos (CTL) reactivos al antígeno.

 El término “procesamiento de antígeno” o “procesamiento” se refiere a un

procedimiento inmunológico que prepara los antígenos para su presentación a través de células, es decir, células presentadoras de antígeno, a células T específicas descomponiendo péptidos, polipéptidos, o proteínas (por ejemplo, se descompone un polipéptido para dar péptidos) y luego asociando los fragmentos resultantes (o partes de ellos) con moléculas de MHC.

Los términos “célula T” y “linfocito T” se usan de manera intercambiable en el presente documento e incluyen células T citotóxicas (CTL, células T CD8+), que son células T de lisis celular, y células T cooperadoras (células T CD4+).

Las células T son un tipo de glóbulo blanco en el grupo de linfocitos y desempeñan un papel central en la inmunidad mediada por células. Se distinguen de otros linfocitos, tales como células B y células citolíticas naturales, por la presencia de un receptor específico en su superficie denominado receptor de células T (TCR). El timo es el órgano principal responsable de la maduración de células T. Se han identificado varios subconjuntos de células T diferentes, cada uno con funciones únicas.

Las células T cooperadoras ayudan a otros glóbulos blancos en procesos inmunológicos, tales como la maduración de células B en células plasmáticas y la activación de células T citotóxicas y macrófagos. Estas células también se denominan células T CD4+ porque expresan la proteína CD4 en su superficie. Las células T cooperadoras se activan cuando se presentan con antígenos peptídicos por moléculas de MHC de clase II en la superficie de células presentadoras de antígeno (APC). Una vez activadas, se dividen rápidamente y secretan proteínas pequeñas denominadas citocinas, que regulan o ayudan a las respuestas inmunitarias activas.

Las células T citotóxicas destruyen las células infectadas con virus y las células tumorales y también están implicadas en el rechazo de trasplante. Estas células también se conocen como células T CD8+ porque expresan la glicoproteína CD8 en su superficie. Reconocen sus dianas mediante la unión a antígenos asociados con MHC de clase I, que está presente en la superficie de casi todas las células en el cuerpo.

La mayoría de las células T poseen un receptor de células T (TCR) que existe

como un complejo de varias proteínas. El propio TCR se forma a partir de genes de TCR α y TCR β independientes e incluye dos cadenas peptídicas distintas denominadas cadenas α y β del TCR. Las células T $\gamma\delta$ representan un pequeño subconjunto de células T que expresan un TCR distinto en su superficie. En las células T $\gamma\delta$, el TCR está compuesto
5 por una cadena γ y una cadena δ . Este grupo de células T es mucho menos frecuente que las células T $\alpha\beta$ (que constituyen aproximadamente el 2 % de las células T totales).

La primera señal para la activación de células T se proporciona cuando el receptor de células T se une a un péptido corto presentado por una molécula de MHC en otra célula. Esto garantiza que sólo se activen células T con un TCR específico para ese
10 péptido. La célula compañera es normalmente una célula presentadora de antígeno (APC), tal como células presentadoras de antígeno profesionales como las células dendríticas en el caso de respuestas indiferenciadas, pero las células B y los macrófagos también son APC importantes.

En el presente documento, se considera que una molécula puede unirse a una
15 diana dada si tiene una afinidad significativa por esa diana y se une a ella en un ensayo convencional. La “afinidad” o “afinidad de unión” puede medirse mediante la constante de disociación en equilibrio (KD). Si una molécula no tiene afinidad significativa por la diana o no se une a la diana de manera significativa en un ensayo convencional, entonces la molécula no se une (sustancialmente) a la diana.

20 La activación específica de células T CD4+ o CD8+ puede confirmarse de diversas maneras. Los métodos para confirmar la activación específica de células T incluyen detectar la proliferación de células T, la producción de citocinas (por ejemplo, linfocinas), o la generación de actividad citolítica. En el caso de células T CD4+, la activación específica de células T puede confirmarse a través de la proliferación de células T. Para
25 las células T CD8+, la activación específica de células T puede confirmarse a través de la actividad citolítica.

Tal como se usa en el presente documento, el término “puntuación” se refiere generalmente al resultado de cualquier prueba o ensayo, normalmente expresado de

manera numérica.

En el presente documento, determinar la puntuación de afinidad de unión de un péptido dado a una molécula de MHC implica detectar la potencial afinidad de unión del péptido a la molécula de MHC.

5 La puntuación de afinidad de unión del péptido a la molécula de MHC puede determinarse usando herramientas de predicción de unión de péptido. Por ejemplo, puede usarse recurso de análisis y base de datos de epítomos inmunes (IEDB-AR: <http://tools.iedb.org>).

Según la presente divulgación, las modificaciones de aminoácidos que van a
10 detectarse por su inmunogenicidad pueden surgir a partir de mutaciones en el ácido nucleico celular. Tales mutaciones también pueden identificarse usando técnicas de secuenciación conocidas.

En una realización, pueden detectarse mutaciones somáticas específicas de cáncer o variaciones de secuencia en el exoma de una muestra tumoral, por ejemplo, en
15 todo el exoma. Por tanto, la presente divulgación puede incluir identificar los signos de mutaciones cancerosas en el exoma, preferiblemente en todo el exoma, de una o más células cancerosas. En una realización, la etapa de identificar una mutación somática específica de cáncer en una muestra tumoral de un paciente con cáncer incluye identificar un perfil de mutaciones cancerosas en todo el exoma.

20 En una realización, pueden detectarse mutaciones somáticas específicas de cáncer o variaciones de secuencia en el transcriptoma de una muestra tumoral, por ejemplo, en todo el transcriptoma. Por tanto, la presente divulgación puede incluir identificar un signo de una mutación cancerosa en el transcriptoma, tal como en todo el transcriptoma, de una o más células cancerosas. En una realización, la etapa de
25 identificar una mutación somática específica de cáncer en una muestra tumoral de un paciente con cáncer incluye identificar un perfil de mutaciones cancerosas en todo el transcriptoma.

En una realización, la etapa de identificar una mutación somática específica de

cáncer o una variación de secuencia incluye la secuenciación de una sola célula de una o más células cancerosas, por ejemplo, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o más células cancerosas. Por tanto, la presente divulgación puede incluir identificar una mutación cancerosa en una o más células cancerosas. En una realización, la célula cancerosa es una célula tumoral circulante. Las células cancerosas, tales como células tumorales circulantes, pueden aislarse antes de realizar la secuenciación de una sola célula.

En una realización, la etapa de identificar una mutación somática específica de cáncer o una variación de secuencia puede llevarse a cabo usando secuenciación de nueva generación (NGS).

En otra realización, la etapa de identificar una mutación somática específica de cáncer o una variación de secuencia puede incluir secuenciar un ARN y/o ADN genómico de una muestra tumoral.

Para revelar mutaciones somáticas específicas de cáncer o variaciones de secuencia, es preferible comparar la información de secuencia obtenida a partir de una muestra tumoral con información de secuencia de referencia, tal como secuencias de ácido nucleico de células no cancerosas normales, incluyendo células de la línea germinal, que pueden obtenerse a partir del mismo paciente u otros individuos. En una realización, puede obtenerse ADN genómico normal de línea germinal a partir de células mononucleares de sangre periférica (PBMC).

El término “genoma” se refiere a pertenecer a una cantidad total de información genética contenida dentro de los cromosomas de un organismo o una célula.

El término “exoma” se refiere a la porción de un genoma que consiste en exones, que son las secciones codificantes de los genes expresados en un organismo. El exoma proporciona el prototipo genético usado para la síntesis de proteínas y otros productos génicos funcionales. Dado que el exoma representa la parte más funcionalmente relevante del genoma, es probable que contribuya significativamente al fenotipo de un organismo. Se estima que el exoma del genoma humano representa el 1,5 % del genoma

total.

El término “transcriptoma” se refiere al conjunto completo de moléculas de ARN, incluyendo ARNm, ARNr, ARNt, y otros ARN no codificantes, producido en una sola célula o un grupo de células. En el presente documento, el transcriptoma puede referirse
5 al conjunto completo de moléculas de ARN producido por una sola célula, un grupo de células, tal como un grupo de células cancerosas, o todas las células de un organismo en un momento dado.

Tal como se usa en el presente documento, el término “ácido nucleico” puede referirse a ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). Los ácidos
10 nucleicos incluyen ADN genómico, ADNc, ARNm, moléculas recombinantes, y moléculas sintetizadas químicamente según la presente divulgación. Los ácidos nucleicos pueden existir en formas monocatenarias o bicatenarias y como moléculas lineales o modificadas de manera covalente.

