

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 871 384**

51 Int. Cl.:

A61P 35/00 (2006.01)

G16B 20/00 (2009.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **07.05.2014 PCT/EP2014/001232**

87 Fecha y número de publicación internacional: **13.11.2014 WO14180569**

96 Fecha de presentación y número de la solicitud europea: **07.05.2014 E 14722117 (0)**

97 Fecha y número de publicación de la concesión europea: **10.03.2021 EP 2994159**

54 Título: **Predicción de inmunogenicidad de los epítomos de células T**

30 Prioridad:

10.05.2013 WO PCT/EP2013/001400

45 Fecha de publicación y mención en BOPI de la traducción de la patente:

28.10.2021

73 Titular/es:

**BIONTECH RNA PHARMACEUTICALS GMBH
(50.0%)**

An der Goldgrube 12

55131 Mainz, DE y

TRON - TRANSLATIONALE ONKOLOGIE AN DER

UNIVERSITÄTSMEDIZIN DER JOHANNES

GUTENBERG- UNIVERSITÄT MAINZ

GEMEINNÜTZIGE GMBH (50.0%)

72 Inventor/es:

SAHIN, UGUR;

TADMOR, ARBEL DAVID;

CASTLE, JOHN CHRISTOPHER;

BOEGEL, SEBASTIAN y

LÖWER, MARTIN

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 871 384 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín Europeo de Patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre Concesión de Patentes Europeas).

DESCRIPCIÓN

Predicción de inmunogenicidad de los epítomos de células T

5 Campo técnico

La presente enseñanza se refiere a métodos para predecir epítomos de células T. En particular, la presente enseñanza se refiere a métodos para predecir si las modificaciones en péptidos o polipéptidos tales como neoantígenos asociados a tumores son inmunogénicas o no. Los métodos de la enseñanza son útiles, en particular, para la provisión de vacunas que son específicas para el tumor de un paciente y, por lo tanto, en el contexto de vacunas personalizadas contra el cáncer.

Antecedentes

15 Las vacunas personalizadas contra el cáncer son vacunas terapéuticas diseñadas a medida para dirigirse a mutaciones específicas de tumores que son exclusivas de un paciente determinado. Este tratamiento ofrece una gran esperanza para los pacientes con cáncer, ya que no daña las células sanas y tiene el potencial de proporcionar una remisión de por vida. Sin embargo, no todas las mutaciones expresadas por el tumor pueden usarse como diana para una vacuna. De hecho, la mayoría de las mutaciones somáticas del cáncer no conducirán a una respuesta inmune cuando se vacunen (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. *Cancer Research* 72, 1081 (2012)). Dado que los tumores pueden codificar hasta 100.000 mutaciones somáticas (MR Stratton, *Science Signaling* 331, 1553 (2011)) mientras que las vacunas se dirigen solo a un puñado de epítomos, es evidente que un objetivo fundamental de la inmunoterapia del cáncer es identificar qué mutaciones es probable que sean inmunogénicas.

25 Desde una perspectiva biológica, para que una mutación somática genere una respuesta inmune, se deben cumplir varios criterios: el alelo que contiene la mutación debe ser expresado por la célula, la mutación debe estar en una región codificadora de proteínas y no sinónima, la proteína traducida debe ser escindida por el proteasoma y un epítomo que contiene la mutación debe ser presentado por el complejo MHC, el epítomo presentado debe ser reconocido por un receptor de células T (TCR) y, finalmente, el complejo TCR-pMHC debe lanzar una señalización en cascada que active las células T (S. Whelan, N. Goldman, *Molecular biology and evolution* 18, 691 (2001)). Hasta ahora no se ha presentado ningún algoritmo que sea capaz de predecir con un alto grado de certeza qué mutaciones es probable que cumplan con todos estos criterios. En el presente informe se consideran varios factores que pueden contribuir a la inmunogenicidad, comparando estos factores con datos experimentales y proponiendo un modelo simple para identificar mutaciones inmunogénicas.

35 Predicción de la unión del MHC: Estado de la técnica

Hace más de 20 años, se estableció que hay posiciones en un péptido de unión al MHC que contribuyen más a la capacidad de unión que otras (por ejemplo, (A. Sette et al., *Proceedings of the National Academy of Sciences* 86, 3296 (1989))). La identificación y descripción de esas posiciones de anclaje permitió encontrar patrones de péptidos de unión a MHC y, por lo tanto, fueron la base para desarrollar métodos de predicción. En los últimos años se han logrado avances significativos en el campo de los modelos *in silico* de la maquinaria de procesamiento de antígenos. Los dos enfoques pioneros, que se desarrollaron a fines de la década de 1990, BIMAS (KC Parker, MA Bednarek, JE Coligan, *The Journal of Immunology* 152, 163 (1994)) y SYFPEITHI (H.-G. Rammensee, J. Bachmann, NPN Emmerich, OA Bachor, S. Stevanovic, *Immunogenetics* 50, 213 (1999)), se basaron en el conocimiento de las posiciones de anclaje y en los motivos específicos de alelos derivados. A medida que se dispuso de más y más datos experimentales de unión de péptidos al MHC, se han desarrollado más herramientas utilizando una amplia variedad de técnicas estadísticas e informáticas (vease la Fig. 1 para una descripción general). Los denominados métodos basados en matrices utilizan matrices de puntuación específicas de la posición para determinar si una secuencia de péptidos coincide con el motivo de unión de un alelo del MHC particular. Otra clase de métodos de predicción de unión al MHC utilizan técnicas de aprendizaje automático, tal como las redes neuronales artificiales o las máquinas de vectores de soporte (véase la Figura 1). El rendimiento de estos algoritmos depende en gran medida de la cantidad y la calidad del conjunto de datos de entrenamiento disponible para cada modelo de alelo (por ejemplo, "HLA-A*02:01", "H2-Db", etc.) para "aprender" patrones/características subyacentes, que tienen capacidad de predicción de la unión. Recientemente, están surgiendo métodos basados en estructuras, que evitan el cuello de botella de tener un gran conjunto de entrenamiento, ya que se basan únicamente en estructuras cristalinas péptido-MHC y funciones de puntuación (por ejemplo, diferentes funciones de energía) para predecir interacciones péptido-MHC, por ejemplo mediante minimización de energía (véase la Fig. 1). Sin embargo, la precisión de esos enfoques todavía está muy por detrás de los métodos basados en secuencias. Los estudios de evaluación comparativa muestran que la herramienta basada en redes neuronales artificiales NetMHC (C. Lundegaard et al., *Nucleic Acids Research* 36, W509 (2008)) y el algoritmo basado en matrices SMM (B. Peters, A. Sette, *BMC bioinformatics* 6, 132 (2005)) obtienen mejores resultados en los datos de evaluación probados (B. Peters, A. Sette, *BMC bioinformatics* 6, 132 (2005); H.H. Lin, S. Ray, S. Tongchusak, EL Reinherz, V. Brusica, *BMC Immunology* 9, 8 (2008)). Ambos enfoques están integrados en los llamados métodos de consenso IEDB, disponibles en la base de datos de epítomos inmunes (Y. Kim et al., *Nucleic Acids Research* 40, W525 (2012)). El modelado de interacciones de unión péptido-MHC II es mucho más complejo que para MHC I, ya que las moléculas del MHC II poseen un surco de unión con extremos abiertos en cada lado,

5 permitiendo la unión de péptidos de diferentes longitudes. Mientras que los péptidos que se unen al MHC I están restringidos principalmente a 8-12 aminoácidos, esta longitud puede diferir dramáticamente para los péptidos del MHC II (9-30 aminoácidos). Un estudio de evaluación comparativa reciente muestra que los métodos de predicción del MHC II disponibles ofrecen una precisión limitada en comparación con la predicción del MHC I (H.H. Lin, S. Ray, S. Tongchusak, EL Reinherz, V. Brusic, BMC Immunology 9, 8 (2008)).