El término “mutación” se refiere a un cambio o una variación en una secuencia de
15 ácido nucleico (sustitución, adición, o delección de nucleótido) en comparación con una referencia, y/o un cambio en la secuencia de aminoácidos de un péptido (proteína) codificado por esa secuencia de ácido nucleico. Una “mutación somática” se produce en cualquier célula del cuerpo excepto en células germinales (espermatozoides y óvulos) y, por tanto, no se transmite a la descendencia. Tales cambios conducen a menudo (pero
20 no siempre) a cáncer u otras enfermedades. Preferiblemente, las mutaciones son no sinónimas. El término “mutación no sinónima” se refiere a una mutación que da como resultado un cambio de aminoácido, tal como una sustitución de aminoácido, preferiblemente debido a una sustitución de nucleótido.

Tal como se usa en el presente documento, el término “mutación” puede incluir
25 mutaciones puntuales, indels, fusiones, cromotripsis, edición de bases, y/o ediciones de ARN.

El término “firma de mutación cancerosa” se refiere a un conjunto de mutaciones presente en células cancerosas en comparación con células de referencia no cancerosas.

El término “referencia” puede usarse para relacionar o comparar los resultados obtenidos usando los métodos de la presente divulgación con los de las muestras tumorales. Generalmente, “referencia” puede basarse en una o más muestras normales, por ejemplo, muestras de individuos de la misma especie sin cáncer.

5 Siempre que sea adecuado para detectar mutaciones, puede emplearse cualquier método de secuenciación, incluyendo, por ejemplo, secuenciación de nueva generación (NGS). Para acelerar la etapa de secuenciación, las futuras generaciones de tecnologías de secuenciación, tales como la denominada secuenciación de tercera generación, pueden reemplazar a NGS. El término “secuenciación de nueva generación” o “NGS” se
10 refiere a todas las nuevas tecnologías de secuenciación de alto rendimiento, a diferencia de los métodos de secuenciación “convencionales” tales como química de Sanger, que lee moldes de ácido nucleico al azar en todo el genoma mediante la fragmentación del genoma en fragmentos más pequeños. Estas tecnologías de NGS (también denominadas tecnologías de secuenciación paralela masiva) pueden suministrar
15 información de secuencia de ácido nucleico para los genomas, exomas, transcriptomas (todas las secuencias transcritas del genoma), o metilomas (todas las secuencias metiladas del genoma) completos en periodos de tiempo muy cortos, por ejemplo, en el plazo de 1-2 semanas, 1-7 días, o incluso 24 horas, y permiten enfoques de secuenciación de una sola célula. Pueden usarse diversas plataformas de NGS
20 disponibles comercialmente o a las que se hace referencia en la bibliografía.

El término “ARN” se refiere a una molécula que contiene al menos un residuo de ribonucleótido y preferiblemente consiste únicamente o sustancialmente en residuos de ribonucleótido. Un “ribonucleótido” se refiere a un nucleótido que tiene un grupo hidroxilo en la posición 2’ del resto β -D-ribofuranosilo. El término “ARN” abarca ARN bicatenario,
25 ARN monocatenario, ARN aislado (por ejemplo, ARN purificado de manera parcial o completa), ARN esencialmente puro, ARN sintético, y ARN producido de manera recombinante, que pueden diferir del ARN que se produce de manera natural a través de la adición, delección, sustitución, y/o modificación de uno o más nucleótidos. Estas

modificaciones pueden incluir la adición de sustancias distintas de nucleótidos en uno o más extremos terminales o de manera interna dentro del ARN, y los nucleótidos en la molécula de ARN pueden incluir nucleótidos no convencionales tales como desoxinucleótidos o nucleótidos sintetizados químicamente. Estos ARN modificados
5 pueden denominarse análogos o miméticos de ARN natural.

El término “ARN” puede incluir “ARNm”. El término “ARNm” se refiere a “ARN mensajero” que se genera usando un molde de ADN y codifica para péptidos o polipéptidos. También se conoce como “transcrito”. Normalmente, el ARNm incluye una 5'-UTR, una región codificante de proteína, y una 3'-UTR. El ARNm tiene una semivida
10 limitada tanto en células como *in vitro*. El ARNm puede generarse a través de transcripción *in vitro* a partir de un molde de ADN, y tales métodos de transcripción *in vitro* son conocidos por los expertos en la técnica.

Los términos “reducir” o “suprimir” se refieren a la capacidad de provocar una reducción global del 5 % o más, del 10 % o más, del 20 % o más, del 50 % o más, o del
15 75 % o más. El término “inhibir” o expresiones similares también abarca la inhibición completa o casi completa, tal como una reducción hasta cero o efectivamente cero.

Los términos “aumentar”, “potenciar”, “fomentar”, o “prolongar” se refieren a aumentos, potenciaciones, fomentos, o prolongaciones en al menos aproximadamente el 10 %, al menos aproximadamente el 20 %, al menos aproximadamente el 30 %, al
20 menos aproximadamente el 40 %, al menos aproximadamente el 50 %, al menos aproximadamente el 80 %, al menos aproximadamente el 100 %, al menos aproximadamente el 200 %, o al menos aproximadamente el 300 %. Estos términos también se refieren a aumentos, potenciaciones, fomentos, o prolongaciones desde cero o niveles indetectables hasta niveles superiores a cero o niveles detectables/medibles.

25 El término “vacuna” se refiere a una preparación (composición) o un producto farmacéutico que induce una respuesta inmunitaria, particularmente una respuesta inmunitaria celular, que reconoce y ataca células enfermas, tales como células cancerosas, o patógenos cuando se administra. Las vacunas pueden usarse para la

prevención o el tratamiento de enfermedades. El término “vacunas contra el cáncer personalizadas” o “vacunas contra el cáncer individualizadas” se refiere a vacunas contra el cáncer adaptadas al entorno específico de un paciente con cáncer individual, ajustadas según sea necesario para las necesidades únicas de la afección del paciente.

5 En una realización, la vacuna proporcionada por la presente divulgación puede incluir uno o más péptidos seleccionados como péptidos inmunogénicos según el método de la presente divulgación o un ácido nucleico (por ejemplo, ARN) que codifica para el/los péptido(s).

La vacuna contra el cáncer proporcionada por la presente divulgación está
10 diseñada para proporcionar uno o más epítomos de células T que son adecuados para la estimulación, sensibilización, y/o expansión de células T específicas de tumor cuando se administra a un paciente. Las células T pueden seleccionar como diana células que expresan el antígeno a partir del cual se deriva el epítomo de células T. Por tanto, las vacunas descritas en el presente documento pueden inducir o potenciar respuestas
15 celulares, tales como actividad de células T citotóxicas, frente al cáncer al presentar uno o más neoantígenos relacionados con tumor a través de MHC de clase I. Dado que la vacuna selecciona como diana mutaciones específicas de cáncer, puede ser específica para el tumor del paciente.

La vacuna proporcionada por la presente divulgación puede incluir uno o más
20 epítomos de células T, por ejemplo, 2 o más, 5 o más, 10 o más, 15 o más, 20 o más, 25 o más, o hasta 60, 55, 50, 45, 40, 35, ó 30 epítomos de células T, cuando se administra a un paciente. Estos epítomos de células T también pueden denominarse “neoepítomos” en la presente divulgación. La presentación de estos epítomos por las células del paciente, especialmente células presentadoras de antígeno, induce preferiblemente células T que
25 seleccionan como diana el epítomo, permitiendo de ese modo seleccionar como diana el tumor del paciente, incluyendo tumores tanto primarios como metastásicos, al reconocer y atacar el mismo epítomo presentado en la superficie de las células tumorales.

El término “tumor” o “enfermedad tumoral” se refiere preferiblemente al crecimiento

anómalo de células (también denominadas células neoplásicas, células tumorales, o células de neoplasia) que forman tumefacciones o lesiones. “Células tumorales” se refieren a células anómalas que crecen de manera persistente a través de proliferación celular desregulada, incluso después de que cesen los estímulos que iniciaron el nuevo
5 crecimiento. Los tumores pueden carecer de conexiones funcionales con los tejidos normales y están desorganizados de manera parcial o completa, normalmente formando masas tisulares distintas que pueden ser benignas, premalignas, o malignas.

El cáncer es una clase de enfermedades en la que los grupos de células muestran un crecimiento descontrolado (división más allá de los límites normales), invasión
10 (infiltración y destrucción de tejido adyacente), y, algunas veces, metástasis (diseminación a otras partes del cuerpo a través de linfa o sangre). Estas tres características malignas distinguen el cáncer de los tumores benignos, que son autolimitados y no invaden ni experimentan metástasis. La mayoría de los cánceres forman tumores, pero algunos, como la leucemia, no lo hacen. Los términos afección neoplásica maligna, neoplasia
15 maligna, y tumor maligno son esencialmente sinónimos de cáncer.

Con los propósitos de la presente divulgación, los términos “cáncer” y “enfermedad cancerosa” se usan de manera intercambiable con “tumor” y “enfermedad tumoral”.

El término “cáncer” puede referirse a carcinomas, adenocarcinomas, blastomas, leucemia, mesoteliomas, melanomas, teratomas, linfomas, neuroblastomas, gliomas,
20 cáncer rectal, cáncer endometrial, cáncer de riñón, cáncer suprarrenal, cáncer de tiroides, cáncer hemático, cáncer cutáneo, cáncer cerebral, cáncer de cuello uterino, cáncer intestinal, cáncer de hígado, cáncer de colon, cáncer de estómago, cáncer de intestino, cáncer de cabeza y cuello, cáncer gastrointestinal, linfoma, cáncer de esófago, cáncer colorrectal, cáncer de páncreas, cáncer de oídos, nariz y garganta (ENT), cáncer de
25 mama, cáncer de próstata, cáncer de útero, cáncer de ovario, cáncer de pulmón, y otros. El cáncer puede ser un cáncer primario o un cáncer metastásico.

El término “tratamiento” se refiere a reducir el tamaño o número de tumores en un sujeto; detener o ralentizar la progresión de la enfermedad; prevenir o retrasar el desarrollo

de nuevas enfermedades en el sujeto; reducir la frecuencia o gravedad de síntomas y/o recidivas en sujetos que actualmente padecen o previamente padecieron la enfermedad; y/o prolongar la esperanza de vida del sujeto.

“Inmunoterapia” se refiere a un método de tratamiento que implica la activación de una respuesta inmunitaria específica. En el contexto de la presente divulgación, términos como “protección”, “prevención”, “preventivo”, “prevenir”, y “protector” se refieren a prevenir o tratar el inicio y/o la diseminación de una enfermedad en un sujeto, o ambos, particularmente minimizando o retrasando la progresión de la enfermedad. Por ejemplo, un individuo en riesgo de desarrollar cáncer sería un candidato para la terapia preventiva del cáncer.

La administración preventiva de inmunoterapia, tal como la administración preventiva de la vacuna de la presente divulgación, preferiblemente protege al receptor frente al desarrollo de la enfermedad. La administración terapéutica de inmunoterapia, tal como la administración terapéutica de la vacuna de la presente divulgación, puede conducir a la supresión de la progresión/crecimiento de la enfermedad. Esto incluye ralentizar o detener la progresión/crecimiento de la enfermedad, por ejemplo, conduciendo a la eliminación de la enfermedad.

La inmunoterapia puede realizarse usando diversas técnicas donde la formulación proporcionada por la presente divulgación funciona eliminando las células enfermas a partir del paciente. Tal eliminación puede producirse como resultado de inducir o potenciar una respuesta inmunitaria en el paciente que es específica para el antígeno o las células que expresan el antígeno.

La formulación y la composición proporcionadas por la presente divulgación pueden usarse solas o en combinación con otros tratamientos, tales como cirugía, radioterapia, quimioterapia, y/o trasplantes de médula ósea (autólogos, singénicos, alogénicos, o no relacionados).

Los términos “inmunización” o “vacunación” describen un procedimiento de tratamiento de un paciente con la intención de inducir una respuesta inmunitaria con

propósitos terapéuticos o profilácticos.

El término “*in vivo*” se refiere a pertenecer a situaciones que se producen dentro del cuerpo de un sujeto.

Los términos “sujeto”, “individuo”, o “paciente” se usan de manera intercambiable y se refieren a vertebrados, tales como mamíferos. Por ejemplo, en el contexto de la presente divulgación, los mamíferos pueden incluir seres humanos, primates no humanos, animales domesticados (por ejemplo, perros, gatos, ovejas, ganado bovino, cabras, cerdos, caballos), animales de laboratorio (por ejemplo, ratones, ratas, conejos, cobayas), y animales cautivos en zoológicos. Tal como se usa en el presente documento, se pretende que el término “animal” abarque seres humanos. El término “sujeto” se refiere a un paciente, es decir, un animal, preferiblemente un ser humano que padece una enfermedad descrita en el presente documento, o a una muestra biológica (por ejemplo, células cancerosas, sangre) aislada a partir del mismo.

La composición de vacuna proporcionada en el presente documento puede incluir además un adyuvante. El término “adyuvante” se refiere a un compuesto que prolonga, potencia, o acelera una respuesta inmunitaria. Preferiblemente, la composición de la presente divulgación ejerce sus efectos sin la adición de ningún adyuvante. Sin embargo, la composición también puede contener un adyuvante conocido. Los adyuvantes pueden incluir grupos heterogéneos de compuestos tales como emulsiones de aceite (por ejemplo, adyuvante de Freund), compuestos minerales (por ejemplo, alumbre), productos bacterianos (por ejemplo, toxina *Bordetella pertussis*), liposomas, y complejos inmunoestimuladores. Los ejemplos de adyuvantes incluyen monofosforil-lípido A (MPL, SmithKline Beecham), saponinas tales como QS21 (SmithKline Beecham), DQS21 (SmithKline Beecham; documento WO 96/33739), QS7, QS17, QS18, y QS-L1, adyuvante incompleto de Freund, adyuvante completo de Freund, vitamina E, Montanide, alumbre, oligonucleótidos CpG, y diversas emulsiones de aceite en agua derivadas de aceites biológicamente degradables tales como escualeno y/o tocoferol.

La composición de vacuna puede incluir además sustancias farmacéuticamente

aceptables que estimulan la respuesta inmunitaria del paciente. Por ejemplo, tales sustancias pueden ser citocinas. Los ejemplos de citocinas incluyen interleucina-12 (IL-12), GM-CSF, e IL-18, que pueden potenciar la actividad protectora de la vacuna.

La composición de vacuna también puede incluir un compuesto que potencia la respuesta inmunitaria. Estos compuestos pueden incluir moléculas coestimuladoras proporcionadas en forma de ácido nucleico o proteína, tales como B7-1 y B7-2 (también conocidos como CD80 y CD86).

La composición de vacuna proporcionada en el presente documento puede administrarse por vías convencionales, tales como inyección o infusión. La administración puede realizarse, por ejemplo, por vía oral, por vía intravenosa, por vía intraperitoneal, por vía intramuscular, por vía subcutánea, o por vía transdérmica. En una realización, la administración puede realizarse por vía intraganglionar, tal como mediante inyección en los ganglios linfáticos.

La composición de vacuna puede administrarse en una cantidad terapéuticamente eficaz. Una “cantidad terapéuticamente eficaz” se refiere a una cantidad suficiente para provocar una respuesta o un efecto deseado, o bien sola o bien en combinación con dosis adicionales. La cantidad terapéuticamente eficaz puede variar dependiendo de diversos factores incluyendo la afección que va a tratarse, la gravedad de la enfermedad, factores del paciente individual tales como la edad, el estado fisiológico, el tamaño y el peso, la duración del tratamiento, terapias concomitantes (si las hay), la vía de administración específica, y similares.

La composición de vacuna puede ser estéril y contener una cantidad eficaz de la sustancia terapéutica activa para producir la respuesta o el efecto deseado.

El término “farmacéuticamente aceptable” se refiere a pertenecer a una sustancia no tóxica que no interfiere en la acción del principio activo en una composición farmacéutica. Tales sustancias pueden incluir sales, agentes tamponantes, conservantes, portadores, adyuvantes, y otros inmunoestimulantes auxiliares, tales como oligonucleótidos CpG, citocinas, quimiocinas, saponinas, GM-CSF, y/o ARN. Cuando sea

apropiado, también pueden incluirse otros compuestos activos terapéuticos.

Un “portador” se refiere a un componente orgánico o inorgánico, natural o sintético, combinado con el principio activo para facilitar, potenciar, o permitir la aplicación del principio activo. En este contexto, el término “portador farmacéuticamente aceptable” se
5 refiere a una o más cargas, diluyentes, o sustancias de encapsulación sólidos o líquidos adecuados que son apropiados para su administración a un paciente.

La composición de vacuna puede contener un agente de tamponamiento adecuado, tal como sal de acetato, sal de citrato, sal de borato, o sal de fosfato.

La composición de vacuna puede contener un conservante adecuado, tal como
10 cloruro de benzalconio, clorobutanol, parabenos, y timerosal.