10 El primer uso sistemático y a gran escala de esos algoritmos para encontrar epítopos de células T fue realizado por Moutaftsi et al., (M. Moutaftsi et al., Nature Biotechnology 24, 817 (2006)), en el que se combinaron diferentes herramientas para predecir posibles candidatos a vacuna de ratones C57BL/6 infectados con el virus vaccinia, se extrajeron esplenocitos y se midieron las respuestas de las células T CD8+ contra el 1% superior de los péptidos predichos. Identificaron 49 (de 2256) péptidos que indujeron una respuesta de células T. Desde entonces se han publicado muchos estudios utilizando diversas herramientas de predicción de la unión del MHC para buscar epítopos de células T como candidatos para una vacuna, principalmente para patógenos, por ejemplo, Leishmania major (C. Herrera-Najera, R. Piña-Aguilar, F. Xacur- García, M.J. Ramírez-Sierra, E. Dumonteil, Proteomics 9, 1293 (2009)). Sin embargo, utilizar únicamente herramientas de predicción de la unión del MHC I para la predicción de la inmunogenicidad es engañoso, ya que esas herramientas están entrenadas para predecir si un péptido dado tiene el potencial de unirse a un alelo del MHC dado. El fundamento de usar predicciones de unión del MHC para predecir la inmunogenicidad es la suposición de que los péptidos que se unen con alta afinidad a un alelo del MHC respectivo tienen más probabilidades de ser inmunogénicos (A. Sette et al., The Journal of Immunology 153, 5586 (1994)). Sin embargo, hay numerosos estudios que indican que también una baja afinidad de unión al MHC puede resultar en una alta inmunogenicidad (M.C. Feltkamp, M.P. Vierboom, W.M. Kast, C.J. Melief, Molecular Immunology 31, 1391 (1994)) y que la estabilidad del péptido-MHC puede ser un mejor predictor de inmunogenicidad que la afinidad por el péptido (M. Harndahl et al., European Journal of Immunology 42, 1405 (2012)). Por esa razón, la predicción de la inmunogenicidad no fue muy precisa hasta ahora, lo que se refleja en las bajas tasas de éxito para predecir la inmunogenicidad. No obstante, la unión de péptidos es una condición necesaria pero no suficiente para el reconocimiento de epítopos de células T, y una predicción eficaz puede reducir drásticamente el número de péptidos que se van a probar experimentalmente. Está claro que el desarrollo de un modelo que predice la inmunogenicidad también debe tener en cuenta el reconocimiento del receptor de células T (TCR) así como la tolerancia central, es decir, la selección negativa y positiva de células T durante el desarrollo en el timo.

30 El documento de la técnica anterior Castle et al., 2012 (Cancer Res; 72: 1081-1091) describe un proceso de selección de mutaciones para pruebas de validación e inmunogenicidad. El documento de la técnica anterior Nielsen et al., 2007 describe un método bioinformático que tiene en cuenta tanto la información de la secuencia de péptidos como de HLA, y genera predicciones cuantitativas de la afinidad de cualquier interacción péptido-HLA-I (PLoS ONE 2: e796). Sin embargo, ninguno de los documentos de la técnica anterior citados anteriormente en este documento enseña un método como se define en las reivindicaciones de la presente solicitud.

40 Existe la necesidad de un modelo predictivo, que sea capaz de modelar todos los aspectos mencionados anteriormente para predecir con precisión la inmunogenicidad de un epítopo, en lugar de solo la unión.

Descripción

Resumen

45 La presente invención está definida por las reivindicaciones adjuntas.

En un aspecto, la presente enseñanza se refiere a un método para predecir modificaciones de aminoácidos inmunogénicos, comprendiendo el método las etapas de:

- 50 a) determinar una puntuación para la unión de un péptido modificado a una o más moléculas del MHC, y
 b) determinar una puntuación para la unión del péptido no modificado a una o más moléculas del MHC, y/o
 55 c) determinar una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T.

60 En una realización, el péptido modificado comprende un fragmento de una proteína modificada, comprendiendo dicho fragmento la modificación o modificaciones presentes en la proteína. En una realización, el péptido o proteína no modificada tiene el aminoácido de la línea germinal en la posición o posiciones correspondientes a la posición o posiciones de la modificación o modificaciones en el péptido o proteína modificada.

En una realización, el péptido o proteína no modificada y el péptido o proteína modificada son idénticos excepto por la modificación o modificaciones. Preferiblemente, el péptido o proteína no modificada y el péptido o proteína modificada tienen la misma longitud y/o secuencia (excepto por la modificación o modificaciones).

65

En una realización, el péptido no modificado y el péptido modificado tienen de 8 a 15, preferiblemente de 8 a 12 aminoácidos de longitud.

5 En una realización, la una o más moléculas del MHC comprenden diferentes tipos de moléculas del MHC, en particular diferentes alelos del MHC. En una realización, la una o más moléculas del MHC son moléculas del MHC de clase I y/o moléculas del MHC de clase II. En una realización, la una o más moléculas del MHC comprenden un conjunto de alelos del MHC tal como un conjunto de alelos del MHC de un individuo o un subconjunto del mismo.

10 En una realización, la puntuación para la unión a una o más moléculas del MHC se determina mediante un proceso que comprende una comparación de secuencias con una base de datos de motivos de unión al MHC.

15 En una realización, la etapa a) comprende determinar si dicho puntaje satisface un umbral predeterminado para la unión a una o más moléculas del MHC y/o la etapa b) comprende determinar si dicho puntaje satisface un umbral predeterminado para unirse a una o más moléculas del MHC. En una realización, el umbral aplicado en la etapa a) es diferente al umbral aplicado en la etapa b). En una realización, el umbral predeterminado de unión a una o más moléculas del MHC refleja una probabilidad de unión a una o más moléculas del MHC.

20 En una realización, el uno o más receptores de células T comprenden un conjunto de receptores de células T, tal como un conjunto de receptores de células T de un individuo o un subconjunto del mismo. En una realización, la etapa c) comprende asumir que dicho conjunto de receptores de células T no incluye receptores de células T que se unen al péptido no modificado cuando están presentes en un complejo MHC-péptido y/o no incluye receptores de células T que se unen al péptido no modificado cuando está presente en un complejo MHC-péptido con alta afinidad.

25 En una realización, la etapa c) comprende determinar una puntuación para las similitudes químicas y físicas entre los aminoácidos modificados y no modificados. En una realización, la etapa c) comprende determinar si dicha puntuación satisface un umbral predeterminado para las similitudes químicas y físicas entre aminoácidos. En una realización, dicho umbral predeterminado para las similitudes químicas y físicas entre aminoácidos refleja una probabilidad de que los aminoácidos sean química y físicamente similares. En una realización, la puntuación de las similitudes químicas y físicas se determina sobre la base de la probabilidad de que los aminoácidos se intercambien en la naturaleza. En una
30 realización, cuanto más frecuentemente se intercambian los aminoácidos en la naturaleza, más similares se consideran los aminoácidos y viceversa. En una realización, las similitudes químicas y físicas se determinan utilizando matrices logarítmicas de probabilidades basadas en la evolución.