La composición de vacuna se proporciona normalmente en una forma de dosificación unitaria y puede prepararse mediante métodos bien conocidos en la técnica. La composición de vacuna de la presente divulgación puede adoptar la forma de inyectables, cápsulas, comprimidos, pastillas para chupar, disoluciones, suspensiones,
15 jarabes, elixires, o emulsiones, por ejemplo.

Las composiciones adecuadas para administración parenteral incluyen generalmente una preparación acuosa o no acuosa estéril del compuesto activo que es isotónica con la sangre del receptor. Los portadores y disolventes adecuados incluyen solución de Ringer y disolución isotónica de cloruro de sodio. Además, como disolvente o
20 medio de suspensión pueden usarse aceites fijos estériles.

【Efectos de la invención】

Cuando se aplican las características de la presente divulgación, el procesamiento de variantes de mutación y HLA de tipo I puede realizarse de manera eficiente a través
25 del procesamiento de datos de NGS para pacientes.

En la presente divulgación, se presenta un método para seleccionar neoantígenos adecuados para vacunas contra el cáncer personalizadas empleando tecnología de IA para el aprendizaje de base de datos. El método incorpora una novedosa aplicación y

validación de un modelo de *afinidad de unión* y un modelo de *inmunogenicidad* para seleccionar candidatos a neoantígeno, seguido de una nueva aplicación del procedimiento de *priorización de neoantígeno*.

5 **【Breve descripción de los dibujos】**

La figura 1 muestra la distribución de puntuaciones de afinidad de unión (BA) y puntuaciones de inmunogenicidad (IMM).

La figura 2a ilustra esquemáticamente el procedimiento de medir puntuaciones de inmunogenicidad según el ejemplo 4.

10 Las figuras 2b a 2d muestran los resultados de confirmar la inmunogenicidad a través de ELISPOT de IFN γ mediante HLA de tipo I en el ejemplo 4 (2b: A*02:01, 2c: A*24:02, 2d: A*11:01).

La figura 3 es una ilustración esquemática del procedimiento de prueba del ejemplo 5.1.

15 La figura 4 es una gráfico de los resultados de ELISPOT de IFN γ de esplenocitos obtenidos en el ejemplo 5.1.

La figura 5 es una ilustración esquemática del procedimiento de prueba del ejemplo 5.2.

20 La figura 6 muestra gráficos de los resultados de ELISPOT de IFN γ de esplenocitos obtenidos en el ejemplo 5.2.

La figura 7 es una representación gráfica del efecto antineoplásico (*in vivo*) del neoantígeno confirmado en el ejemplo 6.

【Modo para la invención】

25 Puede obtenerse una mejor comprensión de la presente divulgación a través de los siguientes ejemplos que se exponen para ilustrar, pero no debe interpretarse que limitan, la presente divulgación.

EJEMPLO 1. Construcción de modelo de predicción de neoantígeno (entrenamiento de modelo) (modelo de predicción de neoantígeno I)

Usando el antígeno leucocitario humano (HLA) de tipo I, específicamente HLA I-A, B, y C, e información de secuencia de aminoácidos de péptido a partir del recurso de análisis y base de datos de epítomos inmunes (IEDB) (<https://www.iedb.org>), se preparó un modelo de inteligencia artificial para predecir la afinidad de unión y la inmunogenicidad de péptidos neoantígenos candidatos. Específicamente, los modelos de predicción de afinidad de unión e inmunogenicidad consistían en una agrupación de 150 modelos diferentes entrenados, cada uno, usando diferentes estructuras de modelo, métodos de división de datos, inclusión de aprendizaje por transferencia, métodos de optimización de entrenamiento, y criterios de parada. Las estructuras de modelo se eligieron de entre GRU (unidad recurrente cerrada) y transformador. Se aplicó un método de validación cruzada de 5 pliegues para la división de datos. Para el aprendizaje por transferencia, el modelo de predicción de afinidad de unión usó un clasificador entrenado con datos cualitativos, mientras que el modelo de predicción de inmunogenicidad utilizó el modelo de predicción de afinidad de unión que tenía el error de predicción más pequeño para los datos de validación. Para el método de optimización de entrenamiento, puede usarse o bien AdamW o promedio estocástico de pesos (SWA). Los criterios de parada se seleccionaron basándose en los errores de predicción más bajos a partir de los datos de validación durante todo el procedimiento de entrenamiento, y se eligieron cinco puntos en orden de los errores más bajos.

El procedimiento de preparación del modelo de predicción de neoantígeno se describe en más detalle a continuación:

1.1. Selección de datos de entrada para el modelo de predicción de neoantígeno

Los datos necesarios para el entrenamiento del modelo de HLA de tipo I se seleccionaron a partir de la base de datos IEDB y se transformaron para su uso en el

entrenamiento del modelo de predicción de neoantígeno. Estos datos se aplicaron tanto al modelo de afinidad de unión como al modelo de inmunogenicidad. Por ejemplo, el procedimiento de selección de datos puede implicar una o más tareas tales como eliminar datos que no corresponden a HLA-A, HLA-B, o HLA-C, retirar datos con mutaciones en moléculas de HLA, excluir péptidos con una longitud de menos de 7 aminoácidos o más de 15 aminoácidos, y eliminar secuencias que incluyen aminoácidos de origen no humano fuera de los 20 aminoácidos convencionales. Por ejemplo, el procedimiento de transformación de datos puede implicar una o más tareas tales como normalizar valores de medición de afinidad de unión (por ejemplo, CI50) a un intervalo de entre 0 y 1 y convertir los datos de puntuación de inmunogenicidad proporcionados como [número de pruebas, número de respuestas] en la media de una distribución beta (L. Guangyuan *et al.*, “DeepImmuno: deep learning-empowered prediction and generation of immunogenic peptides for T-cell immunity” Briefings in Bioinformatics, 22.6, 2021).

Para el entrenamiento del modelo de predicción de inmunogenicidad, se seleccionaron métodos de prueba relacionados con la inmunogenicidad incluyendo los niveles de citocinas, tales como IFN γ , IL-2, TNF α , granzima A, etc., de entre los diversos resultados experimentales en la base de datos IEDB (datos sobre citocinas derivadas de células T, incluyendo IFN- γ , TNF- α , IL-2, IL-1 β , producidos por el tratamiento con PBMC de cada péptido, proliferación de células T desencadenada por activación inmunogénica, quimiocinas tales como CCL4 y CXCL9, y citotoxicidad de células T activadas, representada por granzima B, etc.).

1.2. Método de entrenamiento para el modelo de predicción de neoantígeno

Usando los datos seleccionados en el ejemplo 1.1, se realizaron las siguientes etapas:

Para el modelo de predicción de afinidad de unión, se realizó la selección de uno de dos arquitecturas de modelo (GRU (J. Chung *et al.*, <https://arxiv.org/abs/1412.3555>, 2014) y transformador (Vaswani, Ashish, *et al.*, “Attention is all you need”, Advances in

Neural Information Processing Systems 30, 2017)), uno de los tres métodos de entrenamiento, y uno de los métodos de división de datos por validación cruzada de 5 pliegues. El modelo se guardó en cinco puntos de control de entrenamiento diferentes donde el error de predicción para el conjunto de datos de validación era el más bajo, dando como resultado un total de 150 sesiones de entrenamiento de modelo ($2 \times 3 \times 5 \times 5 = 150$). Se calculó la predicción de afinidad de unión final (valor de predicción 1) promediando los valores de predicción de los 150 modelos entrenados y calculando su varianza.

De manera similar, para el modelo de predicción de inmunogenicidad, se realizó la selección de una de dos arquitecturas de modelo (GRU, transformador), uno de los tres métodos de entrenamiento (incluyendo entrenamiento basándose en el modelo con el error de validación más bajo a partir del modelo de afinidad de unión), y uno de los métodos de división de datos por validación cruzada de 5 pliegues, y se aplicó el mismo enfoque de guardado del modelo en los cinco puntos de control con el error de predicción más bajo para el conjunto de datos de validación, conduciendo a un total de 150 sesiones de entrenamiento de modelo ($2 \times 3 \times 5 \times 5 = 150$). Se calculó la predicción de inmunogenicidad final (valor de predicción 2) promediando las predicciones de los 150 modelos y calculando su varianza.

En el cálculo del valor de predicción 1 y el valor de predicción 2, se determinó el valor de predicción final usando [(el promedio de los 150 valores de predicción de modelo) - (la varianza de los 150 valores de predicción de modelo)]. Se eligió este enfoque porque una varianza más pequeña entre las predicciones de los 150 modelos indica una predicción más estable y fiable.