35 En una realización, si el péptido no modificado tiene una puntuación para unirse a una o más moléculas del MHC que satisfacen un umbral que indica la unión a una o más moléculas del MHC y el péptido modificado tiene una puntuación para unirse a una o más moléculas del MHC que satisface un umbral que indica unión a una o más moléculas del MHC, la modificación o péptido modificado se predice como inmunogénico si los aminoácidos no modificados y modificados tienen una puntuación para las similitudes químicas y físicas que satisfacen un umbral que indica disimilitud química y física.

40 En una realización, si el péptido no modificado se une a una o más moléculas del MHC o tiene una probabilidad de unirse a una o más moléculas del MHC y el péptido modificado se une a una o más moléculas del MHC o tiene una probabilidad de unirse a una o más moléculas del MHC, la modificación o el péptido modificado se predice como inmunogénico si los aminoácidos no modificados y modificados son química y físicamente diferentes o tienen una
45 probabilidad de ser química y físicamente diferentes.

En una realización, la modificación no está en una posición de anclaje para unirse a una o más moléculas del MHC.

50 En una realización, si el péptido no modificado tiene una puntuación para unirse a una o más moléculas del MHC que satisfacen un umbral que indica que no se une a una o más moléculas del MHC y el péptido modificado tiene una puntuación para unirse a una o más moléculas del MHC que satisfacen un umbral que indica la unión a una o más moléculas del MHC, se predice que la modificación o el péptido modificado es inmunogénico.

55 En una realización, si el péptido no modificado no se une a una o más moléculas del MHC o tiene una probabilidad de no unirse a una o más moléculas del MHC y el péptido modificado se une a una o más moléculas del MHC o tiene una probabilidad para unirse a una o más moléculas del MHC, se predice que la modificación o el péptido modificado es inmunogénico.

60 En una realización, la modificación está en una posición de anclaje para unirse a una o más moléculas del MHC.

65 En una realización, el método de la enseñanza comprende realizar la etapa a) en dos o más péptidos modificados diferentes, comprendiendo dichos dos o más péptidos modificados diferentes la misma modificación o modificaciones. En una realización, los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones comprenden fragmentos diferentes de una proteína modificada, comprendiendo dichos fragmentos diferentes la misma modificación o modificaciones presentes en la proteína. En una realización, los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones comprenden todos los fragmentos de

unión al MHC potenciales de una proteína modificada, comprendiendo dichos fragmentos la misma modificación o modificaciones presentes en la proteína. En una realización, el método de la enseñanza comprende además seleccionar (el, los) péptido o péptidos modificados de los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones que tienen una probabilidad o tienen la probabilidad más alta de unirse a una o más moléculas del MHC. En una realización, los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones difieren en la longitud y/o la posición de la modificación o modificaciones.

En una realización, el método de la enseñanza comprende realizar la etapa a) y opcionalmente una o ambas etapas b) y c) en dos o más péptidos modificados diferentes. En una realización, dichos dos o más péptidos modificados diferentes comprenden la misma modificación o modificaciones y/o comprenden modificaciones diferentes. En una realización, las diferentes modificaciones están presentes en la misma y/o en diferentes proteínas. El conjunto de dos o más péptidos modificados diferentes usados en la etapa a) y opcionalmente una o ambas etapas b) y c) pueden ser iguales o diferentes. En una realización, el conjunto de dos o más péptidos modificados diferentes usados en la etapa b) y/o la etapa c) es un subconjunto del conjunto de dos o más péptidos modificados diferentes usados en la etapa a). Preferiblemente, dicho subconjunto incluye el péptido o péptidos que puntúan mejor en la etapa a).

En una realización, el método de la enseñanza comprende comparar las puntuaciones de dos o más de dichos péptidos modificados diferentes. En una realización, el método de la enseñanza comprende clasificar dos o más de dichos péptidos modificados diferentes. En una realización, una puntuación para la unión del péptido modificado a una o más moléculas del MHC se pondera más alta que una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T, preferiblemente una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados y una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T, preferiblemente una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados se ponderan más alto que una puntuación para la unión del péptido no modificado a una o más moléculas del MHC.

En una realización, el método de la enseñanza comprende además identificar mutaciones no sinónimas en una o más regiones codificantes de proteínas.

En una realización, las modificaciones se identifican de acuerdo con la enseñanza secuenciando parcial o completamente el genoma o transcriptoma de una o más células tales como una o más células cancerosas y opcionalmente una o más células no cancerosas e identificando mutaciones en una o más regiones codificadoras de proteínas.

En una realización, dichas mutaciones son mutaciones somáticas. En una realización, dichas mutaciones son mutaciones cancerosas.

En una realización, el método de la enseñanza se usa en la fabricación de una vacuna. En una realización, la vacuna se deriva de (a) modificación o modificaciones o (a) péptido o péptidos modificados predichos como inmunogénicos por los métodos de la enseñanza.

En un aspecto adicional, la presente enseñanza proporciona un método para proporcionar una vacuna que comprende la etapa:

identificar (a) modificación (es) o (a) péptido o péptidos modificados predichos como inmunogénicos por los métodos de la enseñanza.

En una realización, el método comprende además la etapa:
proporcionar una vacuna que comprende un péptido o polipéptido que comprende la modificación o las modificaciones o los péptidos modificados predichos como inmunogénicos, o un ácido nucleico que codifica el péptido o polipéptido.

En un aspecto adicional, la presente enseñanza proporciona una vacuna que se puede obtener usando los métodos de acuerdo con la enseñanza. En este documento se describen realizaciones preferidas de tales vacunas.

Una vacuna proporcionada de acuerdo con la enseñanza puede comprender un vehículo farmacéuticamente aceptable y opcionalmente puede comprender uno o más adyuvantes, estabilizadores, etc. La vacuna puede estar en forma de vacuna terapéutica o profiláctica.

Otro aspecto se refiere a un método para inducir una respuesta inmune en un paciente, que comprende administrar al paciente una vacuna proporcionada de acuerdo con la enseñanza.

Otro aspecto se refiere a un método de tratamiento de un paciente con cáncer que comprende las etapas:

(a) proporcionar una vacuna por los métodos de acuerdo con la enseñanza; y
(b) administrar dicha vacuna al paciente.

Otro aspecto se refiere a un método de tratamiento de un paciente con cáncer que comprende administrar la vacuna de acuerdo con la enseñanza al paciente.

5 En aspectos adicionales, la enseñanza proporciona las vacunas descritas en el presente documento para su uso en los métodos de tratamiento descritos en el presente documento, en particular para su uso en el tratamiento o la prevención del cáncer.

10 Los tratamientos del cáncer descritos en este documento se pueden combinar con resección quirúrgica y/o radiación y/o quimioterapia tradicional.

Otras características y ventajas de las presentes enseñanzas serán evidentes a partir de la siguiente descripción detallada y reivindicaciones.

15 Descripción detallada

Aunque la presente enseñanza se describe en detalle a continuación, debe entenderse que esta enseñanza no se limita a las metodologías, protocolos y reactivos particulares descritos en este documento, ya que estos pueden variar. También debe entenderse que la terminología utilizada en este documento tiene el propósito de describir únicamente realizaciones particulares, y no pretende limitar el alcance de la presente enseñanza, que estará limitada únicamente por las reivindicaciones adjuntas. A menos que se defina de otro modo, todos los términos técnicos y científicos usados en este documento tienen los mismos significados que los entendidos comúnmente por un experto en la técnica.

20 A continuación, se describirán los elementos de la presente enseñanza. Estos elementos se enumeran con realizaciones específicas, sin embargo, debe entenderse que pueden combinarse de cualquier manera y en cualquier número para crear realizaciones adicionales. Los ejemplos descritos de forma diversa y las realizaciones preferidas no deben interpretarse como que limitan la presente enseñanza a solo las realizaciones descritas explícitamente. Debe entenderse que esta descripción apoya y engloba realizaciones que combinan las realizaciones descritas explícitamente con cualquier número de los elementos descritos y/o preferidos. Además, cualquier permutación y combinación de todos los elementos descritos en esta solicitud deben considerarse divulgadas por la descripción de la presente solicitud a menos que el contexto indique lo contrario.

25 Preferiblemente, los términos usados en este documento se definen como se describe en A multilingual glossary of biotechnological terms: (IUPAC Recommendations)", H.G.W. Leuenberger, B. Nagel y H. Kölbl, Eds., (1995) Helvetica Chimica Acta, CH-4010 Basilea, Suiza.

35 La práctica de la presente enseñanza empleará, a menos que se indique lo contrario, métodos convencionales de bioquímica, biología celular, inmunología y técnicas de ADN recombinante que se explican en la literatura en el campo (véase, por ejemplo, Molecular Cloning: A laboratory Manual, 2ª edición, J. Sambrook y et al., Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

40 A lo largo de esta memoria descriptiva y las reivindicaciones que siguen, a menos que el contexto requiera lo contrario, la palabra "comprende", y variaciones tales como "que comprende", se entenderá que implica la inclusión de un miembro, número entero o etapa o grupo de miembros, números enteros o etapas pero no la exclusión de cualquier otro miembro, número entero o etapa o grupo de miembros, números enteros o etapas aunque en algunas realizaciones dicho otro miembro, número entero o etapa o grupo de miembros, números enteros o etapas establecidos pueden ser excluidos, es decir, el objeto consiste en la inclusión de un miembro, número entero o etapa o grupo de miembros, números enteros o etapas establecidos. Los términos "un" y "uno, una" y "el, la" y referencias similares utilizadas en el contexto de la descripción de la enseñanza (especialmente en el contexto de las reivindicaciones) deben interpretarse para cubrir tanto el singular como el plural, a menos que se indique lo contrario en este documento o sea claramente contradicho por el contexto. La mención de intervalos de valores en el presente documento está destinada simplemente a servir como un método abreviado para hacer referencia individualmente a cada valor separado que se encuentre dentro del intervalo. A menos que se indique lo contrario en el presente documento, cada valor individual se incorpora en la memoria descriptiva como si se mencionara individualmente en el presente documento.

55 Todos los métodos descritos en el presente documento se pueden realizar en cualquier orden adecuado a menos que se indique lo contrario en el presente documento o que el contexto lo contradiga claramente de otro modo. El uso de todos y cada uno de los ejemplos, o ejemplo de expresión (por ejemplo, "tal como"), proporcionado en el presente documento está destinado simplemente a ilustrar mejor la enseñanza y no supone una limitación en el alcance de la enseñanza reivindicada. Ninguna expresión en la memoria descriptiva debe interpretarse como que indica algún elemento no reivindicado esencial para la práctica de la enseñanza.

60 De acuerdo con la presente enseñanza, el término "péptido" se refiere a sustancias que comprenden dos o más, preferiblemente 3 o más, preferiblemente 4 o más, preferiblemente 6 o más, preferiblemente 8 o más, preferiblemente 10 o más, preferiblemente 13 o más, preferiblemente 16 más, preferiblemente 21 o más y hasta preferiblemente 8, 10, 20, 30, 40 o 50, en particular 100 aminoácidos unidos covalentemente por enlaces peptídicos. El término "polipéptido"

o "proteína" se refiere a péptidos grandes, preferiblemente a péptidos con más de 100 residuos de aminoácidos, pero en general los términos "péptido", "polipéptido" y "proteína" son sinónimos y se usan indistintamente en este documento.

5 De acuerdo con la enseñanza, el término "modificación" con respecto a péptidos, polipéptidos o proteínas se refiere a un cambio de secuencia en un péptido, polipéptido o proteína en comparación con una secuencia parental tal como la secuencia de un péptido, polipéptido o proteína de tipo silvestre. El término incluye variantes de inserción de aminoácidos, variantes de adición de aminoácidos, variantes de eliminación de aminoácidos y variantes de sustitución de aminoácidos, preferiblemente variantes de sustitución de aminoácidos. Todos estos cambios de secuencia de
10 acuerdo con la enseñanza pueden potencialmente crear nuevos epítomos.

Las variantes de inserción de aminoácidos comprenden inserciones de uno o dos o más aminoácidos en una secuencia de aminoácidos particular.

15 Las variantes de adición de aminoácidos comprenden fusiones del terminal amino y/o carboxilo de uno o más aminoácidos, tales como 1, 2, 3, 4 o 5, o más aminoácidos.

Las variantes de eliminación de aminoácidos se caracterizan por la eliminación de uno o más aminoácidos de la secuencia, tal como por la eliminación de 1, 2, 3, 4 o 5, o más aminoácidos.

20 Las variantes de sustitución de aminoácidos se caracterizan porque se elimina al menos un residuo en la secuencia y se inserta otro residuo en su lugar.

De acuerdo con la enseñanza, una modificación o un péptido modificado usado para ensayar en los métodos de la enseñanza puede derivarse de una proteína que comprende una modificación.

El término "derivado" significa de acuerdo con la enseñanza que una entidad particular, en particular una secuencia peptídica particular, está presente en el objeto del que se deriva. En el caso de secuencias de aminoácidos, especialmente regiones de secuencia particulares, "derivadas" significa en particular que la secuencia de aminoácidos
30 relevante se deriva de una secuencia de aminoácidos en la que está presente.

Una proteína que comprende una modificación a partir de la cual se puede derivar una modificación o un péptido modificado usado para probar en los métodos de la enseñanza puede ser un neoantígeno.

35 De acuerdo con la enseñanza, el término "neoantígeno" se refiere a un péptido o proteína que incluye una o más modificaciones de aminoácidos en comparación con el péptido o proteína parental. Por ejemplo, el neoantígeno puede ser un neoantígeno asociado a tumor, en el que el término "neoantígeno asociado a tumor" incluye un péptido o proteína que incluye modificaciones de aminoácidos debido a mutaciones específicas del tumor.

40 De acuerdo con la enseñanza, el término "mutación específica del tumor" o "mutación específica de cáncer" se refiere a una mutación somática que está presente en el ácido nucleico de un tumor o célula cancerosa pero ausente en el ácido nucleico de una célula correspondiente normal, es decir, no tumoral o no cancerosa. Los términos "mutación específica del tumor" y "mutación del tumor" y los términos "mutación específica del cáncer" y "mutación del cáncer" se usan indistintamente en el presente documento.

45 El término "respuesta inmune" se refiere a una respuesta corporal integrada a una diana tal como un antígeno y preferiblemente se refiere a una respuesta inmune celular o una respuesta inmune celular así como también humoral. La respuesta inmune puede ser protectora/preventiva/profiláctica y/o terapéutica.

50 "Inducir una respuesta inmune" puede significar que no hubo respuesta inmune antes de la inducción, pero también puede significar que hubo un cierto nivel de respuesta inmune antes de la inducción y después de la inducción se potencia dicha respuesta inmune. Por lo tanto, "inducir una respuesta inmunitaria" también incluye "potenciar una respuesta inmunitaria". Preferiblemente, después de inducir una respuesta inmune en un sujeto, dicho sujeto está protegido de desarrollar una enfermedad tal como una enfermedad cancerosa o la condición de la enfermedad mejora induciendo una respuesta inmune. Por ejemplo, se puede inducir una respuesta inmune contra un antígeno expresado
55 en un tumor en un paciente que padece una enfermedad cancerosa o en un sujeto que tiene riesgo de desarrollar una enfermedad cancerosa. Inducir una respuesta inmune en este caso puede significar que el estado de enfermedad del sujeto mejora, que el sujeto no desarrolla metástasis o que el sujeto que está en riesgo de desarrollar una enfermedad cancerosa no desarrolla una enfermedad cancerosa.