EJEMPLO 2: Fase de predicción y selección de neoantígeno

Se seleccionaron neoantígenos usando el modelo de predicción de neoantígeno construido en el ejemplo 1, y el procedimiento de selección avanzó a través de las siguientes etapas:

1) Recopilación de datos de NGS de paciente con cáncer a partir de bases de datos públicas;

2) Determinación del tipo de HLA e identificación de péptidos derivados de células cancerosas usando los datos de NGS del paciente con cáncer, seguido del cálculo de TPM;

3) Cálculo de puntuaciones de predicción usando el modelo de predicción de neoantígeno del ejemplo 1;

4) Cálculo de priorización de neoantígeno y selección de péptidos experimentales usando la fórmula del ejemplo 3;

5) Realización de experimentos de afinidad de unión e inmunogenicidad usando PBMC con el mismo tipo de HLA identificado en la etapa 2).

Para la verificación del modelo de predicción de neoantígeno, se utilizaron datos disponibles públicamente de cuatro pacientes con melanoma que lograron una respuesta completa tras el tratamiento con pembrolizumab (tipo de HLA: A*02:01; véase “John Wiley & Sons A/S, Tissue Antigens 73, 95–170”) como datos de muestra de paciente (a continuación en el presente documento “datos de muestra de paciente” o “conjunto de datos de muestra de paciente”). La verificación se realizó con referencia al estudio: [Hugo, Willy, *et al.* “Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma”. *Cell* 165,1 (2016): 35-44.].

2.1. Preparación de datos para la selección de neoantígeno (procesamiento de datos de secuenciación de ARN y secuenciación de exoma completo derivados de paciente)

Para el conjunto de datos de muestra de paciente, se configuró una combinación de códigos para realizar procesamiento previo de datos, descubrimiento de variantes, y refinamiento de conjunto de llamadas, lo que permite la extracción de secuencias de mutación e información de tipo de HLA. Específicamente:

1) Se convirtió información de mutación de ADN en información de mutación de

péptido, y se concatenaron datos de abundancia de ARN y datos de tipo de HLA. Además, se llevó a cabo un procedimiento para comprobar si los péptidos derivados de estas mutaciones se encontraban en secuencias de proteínas conocidas de *Homo sapiens* usando la base de datos SwissProt de Uniprot.

5 2) Se obtuvieron datos de mutación usando herramientas como Mutect2, incluida en el paquete GATK desarrollado por Broad Institute, y Strelka, desarrollada por Illumina, que se usan para identificar mutaciones genéticas específicas de cáncer. Después de procesar la información de mutación de ADN, se organizaron las mutaciones de aminoácido y las deleciones en marco a nivel de proteína, y se unificó cualquier péptido
10 identificado tanto en Mutect2 como en Strelka en una única entrada. (Koboldt, D. C. *et al.*, VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research* 22, 568–576 (2012); Cai, L., Yuan, W., Zhang, Z., He, L. y Chou, K. C., In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Scientific Reports* 6, 9,
15 <https://doi.org/10.1038/srep36540> (2016); Kim, S. *et al.*, Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* 15, 591–594, <https://doi.org/10.1038/s41592-018-0051-x> (2018); McKenna, A. *et al.*, The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303, <https://doi.org/10.1101/gr.107524.110> (2010);
20 Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., y Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3), 213–219. doi:10.1038/nbt.2514).

3) Se calculó la hidrofobia de los péptidos para preparar datos para la posterior producción de neoantígenos. Aunque las consideraciones de hidrofobia pueden no ser
25 necesarias cuando se desarrollan neoantígenos para vacunas contra el cáncer personalizadas usando ARNm, se vuelven cruciales cuando se producen vacunas contra el cáncer a base de péptidos o se realizan pruebas de inmunogenicidad. Una alta hidrofobia (por ejemplo, un alto contenido de aminoácidos hidrófobos) puede hacer que

la producción de péptido no sea viable. En este ejemplo, se evaluó la hidrofobia calculando la puntuación de Gravy (promedio global de hidropatía).

2.2. Selección de neoantígeno usando el modelo de predicción de neoantígeno (afinidad de unión e inmunogenicidad) (verificación para la aplicación de vacunas contra el cáncer personalizadas)

Usando los datos preparados en el ejemplo 2.1, se establecieron umbrales para la afinidad de unión y la inmunogenicidad para aplicar el modelo de predicción de neoantígeno, y se verificaron los neoantígenos seleccionados tal como sigue:

10 - Se comparó la afinidad de unión de los neoantígenos predichos con un péptido patrón, NY-ESO-1 (péptido de carcinoma de células escamosas de esófago de Nueva York de tipo 1 (157-165, SLLMWITQC; sintetizado por AnyGen)). Se validó la afinidad de unión de los neoantígenos predichos comparando los valores de CI50 de los mismos frente al patrón conocido.

15 - Se realizó un ensayo de inmunogenicidad por ELISPOT de IFN γ para evaluar la inmunogenicidad de los neoantígenos predichos.

 - Se dejaron reposar PBMC normales de un donante HLA-A*02:01 durante 24 horas, y se aplicaron pulsos de péptidos candidatos (50 μ M) con IL-2 (10 UI/ml) (72 horas, 37 °C, 5 % de CO₂). Luego se lavaron las PBMC y se transfirieron a una placa de
20 ensayo de ELISPOT, seguido de reestimulación con el mismo péptido (50 μ M, 24 horas, 37 °C, 5 % de CO₂). Se realizó el ensayo de ELISPOT según el protocolo de inmunotransferencia por manchas, y se repitió el experimento con PBMC de tres donantes adicionales para verificar la reproducibilidad y la significación estadística.

25 La secuencia usada para la porción de cadena alfa de HLA-A*02:01 es tal como sigue:

SHSMRYFFTSVSRPGRGEPRIAVGYVDDTQFVRFDSDAASQRMEPRAPWI
EQEGPEYWDGETRKVKAHSQTHRVDLGLRGYYNQSEAGSHTVQRMYGCDVGSD

WRFLRGYHQYAYDGKDYIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQLRAYLE
GTCVEWLRRLRYLENGKETLQRT (SEQ ID NO: 1)

Dada la variabilidad en las pruebas de inmunogenicidad debido a la naturaleza de los experimentos basados en células, se aplicaron normalización Min-Max y valores ponderados (detección de 1 lote: x1, detección de 2 lotes: x2, detección de 3 lotes: x3) para clasificar los neoantígenos y establecer un método para seleccionar neoantígenos adecuado para la aplicación de vacunas contra el cáncer personalizadas.

Se procesaron los datos de péptido mutante de 118.057 péptidos derivados de muestras de paciente (cuatro pacientes con melanoma tratados con pembrolizumab que lograron una respuesta completa, tipo de HLA: A*02:01) usando el modelo de predicción de afinidad de unión (BA) y el modelo de predicción de inmunogenicidad (IMM) del ejemplo 1. Los resultados se muestran en la figura 1.

Adicionalmente, a partir de los resultados en la figura 1, se identificaron dos grupos: grupo de alta BA (G1; $0,15 \leq$ puntuación de IMM $< 0,25$ y $0,6 \leq$ puntuación de BA) y grupo de alta IMM (G2: puntuación de IMM $\geq 0,25$ y $0,6 >$ puntuación de BA $\geq 0,1$). Se midieron la afinidad de unión y la inmunogenicidad reales para estos grupos y se muestran en la tabla 1:

TABLA 1

HLA-I	Afinidad de unión (BA)	Inmunogenicidad (IMM)	Prueba de selección de neoantígeno
G1	100 %	5,6 %	Cuatro pacientes con melanoma (CR) con tratamiento con pembrolizumab
G2	49 %	23,6 %	

Tal como se muestra en la tabla 1:

Cuando se estableció el umbral para que fuera de $0,15 \leq$ puntuación de IMM $< 0,25$ y $0,6 \leq$ puntuación de BA (G1), la precisión de predicción para la afinidad de unión positiva fue del 100 %.

Cuando se estableció el umbral para que fuera de puntuación de $IMM \geq 0,25$ y $0,6 > \text{puntuación de BA} \geq 0,1$ (G2), la precisión de predicción para la inmunogenicidad positiva fue del 23,6 %.