60 Los términos "respuesta inmune celular" y "respuesta celular" o términos similares se refieren a una respuesta inmune dirigida a células caracterizada por la presentación de un antígeno con MHC clase I o clase II que involucra células T o linfocitos T que actúan como "colaboradoras" o "asesinos". Las células T colaboradoras (también denominadas células T CD4⁺) desempeñan un papel central al regular la respuesta inmune y las células asesinas (también denominadas células T citotóxicas, células T citolíticas, células T CD8⁺ o CTL) matan las células enfermas tal como las células cancerosas, evitando la producción de más células enfermas. En realizaciones preferidas, la presente
65

enseñanza implica la estimulación de una respuesta CTL antitumoral contra células tumorales que expresan uno o más antígenos expresados en tumores y que presentan preferiblemente tales antígenos expresados en tumores con MHC clase I.

5 Un "antígeno" de acuerdo con la enseñanza cubre cualquier sustancia, preferiblemente un péptido o proteína, que es un objetivo de y/o induce una respuesta inmune tal como una reacción específica con anticuerpos o linfocitos T (células T). Preferiblemente, un antígeno comprende al menos un epítipo tal como un epítipo de células T. Preferiblemente, un antígeno en el contexto de la presente enseñanza es una molécula que, opcionalmente después del procesamiento, induce una reacción inmune, que preferiblemente es específica para el antígeno (incluidas las células que expresan el antígeno). El antígeno o un epítipo de células T del mismo se presenta preferiblemente por una célula, preferiblemente por una célula presentadora de antígeno que incluye una célula enferma, en particular una célula cancerosa, en el contexto de moléculas del MHC, lo que da como resultado una respuesta inmune contra el antígeno (incluidas las células que expresan el antígeno).

15 En una realización, un antígeno es un antígeno tumoral (también denominado antígeno expresado por tumor en este documento), es decir, una parte de una célula tumoral tal como una proteína o péptido expresado en una célula tumoral que puede derivarse del citoplasma, la superficie celular o el núcleo celular, en particular aquellos que se encuentran principalmente intracelularmente o como antígenos de superficie de células tumorales. Por ejemplo, los antígenos tumorales incluyen el antígeno carcinoembrionario, α 1-fetoproteína, isoferritina y sulfoglicoproteína fetal, α 2-H-ferroproteína y γ -fetoproteína. De acuerdo con la presente enseñanza, un antígeno tumoral comprende preferiblemente cualquier antígeno que se exprese y, opcionalmente, sea característico con respecto al tipo y/o nivel de expresión para tumores o cánceres así como para células tumorales o cancerosas, es decir, un antígeno asociado a tumores. En una realización, el término "antígeno asociado a tumores" se refiere a proteínas que están en condiciones normales expresadas específicamente en un número limitado de tejidos y/u órganos o en etapas específicas de desarrollo, por ejemplo, los antígenos asociados a tumores pueden estar por debajo de las condiciones normales expresadas específicamente en el tejido del estómago, preferiblemente en la mucosa gástrica, en los órganos reproductores, por ejemplo, en los testículos, en el tejido trofoblástico, por ejemplo, en la placenta, o en las células de la línea germinal, y se expresan o expresan de manera aberrante en uno o más tumores o tejidos cancerosos. En este contexto, "un número limitado" significa preferiblemente no más de 3, más preferiblemente no más de 2. Los antígenos tumorales en el contexto de la presente enseñanza incluyen, por ejemplo, antígenos de diferenciación, preferiblemente antígenos de diferenciación específicos del tipo celular, es decir, proteínas que están en condiciones normales expresadas específicamente en un cierto tipo de células en una determinada etapa de diferenciación, antígenos de cáncer/testículo, es decir, proteínas que están en condiciones normales expresadas específicamente en testículos y, a veces, en placenta, y antígenos específicos de la línea germinal. Preferiblemente, el antígeno tumoral o la expresión aberrante del antígeno tumoral identifica células cancerosas. En el contexto de la presente enseñanza, el antígeno tumoral que expresa una célula cancerosa en un sujeto, por ejemplo, un paciente que padece una enfermedad cancerosa, es preferiblemente una autoproteína en dicho sujeto. En realizaciones preferidas, el antígeno tumoral en el contexto de la presente enseñanza se expresa en condiciones normales específicamente en un tejido u órgano que no es esencial, es decir, tejidos u órganos que cuando son dañados por el sistema inmunológico no conducen a la muerte del sujeto, o en órganos o estructuras del cuerpo que no son o sólo difícilmente accesibles por el sistema inmunológico.

De acuerdo con la enseñanza, los términos "antígeno tumoral", "antígeno expresado por el tumor", "antígeno canceroso" y "antígeno expresado por cáncer" son equivalentes y se usan indistintamente en el presente documento.

45 El término "inmunogenicidad" se refiere a la efectividad relativa para inducir una respuesta inmune que se asocia preferiblemente con tratamientos terapéuticos, tales como tratamientos contra cánceres. Como se usa en este documento, el término "inmunogénico" se refiere a la propiedad de tener inmunogenicidad. Por ejemplo, el término "modificación inmunogénica" cuando se usa en el contexto de un péptido, polipéptido o proteína se refiere a la efectividad de dicho péptido, polipéptido o proteína para inducir una respuesta inmune causada por y/o dirigida contra dicha modificación. Preferiblemente, el péptido, polipéptido o proteína no modificados no induce una respuesta inmune, induce una respuesta inmune diferente o induce un nivel diferente, preferiblemente un nivel más bajo, de respuesta inmune.

55 Los términos "complejo mayor de histocompatibilidad" y la abreviatura "MHC" incluyen moléculas del MHC de clase I y MHC de clase II y se refieren a un complejo de genes que se encuentra en todos los vertebrados. Las proteínas o moléculas del MHC son importantes para la señalización entre linfocitos y células presentadoras de antígeno o células enfermas en reacciones inmunes, en las que las proteínas o moléculas del MHC se unen a péptidos y los presentan para su reconocimiento por receptores de células T. Las proteínas codificadas por el MHC se expresan en la superficie de las células y presentan tanto antígenos propios (fragmentos de péptidos de la propia célula) como antígenos no propios (por ejemplo, fragmentos de microorganismos invasores) a una célula T.

65 La región del MHC se divide en tres subgrupos, clase I, clase II y clase III. Las proteínas del MHC de clase I contienen una cadena α y una microglobulina β 2 (que no forma parte del MHC codificado por el cromosoma 15). Presentan fragmentos de antígeno a las células T citotóxicas. En la mayoría de las células del sistema inmunológico, específicamente en las células presentadoras de antígenos, las proteínas del MHC de clase II contienen cadenas α

β y presentan fragmentos de antígeno a las células T colaboradoras. La región del MHC de clase III codifica otros componentes inmunitarios, tales como los componentes del complemento y algunos que codifican citocinas.

5 El MHC es poligénico (hay varios genes del MHC de clase I y MHC de clase II) y polimórfico (hay múltiples alelos de cada gen).

10 Como se usa en el presente documento, el término "haplotipo" se refiere a los alelos de HLA que se encuentran en un cromosoma y las proteínas codificadas por ellos. El haplotipo también puede referirse al alelo presente en cualquier locus dentro del MHC. Cada clase del MHC está representada por varios loci: por ejemplo, HLA-A (antígeno leucocitario humano A), HLA-B, HLA-C, HLA-E, HLA-F, HLA-G, HLA-H, HLA-J, HLA-K, HLA-L, HLA-P y HLA-V para clase I y HLA-DRA, HLA-DRB 1-9, HLA, HLA-DQA1, HLA-DQB 1, HLA-DPA1, HLA-DPB1, HLA-DMA, HLA-DMB, HLA-DOA y HLA-DOB para la clase II. Los términos "alelo del HLA" y "alelo del MHC" se usan indistintamente en el presente documento.

15 Los MHC exhiben un polimorfismo extremo: dentro de la población humana hay, en cada locus genético, un gran número de haplotipos que comprenden alelos distintos. Diferentes alelos polimórficos del MHC, tanto de clase I como de clase II, tienen diferentes especificidades de péptidos: cada alelo codifica proteínas que se unen a péptidos que exhiben patrones de secuencia particulares.

20 En una realización preferida de todos los aspectos de la enseñanza, una molécula del MHC es una molécula de HLA.

25 En el contexto de la presente enseñanza, el término "péptido de unión a MHC" incluye péptidos de unión a MHC de clase I y/o clase II o péptidos que pueden procesarse para producir péptidos de unión a MHC de clase I y/o clase II. En el caso de complejos del MHC de clase I/péptido, los péptidos de unión son típicamente de 8 a 12, preferiblemente de 8 a 10 aminoácidos de longitud, aunque los péptidos más largos o más cortos pueden ser eficaces. En el caso de complejos del MHC de clase II/péptido, los péptidos de unión tienen típicamente de 9 a 30, preferiblemente de 10 a 25 aminoácidos de longitud y en particular de 13 a 18 aminoácidos de longitud, mientras que los péptidos más largos y más cortos pueden ser eficaces.

30 Si un péptido se va a presentar directamente, es decir, sin procesamiento, en particular sin escisión, tiene una longitud que es adecuada para unirse a una molécula del MHC, en particular a una molécula del MHC de clase I, y preferiblemente es de 7-30 aminoácidos de longitud tal como 7-20 aminoácidos de longitud, más preferiblemente 7-12 aminoácidos de longitud, más preferiblemente 8-11 aminoácidos de longitud, en particular 9 o 10 aminoácidos de longitud.

35 Si un péptido es parte de una entidad más grande que comprende secuencias adicionales, por ejemplo de una secuencia de vacuna o polipéptido, y se presentará después del procesamiento, en particular después de la escisión, el péptido producido por procesamiento tiene una longitud que es adecuada para unirse a una molécula del MHC, en particular una molécula del MHC de clase I, y preferiblemente es de 7-30 aminoácidos de longitud tal como de 7-20 aminoácidos de longitud, más preferiblemente de 7-12 aminoácidos de longitud, más preferiblemente de 8-11 aminoácidos de longitud, en particular de 9 o 10 aminoácidos de longitud. Preferiblemente, la secuencia del péptido que se presentará después del procesamiento se deriva de la secuencia de aminoácidos de un antígeno o polipéptido usado para la vacunación, es decir, su secuencia corresponde sustancialmente y es preferiblemente completamente idéntica a un fragmento del antígeno o polipéptido.

45 Por lo tanto, un péptido de unión al MHC en una realización comprende una secuencia que corresponde sustancialmente y es preferiblemente completamente idéntica a un fragmento de un antígeno.

50 El término "epítipo" se refiere a un determinante antigénico en una molécula tal como un antígeno, es decir, a una parte o fragmento de la molécula que es reconocida por el sistema inmunológico, por ejemplo, que es reconocida por una célula T, en particular cuando se presenta en el contexto de moléculas del MHC. Un epítipo de una proteína tal como un antígeno tumoral comprende preferiblemente una porción continua o discontinua de dicha proteína y está preferiblemente entre 5 y 100, preferiblemente entre 5 y 50, más preferiblemente entre 8 y 30, lo más preferiblemente entre 10 y 25 aminoácidos de longitud, por ejemplo, el epítipo puede tener preferiblemente 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 o 25 aminoácidos de longitud. Se prefiere particularmente que el epítipo en el contexto de la presente enseñanza sea un epítipo de células T.

60 De acuerdo con la enseñanza, un epítipo puede unirse a moléculas del MHC tales como moléculas del MHC en la superficie de una célula y, por lo tanto, puede ser un "péptido de unión al MHC".

65 Como se usa en el presente documento, el término "neoepítipo" se refiere a un epítipo que no está presente en una referencia tal como una célula normal no cancerosa o de línea germinal pero que se encuentra en células cancerosas. Esto incluye, en particular, situaciones en las que en una célula normal no cancerosa o de la línea germinal se encuentra un epítipo correspondiente; sin embargo, debido a una o más mutaciones en una célula cancerosa, la secuencia del epítipo se cambia para dar como resultado el neoepítipo.

Como se usa en este documento, el término "epítipo de células T" se refiere a un péptido que se une a una molécula del MHC en una configuración reconocida por un receptor de células T. Normalmente, los epítipos de células T se presentan en la superficie de una célula presentadora de antígeno.

5 Como se usa en este documento, el término "predecir epítipos de células T" se refiere a una predicción de si un péptido se unirá a una molécula del MHC y será reconocido por un receptor de células T. El término "predecir epítipos de células T" es esencialmente sinónimo de la frase "predecir si un péptido es inmunogénico".

10 De acuerdo con la enseñanza, un epítipo de células T puede estar presente en una vacuna como parte de una entidad mayor, tal como una secuencia de vacuna y/o un polipéptido que comprende más de un epítipo de células T. El péptido o epítipo de células T presentado se produce siguiendo un procesamiento adecuado.

15 Los epítipos de células T pueden modificarse en uno o más residuos que no son esenciales para el reconocimiento de TCR o para la unión al MHC. Dichos epítipos de células T modificados pueden considerarse inmunológicamente equivalentes.

20 Preferiblemente, un epítipo de células T cuando se presenta por el MHC y es reconocido por un receptor de células T es capaz de inducir en presencia de señales coestimuladoras apropiadas, expansión clonal de la célula T que transporta al receptor de células T reconociendo específicamente el complejo péptido/MHC.

Preferiblemente, un epítipo de células T comprende una secuencia de aminoácidos que corresponde sustancialmente con la secuencia de aminoácidos de un fragmento de un antígeno. Preferiblemente, dicho fragmento de un antígeno es un péptido presentado por el MHC de clase I y/o clase II.

25 Un epítipo de células T de acuerdo con la enseñanza se refiere preferiblemente a una porción o fragmento de un antígeno que es capaz de estimular una respuesta inmune, preferiblemente una respuesta celular contra el antígeno o células caracterizadas por la expresión del antígeno y preferiblemente por la presentación del antígeno tal como células enfermas, en particular células cancerosas. Preferiblemente, un epítipo de células T es capaz de estimular una respuesta celular contra una célula caracterizada por la presentación de un antígeno con MHC de clase I y preferiblemente es capaz de estimular un linfocito T citotóxico sensible al antígeno (CTL).

30 "Procesamiento de antígenos" o "procesamiento" se refiere a la degradación de un péptido, polipéptido o proteína en productos de procesación, que son fragmentos de dicho péptido, polipéptido o proteína (por ejemplo, la degradación de un polipéptido en péptidos) y la asociación de uno o más de estos fragmentos (por ejemplo, mediante unión) con moléculas del MHC para su presentación por células, preferiblemente células presentadoras de antígenos, a células T específicas.