5 **EJEMPLO 2.3. Medición del nivel de expresión de ARN (TPM) como variable adicional**

Para evaluar adicionalmente los neoantígenos predichos, se midió TPM (nivel de expresión de ARN; transcritos por millón) como variable adicional. TPM indica la tasa de expresión de secuencias de ARN a partir de los candidatos a neoantígeno, que refleja la

10 probabilidad de que las mutaciones en las células cancerosas del paciente se traduzcan a péptidos reales. Sin embargo, dado que la base de datos IEDB no proporciona datos de TPM, se usó para la priorización de los resultados de predicción en lugar de para el entrenamiento del modelo. Para desarrollar criterios de evaluación usando TPM, se

15 de datos TESLA; Wells, Daniel K., *et al.* "Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction". *Cell* 183.3 (2020): 818-834) para seleccionar candidatos a neoantígeno (candidatos a péptido). Se calcularon los valores de predicción (valor de predicción 1 y valor de predicción 2) de estos

20 candidatos a neoantígeno usando el modelo de predicción (véase el ejemplo 1.2). Se procesaron los datos de secuenciación de ARN incluidos en los datos de NGS disponibles públicamente de pacientes con cáncer para calcular los niveles de expresión de ARN (TPM) de los candidatos a neoantígeno.

25 **EJEMPLO 3: Evaluación de los neoantígenos seleccionados (priorización de neoantígeno)**

Con respecto a los valores de predicción 1 y 2 obtenidos a partir del ejemplo 1.2, se realizó la optimización de las variables en el modelo de afinidad de unión y el modelo de inmunogenicidad a través de análisis de regresión lineal multivariante. Finalmente, se

confirmaron TPM (nivel de expresión de ARN), la puntuación de predicción de afinidad de unión (BA; valor de predicción 1 del ejemplo 3), y la puntuación de predicción de inmunogenicidad (IMM; valor de predicción 2 del ejemplo 3) como variables, y se determinó la clasificación del neoantígeno.

5 El conjunto de datos TESLA incluye los tipos de HLA y datos de NGS de 8 pacientes, 845 secuencias de péptido como neoantígenos candidatos derivadas de esos pacientes, y los resultados experimentales de inmunogenicidad. Entre los 845 candidatos, se verificó que 38 péptidos tenían inmunogenicidad. Para comparar la excelencia de los criterios de predicción de neoantígeno, se seleccionaron
10 20 neoantígenos candidatos para cada uno de los 8 pacientes (un total de 160) aplicando TPM y variando la ponderación de la puntuación de predicción de afinidad de unión (BA) y la puntuación de predicción de inmunogenicidad (IMM). Se registró el número de candidatos a neoantígeno con inmunogenicidad verificada de entre los 38 en la tabla 2 a continuación (umbral: TPM (transcritos por millón) > 1):

15 TABLA 2

Distribución de ponderación	W1: 1,0 W2: 0,0	W1: 0,8 W2: 0,2	W1: 0,5 W2: 0,5	W1: 0,2 W2: 0,8	W1: 0,0 W2: 1,0
TESLA	14	18	26	23	23

(En la tabla 2, W1 se refiere a la ponderación de BA y W2 se refiere a la ponderación de IMM)

Tal como se confirma en la tabla 2, cuando las ponderaciones de la puntuación de predicción de afinidad de unión (BA) y la puntuación de predicción de inmunogenicidad
20 (IMM) eran iguales (1:1), se identificó el número más alto de neoantígenos inmunogénicos.

Basándose en los resultados anteriores, se estableció la fórmula para clasificar los neoantígenos tal como sigue:

Esquema de puntuación de priorización de neoantígeno: $[\min(\text{TPM}, 1) * (\text{BA} + \text{IMM})/2]$

El número de neoantígenos detectados antes y después de aplicar la fórmula de priorización se muestra en la tabla 3 a continuación:

TABLA 3

	Paciente	Positivo	Antes de la priorización	Después de la priorización
Ronda piloto	#4	1	1	1
	#8	1	0	1
	#9	2	1	2
	[310]	4	2	4
Ronda principal	#1	9	3	7
	#2	4	3	2
	#3	13	2	9
	#12	4	2	2
	#16	4	1	2
	[535]	34	11	22

5

Tal como se muestra en la tabla 3, la aplicación del método de priorización aumentó el número de neoantígenos seleccionables, confirmando una mejora de la precisión de predicción de neoantígeno.

10 EJEMPLO 4. Modelo de predicción de neoantígeno II y selección de neoantígeno

Haciendo referencia a los métodos descritos en los ejemplos 1 a 3, se seleccionaron pacientes con melanoma que poseen tres tipos de HLA diferentes (A*02:01, A*24:02, A*11:01), y se predijeron 20 neoantígenos usando el modelo de predicción de
15 neoantígeno. Se confirmó la inmunogenicidad de los neoantígenos predichos usando ensayos de ELISPOT de IFN γ .

Más específicamente, los criterios para seleccionar datos de paciente fueron tal como sigue:

- Tumor con alta carga mutacional → Melanoma;
- WES completa (secuenciación de exosoma completo; selección de mutación), conjunto de datos de secuenciación de ARN (nivel de expresión) (convertido en secuencias de péptido para su uso);

5 - Tratamiento con ICI (inhibidor de punto de control inmunitario) (terapia con anticuerpos anti-PD-1: pembrolizumab, nivolumab) → CR (respuesta completa), PR (respuesta parcial), PD (enfermedad progresiva);

Tipo de HLA: A*02:01, A*24:02, A*11:01 ([HLA-A*02:01]:
 SHSMRYFFTSVSRPGRGEPRFIAVGYVDDTQFVRFSDAASQRMEPRAPWIEQEG
 10 PEYWDGETRKYKAHSQTHRVDLGLRGYYNQSEAGSHTVQRMYGCDVGS DWRFL
 RGYHQYAYDGKDYIALKEDLRSWTAADMAAQTTKHKWEAAHVAEQLRAYLEGTCV
 EWLRRYLENGKETLQRT (SEQ ID NO: 1), [HLA-A*11:01]:
 SHSMRYFYTSVSRPGRGEPRFIAVGYVDDTQFVRFSDAASQRMEPRAPWIEQEG
 PEYWDQETRNVKAQSQTDRVDLGLRGYYNQSE DGSHTIQIMYGCDVGP DGRFLR
 15 GYRQDAYDGKDYIALNEDLRSWTAADMAAQITKRKWEAAHAAEQQRAYLEGRCVE
 WLRRYLENGKETLQRT (SEQ ID NO: 2), [HLA-A*24:02]:
 SHSMRYFSTSVSRPGRGEPRFIAVGYVDDTQFVRFSDAASQRMEPRAPWIEQEG
 PEYWDEETGKVKHSQTDRENLR LRYYNQSEAGSHTLQMMFGCDVGS DGRFLR
 GYHQYAYDGKDYIALKEDLRSWTAADMAAQITKRKWEAAHVAEQQRAYLEGTCVD
 20 GLRRYLENGKETLQRT (SEQ ID NO: 3);

- Fuente: Hugo *et al.*, 2016, Cell 165, 35–44, Riaz *et al.*, 2017, Cell 171, 934–949 (incorporados en el presente documento mediante referencia en su totalidad).

El modelo de predicción de neoantígeno aplicado a los datos de paciente
 25 seleccionados fue tal como sigue:

- Puntuación de afinidad de unión (BA) = (puntuación de afinidad de unión media del grupo de pacientes seleccionado) – (varianza de la puntuación de afinidad de unión del grupo de pacientes seleccionado);

- Puntuación de inmunogenicidad (IMM) = (puntuación de inmunogenicidad media del grupo de pacientes seleccionado) – (Varianza de la puntuación de inmunogenicidad del grupo de pacientes seleccionado).

5 Se obtuvo la puntuación de BA combinando el neoantígeno de cada paciente y sus A*02:01, A*24:02, y A*11:01 únicos usando el modelo de predicción de afinidad de unión entrenado para el valor de predicción 1 en el ejemplo 1.2 (véase el ejemplo 2.1). Se obtuvo la puntuación de IMM prediciendo la inmunogenicidad del neoantígeno de cada paciente y sus tipos de HLA únicos usando el modelo de predicción de inmunogenicidad
10 entrenado para el valor de predicción 2 en el ejemplo 1.2 (véase el ejemplo 2.1).

Usando la puntuación de TPM (nivel de expresión de ARN; transcritos por millón), se calculó la puntuación de priorización de neoantígeno (umbral de TPM: 1; véase el ejemplo 3):

15 Puntuación de priorización = $\min(\text{TPM}, 1) * [(BA + IMM)/2]$

$\min(\text{TPM}, 1)$ significa que si $\text{TPM} > 1$, el valor de TPM se reemplaza por 1; si $\text{TPM} < 1$, se usa el valor de TPM real.

Se clasificaron los neoantígenos según la puntuación de priorización calculada, y
20 se predijeron los 20 mejores péptidos como neoantígenos (véanse las tablas 4 a 6).

TABLA 4

Clasificación	Secuencia	SEQ ID NO	Tipo de HLA	Puntuación
1	YLFNFVNV	4	A*02:01	0,557830519
2	RLAEFIIGM	5	A*02:01	0,551505094
3	YLFNFVNVG	6	A*02:01	0,501269973
4	RLAEFIIGMM	7	A*02:01	0,499001865
5	LLTDIIPDGV	8	A*02:01	0,494219272
6	FLALCADA	9	A*02:01	0,493808672
7	YIAGILARI	10	A*02:01	0,490358034

8	RLAEFIIGMMG	11	A*02:01	0,477305298
9	YLMVALASV	12	A*02:01	0,471998357
10	RLAEFIIGMMGT	13	A*02:01	0,459548693
11	GLVECLPEI	14	A*02:01	0,456333267
12	AYLFNFVNV	15	A*02:01	0,448139594
13	VLFYLSAV	16	A*02:01	0,448117179
14	RLFFCFMNV	17	A*02:01	0,443291682
15	CILAVLPSL	18	A*02:01	0,440896908
16	IMFNKGFLA	19	A*02:01	0,438771916
17	KTDPWFLLAV	20	A*02:01	0,437365302
18	SLIGVSFFALPV	21	A*02:01	0,437040423
19	SLWAESLVHV	22	A*02:01	0,436619991
20	FLLAVLGAPV	23	A*02:01	0,435148758

TABLA 5

Clasificación	Secuencia	SEQ ID NO	Tipo de HLA	Puntuación
1	TFYWTFVNAF	24	A*24:02	0,455910057
2	VFYQFSKLIF	25	A*24:02	0,403493367
3	RMYSWRLLNL	26	A*24:02	0,371202482
4	TWNVHQQFLW	27	A*24:02	0,366847255
5	RMYSWRLL	28	A*24:02	0,36676025
6	FYWTFVNAFSIF	29	A*24:02	0,363375765
7	QWILKTRAYFL	30	A*24:02	0,338967703
8	STWNVHQQF	31	A*24:02	0,338013999
9	QWILKTRAYF	32	A*24:02	0,330919878
10	STWNVHQQFLW	33	A*24:02	0,330000281
11	FYWTFVNAF	34	A*24:02	0,328668696
12	STWNVHQQFLW	35	A*24:02	0,327606417
13	FYWTFVNAF	36	A*24:02	0,324465905
14	STWNVHQQFLW	37	A*24:02	0,320806235
15	FYWTFVNAF	38	A*24:02	0,319582139
16	STWNVHQQFLW	39	A*24:02	0,316823043
17	FYWTFVNAF	40	A*24:02	0,316026847
18	STWNVHQQFLW	41	A*24:02	0,314389744
19	FYWTFVNAF	42	A*24:02	0,313212986

20	LYHFSDRMYSW	43	A*24:02	0,312662407
----	-------------	----	---------	-------------

TABLA 6

Clasificación	Secuencia	SEQ ID NO	Tipo de HLA	Puntuación
1	STACFHPVTK	44	A*11:01	0,463350479
2	TTFGIQVRYLK	45	A*11:01	0,436937466
3	KIGDFGLATMK	46	A*11:01	0,42843179
4	KIGDFGLATEK	47	A*11:01	0,427919868
5	SSSGSINLLK	48	A*11:01	0,420742704
6	AVAVDFQFSK	49	A*11:01	0,408470424
7	KTAPVEWRK	50	A*11:01	0,392922075
8	STSYLQHREK	51	A*11:01	0,390245612
9	VAVDFQFSK	52	A*11:01	0,382257777
10	SAFTIQEYFVK	53	A*11:01	0,38028147
11	SVSNTTVGSK	54	A*11:01	0,378114721
12	RVQTGTIPK	55	A*11:01	0,377405025
13	SSGSINLLK	56	A*11:01	0,373327626
14	GTAAISVTGALK	57	A*11:01	0,373120201
15	STPNFSRLRRY	58	A*11:01	0,372708181
16	RSVLSSGYQK	59	A*11:01	0,372577947
17	VVSPVSLPRR	60	A*11:01	0,370444394
18	AAISVTGALK	61	A*11:01	0,36963003
19	GSQPPSLGRK	62	A*11:01	0,369425624
20	TVYCRYHNY	63	A*11:01	0,368830969

Para validar los neoantígenos predichos, se obtuvieron PBMC (células mononucleares de sangre periférica) de tres donantes adultos con tipos de HLA conocidos y se usaron para la evaluación de inmunogenicidad a través de ELISPOT de IFN γ (los tres donantes tenían, cada uno, tipos de HLA diferentes, y las PBMC obtenidas portaban o bien A*02:01, o bien A*24:02, o bien A*11:01). El procedimiento experimental se ilustra esquemáticamente en la figura 2.

Los resultados de ELISPOT obtenidos se normalizaron usando la siguiente fórmula para su análisis:

$$Normalización = \frac{X - X(medio)}{X(CPI) - X(medio)}$$

(Umbral establecido a >10 para determinar positivo para inmunogenicidad;

X: Número de manchas (ELISPOT) en el grupo de prueba;

X(medio): Número de manchas en el grupo tratados con medio (grupo no tratado;

5 control negativo);

X(CPI): Número de manchas en el grupo de control positivo tratado con antígenos proteicos de los virus CMV, paragripal, y gripal.

Se realizó normalización para mantener la consistencia entre las diferentes pruebas.

10

Para 8 pacientes (A*02:01), 7 pacientes (A*24:02), y 5 pacientes (A*11:01) con tipos de HLA coincidentes, se predijeron 20 neoantígenos mejores y 10 neoantígenos mejores usando el modelo de predicción de neoantígeno descrito en los ejemplos 1 y 2. Se realizó la evaluación de inmunogenicidad para los 20 neoantígenos predichos usando el método de ELISPOT de IFN γ tal como se describe en los ejemplos 3 y 4. Se determinó el número de neoantígenos inmunogénicos en los 20 neoantígenos predichos de cada paciente.

15

Los resultados obtenidos se muestran en la tabla 7 y las figuras 2b-2d:

20

TABLA 7

Tipo de HLA-I	N.º de pacientes	20 mejores (%)	10 mejores (%)
A*02:01	8	26,8	22,5
A*24:02	7	20,7	14,3
A*11:01	5	21,0	18,0

La evaluación experimental de la inmunogenicidad para A*02:01 (8 pacientes), A*24:02 (7 pacientes), y A*11:01 (5 pacientes) demostró tasas de predicción de neoantígeno inmunogénico del 26,8 %, del 20,7 %, y del 21,0 %, respectivamente.

EJEMPLO 5. Inmunogenicidad ex vivo

5.1. Modelo de xenoinjerto derivado de línea celular humana (línea celular MCF7)

5 La línea celular de cáncer de mama MCF7 es una línea celular de cáncer humana que expresa HLA-A*02:01. Haciendo referencia al ejemplo 4, se realizó secuenciación NGS, incluyendo secuenciación de exoma completo (WES; selección de mutación) y secuenciación de ARN (nivel de expresión), sobre la línea celular MCF7 (ATCC, n.º de cat. HTB-22). Usando el modelo de predicción de neoantígeno (BA, IMM, TPM), se
10 seleccionaron los 20 neoantígenos mejores, y se identificaron 20 neoantígenos (ARNm). Estos se formularon para dar una vacuna contra el cáncer de ARNm-lipolejo (véase el ejemplo A(1) de las patentes n.ºs KR 10-2022-0149461 A y KR 10-2022-0149460 A) y se administraron a ratones transgénicos A02:01 (ratones C57BL/6 modificados por ingeniería para expresar HLA-A2,1; Taconic biosciences, n.º de cat. 8906-F). Se
15 administró la vacuna a 60 µg por dosis, tres veces a la semana (QW), a través de inyección intravenosa (i.v.) en la vena de la cola. Se evaluó la inmunogenicidad usando ELISPOT de IFNγ de esplenocitos.