35 Las "células presentadoras de antígeno" (APC) son células que presentan fragmentos peptídicos de antígenos proteicos en asociación con moléculas del MHC en su superficie celular. Algunas APC pueden activar células T específicas de antígeno.

40 Las células presentadoras de antígeno profesionales son muy eficaces para internalizar el antígeno, ya sea por fagocitosis o por endocitosis mediada por receptor, y luego exhiben un fragmento del antígeno, unido a una molécula del MHC de clase II, en su membrana. La célula T reconoce e interactúa con el complejo antígeno-molécula del MHC de clase II en la membrana de la célula presentadora de antígeno. Luego, la célula presentadora de antígeno produce una señal coestimuladora adicional, que conduce a la activación de la célula T. La expresión de moléculas coestimuladoras es una característica definitoria de las células presentadoras de antígenos profesionales.

45 Los principales tipos de células presentadoras de antígeno profesionales son las células dendríticas, que tienen la gama más amplia de presentación de antígenos, y son probablemente las células presentadoras de antígenos, macrófagos, células B y ciertas células epiteliales activadas más importantes. Las células dendríticas (DC) son poblaciones de leucocitos que presentan antígenos capturados en tejidos periféricos a las células T a través de las vías de presentación de antígenos del MHC de clase II y I. Es bien sabido que las células dendríticas son potentes inductores de respuestas inmunes y la activación de estas células es una etapa crítica para la inducción de inmunidad antitumoral. Las células dendríticas se clasifican convenientemente como células "inmaduras" y "maduras", que pueden usarse como una forma sencilla de discriminar entre dos fenotipos bien caracterizados. Sin embargo, no debe interpretarse que esta nomenclatura excluye todas las posibles etapas intermedias de diferenciación. Las células dendríticas inmaduras se caracterizan por ser células presentadoras de antígenos con una alta capacidad de captación y procesamiento de antígenos, lo que se correlaciona con la alta expresión del receptor de Fcγ y del receptor de manosa. El fenotipo maduro se caracteriza típicamente por una menor expresión de estos marcadores, pero una alta expresión de las moléculas de la superficie celular responsables de la activación de las células T, tal como MHC de clase I y II, moléculas de adhesión (por ejemplo, CD54 y CD11) y moléculas coestimuladoras (por ejemplo, CD40, CD80, CD86 y 4-1 BB). La maduración de las células dendríticas se conoce como el estado de activación de las células dendríticas en el que dichas células dendríticas presentadoras de antígenos conducen al cebado de las células T, mientras que la presentación por las células dendríticas inmaduras produce tolerancia. La maduración de las células dendríticas es causada principalmente por biomoléculas con características microbianas detectadas por receptores

innatos (ADN bacteriano, ARN viral, endotoxina, etc.), citocinas proinflamatorias (TNF, IL-1, IFN), ligadura de CD40 en la superficie de las células dendríticas por CD40L, y sustancias liberadas por las células que sufren una muerte celular estresante. Las células dendríticas pueden obtenerse cultivando células de médula ósea *in vitro* con citocinas, tales como el factor estimulante de colonias de granulocitos y macrófagos (GM-CSF) y el factor de necrosis tumoral alfa.

Las células presentadoras de antígenos no profesionales no expresan constitutivamente las proteínas del MHC de clase II necesarias para la interacción con células T no alteradas; estos se expresan solo tras la estimulación de las células presentadoras de antígenos no profesionales por ciertas citocinas tales como IFN γ .

Las células presentadoras de antígenos se pueden cargar con péptidos presentados por el MHC de clase I transduciendo las células con ácido nucleico, preferiblemente ARN, que codifica un péptido o polipéptido que comprende el péptido que se va a presentar, por ejemplo, un ácido nucleico que codifica un antígeno o polipéptido utilizado para la vacunación.

En algunas realizaciones, se puede administrar a un paciente una composición farmacéutica o vacuna que comprende un vehículo de suministro de ácido nucleico que se dirige a una célula dendrítica u otra célula presentadora de antígeno, lo que da como resultado una transfección que se produce *in vivo*. La transfección *in vivo* de células dendríticas, por ejemplo, se puede realizar generalmente usando cualquier método conocido en la técnica, tal como los descritos en el documento WO 97/24447, o el enfoque de pistola de genes descrito por Mahvi et al., *Immunology and Cell Biology* 75: 456-460, 1997.

De acuerdo con la enseñanza, el término "célula presentadora de antígeno" también incluye células diana.

"Célula diana" significará una célula que es diana para una respuesta inmune tal como una respuesta inmune celular. Las células diana incluyen células que presentan un antígeno, es decir, un fragmento de péptido derivado de un antígeno, e incluyen cualquier célula indeseable, tal como una célula cancerosa. En realizaciones preferidas, la célula diana es una célula que expresa un antígeno como se describe en el presente documento y que presenta preferiblemente dicho antígeno con el MHC de clase I.

El término "porción" se refiere a una fracción. Con respecto a una estructura particular tal como una secuencia de aminoácidos o proteína, el término "porción" de la misma puede designar una fracción continua o discontinua de dicha estructura. Preferiblemente, una porción de una secuencia de aminoácidos comprende al menos 1%, al menos 5%, al menos 10%, al menos 20%, al menos 30%, preferiblemente al menos 40%, preferiblemente al menos 50%, más preferiblemente al menos al menos el 60%, más preferiblemente al menos el 70%, incluso más preferiblemente al menos el 80% y lo más preferiblemente al menos el 90% de los aminoácidos de dicha secuencia de aminoácidos. Preferiblemente, si la porción es una fracción discontinua, dicha fracción discontinua está compuesta por 2, 3, 4, 5, 6, 7, 8 o más partes de una estructura, siendo cada parte un elemento continuo de la estructura. Por ejemplo, una fracción discontinua de una secuencia de aminoácidos puede estar compuesta por 2, 3, 4, 5, 6, 7, 8 o más, preferiblemente no más de 4 partes de dicha secuencia de aminoácidos, en la que cada parte preferiblemente comprende en al menos 5 aminoácidos continuos, al menos 10 aminoácidos continuos, preferiblemente al menos 20 aminoácidos continuos, preferiblemente al menos 30 aminoácidos continuos de la secuencia de aminoácidos.

Los términos "parte" y "fragmento" se usan indistintamente en este documento y se refieren a un elemento continuo. Por ejemplo, una parte de una estructura tal como una secuencia de aminoácidos o una proteína se refiere a un elemento continuo de dicha estructura. Una porción, una parte o un fragmento de una estructura preferiblemente comprende una o más propiedades funcionales de dicha estructura. Por ejemplo, una porción, una parte o un fragmento de un epítipo, péptido o proteína es preferiblemente inmunológicamente equivalente al epítipo, péptido o proteína del que deriva. En el contexto de la presente enseñanza, una "parte" de una estructura tal como una secuencia de aminoácidos preferiblemente comprende, preferiblemente consta de al menos 10%, al menos 20%, al menos 30%, al menos 40%, al menos 50 %, al menos 60%, al menos 70%, al menos 80%, al menos 85%, al menos 90%, al menos 92%, al menos 94%, al menos 96%, al menos 98%, al menos 99% de la estructura completa o secuencia de aminoácidos.

El término "célula inmunorreactiva" en el contexto de la presente enseñanza se refiere a una célula que ejerce funciones efectoras durante una reacción inmunitaria. Una "célula inmunorreactiva" preferiblemente es capaz de unirse a un antígeno o una célula caracterizada por la presentación de un antígeno o un fragmento de péptido del mismo (por ejemplo, un epítipo de células T) y mediar una respuesta inmune. Por ejemplo, tales células secretan citocinas y/o quimiocinas, secretan anticuerpos, reconocen células cancerosas y, opcionalmente, eliminan tales células. Por ejemplo, las células inmunorreactivas comprenden células T (células T citotóxicas, células T colaboradoras, células T que infiltran tumores), células B, células asesinas naturales, neutrófilos, macrófagos y células dendríticas. Preferiblemente, en el contexto de la presente enseñanza, "células inmunorreactivas" son células T, preferiblemente células T CD4⁺ y/o CD8⁺.