Se realizó el ensayo de inmunogenicidad tal como sigue: se recubrió una placa de ELISPOT de 96 pocillos con anticuerpos anti-IFNγ. Después de lavar la placa con PBS,
20 se cargaron esplenocitos recogidos de los ratones inmunizados en cada pocillo a $1,0 \times 10^6$ células por pocillo. Para cada grupo inmune, se sembraron las células en pocillos por triplicado. Se usó un volumen total de 200 µl por pocillo, o bien con 10 µM del péptido o bien con el medio solo para la reestimulación. Luego se incubó la placa a 37 °C durante 18 horas. Después de lavar la placa tres veces con PBST, se trató cada pocillo con
25 anticuerpos de detección de IFNγ de ratón marcados con biotina y se incubó durante 2 horas a temperatura ambiente. Se lavaron las células de nuevo, y se añadió estreptavidina conjugada con AP a cada pocillo y se incubaron durante 30 minutos a temperatura ambiente. Se desarrolló la reacción de AP usando conjuntos de reactivos de

sustrato (inmunotransferencia por manchas). Se contaron automáticamente las unidades formadoras de manchas (UFM) por pocillo usando un lector de ELISPOT (lector de ELISPOT CTL).

El procedimiento experimental, incluyendo la administración de vacuna, se ilustra esquemáticamente en la figura 3.

Los 20 neoantígenos seleccionados se muestran en la tabla 8:

TABLA 8

Clasificación	Nombre	Neoantígeno	SEQ ID NO
1	MCF7-1	LLFDMTTHSV	64
2	MCF7-2	FQSFFSLIV	65
3	MCF7-3	FLGSGVHGGV	66
4	MCF7-4	YLFFCGHMFHA	67
5	MCF7-5	ITFEHILAL	68
6	MCF7-6	NLANEITYI	69
7	MCF7-7	VLCEPVHGL	70
8	MCF7-8	MLLESLRHAV	71
9	MCF7-9	SLFGLGFEEHV	72
10	MCF7-10	HLLTDELGYV	73
11	MCF7-11	KLSEASMIV	74
12	MCF7-12	ALLDMRSLL	75
13	MCF7-13	GLWEALTPCNV	76
14	MCF7-14	SLARLHLTV	77
15	MCF7-15	TLLPLPLLL	78
16	MCF7-16	HLFWFADLNYRL	79
17	MCF7-17	STMSFLRSLYL	80
18	MCF7-18	TLNHEIYEA	81
19	MCF7-19	ILWGSISFNLAV	82
20	MCF7-20	MVCGAAFFI	83

Los resultados de ELISPOT de IFN γ de esplenocitos obtenidos se muestran en la figura 4. En la figura 4, el control se refiere al grupo no tratado sin tratamiento con péptido neoantígeno, y “MCF7 NEO” se refiere al grupo tratado con la vacuna contra el cáncer de ARNm-lipoplejo que contiene los 20 neoantígenos seleccionados. El umbral para

considerar un efecto antineoplásico se estableció en tres veces el recuento de ELISPOT promedio del grupo no tratado (control). Como resultado, se predijo que el 25 % de los 20 neoantígenos sometidos a prueba tenían eficacia antineoplásica, tal como indica la marca * en la figura 4 (5/20).

5

5.2. Neoantígenos predichos mediante modelo de IA de B16F10 (melanoma)

Hacienda referencia al ejemplo 4, se realizó NGS sobre la línea celular B16F10 (melanoma de ratón; ATCC, n.º de cat. CRL-6475) (convertida en secuencias de aminoácidos), y se predijeron los 50 neoantígenos mejores y los 20 neoantígenos mejores usando el modelo de predicción de neoantígeno (BA, IMM, TPM). Se formularon los neoantígenos predichos para dar vacunas contra el cáncer de ARNm-lipoplejo y se administraron a ratones C57BL/6 (60 µg por dosis, dos veces a la semana a través de inyección i.v.) para evaluar la inmunogenicidad a través de ELISPOT de IFNγ de esplenocitos (véase la figura 5). Como control positivo, se administró mTRP2 (proteína relacionada con tirosinasa 2) (antígeno asociado a tumor de melanoma; SVYDFFVWL (SEQ ID NO: 134)) a un modelo de ratón usando el mismo método descrito previamente. Se estableció el umbral para determinar la inmunogenicidad de los neoantígenos predichos.

Las 50 secuencias mejores de neoantígeno usadas en este ejemplo se enumeran en la tabla 9:

TABLA 9

50 secuencias mejores de B16F10

Clasificación	Secuencia de neoantígeno de B16F10	SEQ ID NO
1	MSLPFSFL	84
2	ANFIFRQL	85
3	STYKLTNL	86
4	NMYGFRNV	87

5	VIYSHLQV	88
6	SGFRYNVL	89
7	VVMEYENL	90
8	IVHYFIAL	91
9	VSFAPLVQL	92
10	ATFYFHHPV	93
11	VHYLRSL	94
12	IGIFFQMRV	95
13	VAHVAAFL	96
14	VFLRVRAL	97
15	TIFMMLRL	98
16	KGFLYSAL	99
17	SSVLFKGL	100
18	MQFYLRHL	101
19	ARMMFSGL	102
20	KSLYYASCL	103
21	LVFVVGRL	104
22	SIGVYIPL	105
23	TVYRCGHL	106
24	SLFTLHQSV	107
25	FTFVYPTI	108
26	CMLRYQTCL	109
27	VCFDYTFAL	110
28	LVYYLPL	111
29	VPYAVFPPL	112
30	VSFANAVV	113
31	VVPSFSAEI	114
32	CNFCYQRNQL	115
33	QMYPAFKRV	116
34	ILVSTYRPL	117
35	SFLAGFRPL	118
36	AIYHHASRAI	119
37	IAMGFPVERL	120
38	VLILHAAAL	121
39	SLVAPFGSV	122
40	MYRYFVTL	123

41	VMMDNFDYL	124
42	VTFQAFIDV	125
43	RAFAAVRL	126
44	KNFEFNKQRSIV	127
45	LMYVYEMLGAGL	128
46	CMLRYQTCLV	129
47	AIALHLHL	130
48	TSYSTAAPL	131
49	SMPHYSSPAI	132
50	TIPFYDNPI	133

Los resultados obtenidos se muestran en la figura 6. Tal como se representa en la figura 6, se confirmó la inmunogenicidad para el 45 % (9/20) de los 20 neoantígenos predichos mejores (véase la parte superior de la figura 6). De los 12 neoantígenos con alta inmunogenicidad de entre los 50 mejores, 9 estaban dentro de los 20 mejores, lo que indica una alta probabilidad de seleccionar como diana neoantígenos clave cuando se centra en los 20 mejores (véase la tabla 10).

TABLA 10

Clasificación	Positivo
20 mejores	9 (45 %)
21 mejores ~ 50 mejores	3 (10 %)
Total	12 (24 %)

(Umbral de inmunogenicidad: > 3x el promedio de ELISPOT con medio)

10

EJEMPLO 6: Eficacia *in vivo*: efectividad antitumoral de neoantígeno predicho

Se implantaron por vía subcutánea un total de 5×10^4 células tumorales B16F10 en ratones C57BL/6, y se realizó la separación de grupos cuando el tamaño de tumor alcanzó 50 mm³. Los grupos experimentales incluían un grupo que recibió ARNm-lipoplejo cargado con los 10 neoantígenos de B16F10 predichos mejores (seleccionados de los 12 neoantígenos mejores en el ejemplo 5.2), un grupo de control positivo que

recibió ARNm-lipoplejo cargado con mTRP2 (proteína relacionada con tirosinasa 2) (antígeno asociado a tumor de melanoma) (SEQ ID NO: 134), y un grupo de control negativo que recibió lipoplejo sin ningún ARNm de neoantígeno cargado. Se administró ARNm-lipoplejo a una dosis de 60 µg por inyección, tres veces en intervalos de cinco días, registrándose los tamaños de tumor durante tres semanas.

Los resultados obtenidos (tamaño de tumor) se muestran en la figura 7 (vacuna contra el cáncer de neoantígeno: grupo que recibió 10 neoantígenos predichos; control positivo: grupo que recibió mTRP2; control: grupo que recibió únicamente lipoplejo (control negativo)). Tal como se muestra en la figura 7, se observó una reducción del 63 % en el tamaño de tumor en el grupo tratado con ARNm-lipoplejo que contenía los 10 neoantígenos de B16F10 predichos mejores.

Los expertos habituales en la técnica reconocerán que la presente invención puede realizarse de otras formas específicas sin apartarse de su espíritu o características esenciales. Las realizaciones descritas deben considerarse en todos los aspectos únicamente como ilustrativas y no restrictivas. Por tanto, el alcance de la presente invención se indica por las reivindicaciones adjuntas en lugar de por la descripción anterior. Todos los cambios que entran dentro del significado e intervalo de equivalencia de las reivindicaciones deben englobarse dentro del alcance de la presente invención.