Preferiblemente, una "célula inmunorreactiva" reconoce un antígeno o un fragmento de péptido del mismo con cierto grado de especificidad, en particular si se presenta en el contexto de moléculas del MHC tales como en la superficie

de células presentadoras de antígenos o células enfermas tales como células cancerosas. Preferiblemente, dicho reconocimiento permite que la célula que reconoce un antígeno o un fragmento de péptido del mismo sea sensible o reactiva. Si la célula es una célula T colaboradora (célula T CD4⁺) que porta receptores que reconocen un antígeno o un fragmento peptídico del mismo en el contexto de moléculas del MHC de clase II, dicha capacidad de respuesta o reactividad puede implicar la liberación de citocinas y/o la activación de linfocitos CD8⁺ (CTL) y/o células B. Si la célula es un CTL, dicha capacidad de respuesta o reactividad puede implicar la eliminación de células presentadas en el contexto de moléculas del MHC de clase I, es decir, células caracterizadas por la presentación de un antígeno con MHC de clase I, por ejemplo, mediante apoptosis o lisis de células mediadas por perforina. De acuerdo con la enseñanza, la capacidad de respuesta de los CTL puede incluir flujo de calcio sostenido, división celular, producción de citocinas tales como IFN- γ y TNF- α , sobreexpresión de marcadores de activación tales como CD44 y CD69, y destrucción citolítica específica de células diana que expresan antígenos. La capacidad de respuesta de CTL también se puede determinar usando un informador artificial que indique con precisión la capacidad de respuesta de CTL. Dichos CTL que reconocen un antígeno o un fragmento de antígeno y son sensibles o reactivos también se denominan en este documento "CTL sensibles a antígenos". Si la célula es una célula B, dicha capacidad de respuesta puede implicar la liberación de inmunoglobulinas.

Los términos "célula T" y "linfocito T" se utilizan indistintamente en este documento e incluyen células T colaboradoras (células T CD4⁺) y células T citotóxicas (CTL, células T CD8⁺) que comprenden células T citolíticas.

Las células T pertenecen a un grupo de glóbulos blancos conocidos como linfocitos y desempeñan un papel central en la inmunidad mediada por células. Se pueden distinguir de otros tipos de linfocitos, tales como las células B y las células asesinas naturales por la presencia de un receptor especial en su superficie celular llamado receptor de células T (TCR). El timo es el principal órgano responsable de la maduración de las células T. Se han descubierto varios subconjuntos diferentes de células T, cada uno con una función distinta.

Las células T colaboradoras ayudan a otros glóbulos blancos en los procesos inmunológicos, incluida la maduración de las células B en células plasmáticas y la activación de células T citotóxicas y macrófagos, entre otras funciones. Estas células también se conocen como células T CD4⁺ porque expresan la proteína CD4 en su superficie. Las células T colaboradoras se activan cuando se les presentan antígenos peptídicos mediante moléculas del MHC de clase II que se expresan en la superficie de las células presentadoras de antígenos (APC). Una vez activadas, se dividen rápidamente y secretan pequeñas proteínas llamadas citocinas que regulan o ayudan en la respuesta inmune activa.

Las células T citotóxicas destruyen las células infectadas por virus y las células tumorales, y también están implicadas en el rechazo de trasplantes. Estas células también se conocen como células T CD8⁺, ya que expresan la glicoproteína CD8 en su superficie. Estas células reconocen sus dianas al unirse al antígeno asociado con el MHC de clase I, que está presente en la superficie de casi todas las células del cuerpo.

La mayoría de las células T tienen un receptor de células T (TCR) que existe como un complejo de varias proteínas. El receptor de células T real está compuesto por dos cadenas de péptidos separadas, que se producen a partir de los genes alfa y beta del receptor de células T independientes (TCR α y TCR β) y se denominan cadenas α y β del TCR. Las células T $\gamma\delta$ (células T gamma delta) representan un pequeño subconjunto de células T que poseen un receptor de células T (TCR) distinto en su superficie. Sin embargo, en las células T $\gamma\delta$, el TCR está formado por una cadena γ y una cadena δ . Este grupo de células T es mucho menos común (2% del total de células T) que las células T $\alpha\beta$.

La primera señal en la activación de las células T se proporciona mediante la unión del receptor de las células T a un péptido corto presentado por el MHC en otra célula. Esto asegura que solo se active una célula T con un TCR específico para ese péptido. La célula asociada suele ser una célula presentadora de antígeno, tal como una célula presentadora de antígeno profesional, generalmente una célula dendrítica en el caso de respuestas no alteradas, aunque las células B y los macrófagos pueden ser APC importantes.

De acuerdo con la presente enseñanza, una molécula es capaz de unirse a una diana si tiene una afinidad significativa por dicha diana predeterminada y se une a dicha diana predeterminada en ensayos estándar. La "afinidad" o "afinidad de unión" se mide a menudo mediante la constante de disociación de equilibrio (K_D). Una molécula no es (sustancialmente) capaz de unirse a una diana si no tiene una afinidad significativa por dicha diana y no se une significativamente a dicha diana en ensayos estándar.

Los linfocitos T citotóxicos se pueden generar *in vivo* mediante la incorporación de un antígeno o un fragmento de péptido del mismo en células presentadoras de antígeno *in vivo*. El antígeno o un fragmento de péptido del mismo puede representarse como proteína, como ADN (por ejemplo, dentro de un vector) o como ARN. El antígeno puede procesarse para producir un péptido asociado para la molécula del MHC, mientras que puede presentarse un fragmento del mismo sin necesidad de procesamiento adicional. Este último es el caso en particular, si estos pueden unirse a moléculas del MHC. En general, es posible la administración a un paciente mediante inyección intradérmica. Sin embargo, la inyección también puede realizarse intranodalmente en un ganglio linfático (Maloy et al. (2001), Proc Natl Acad Sci USA 98: 3299-303). Las células resultantes presentan el complejo de interés y son reconocidas por linfocitos T citotóxicos autólogos que luego se propagan.

La activación específica de células T CD4+ o CD8+ puede detectarse de diversas formas. Los métodos para detectar la activación de células T específicas incluyen detectar la proliferación de células T, la producción de citocinas (por ejemplo, linfocinas) o la generación de actividad citolítica. Para las células T CD4+, un método preferido para detectar la activación de células T específicas es la detección de la proliferación de células T. Para las células T CD8+, un método preferido para detectar la activación específica de las células T es la detección de la generación de actividad citolítica.

Por "célula caracterizada por la presentación de un antígeno" o "célula que presenta un antígeno" o expresiones similares se entiende una célula tal como una célula enferma, por ejemplo, una célula cancerosa, o una célula presentadora de antígeno que presenta el antígeno que expresa o un fragmento derivado de dicho antígeno, por ejemplo, por procesamiento del antígeno, en el contexto de moléculas del MHC, en particular moléculas del MHC clase I. De manera similar, los términos "enfermedad caracterizada por la presentación de un antígeno" denotan una enfermedad que involucra células caracterizadas por la presentación de un antígeno, en particular con MHC de clase I. La presentación de un antígeno por una célula puede efectuarse transfectando la célula con un ácido nucleico como el ARN que codifica el antígeno.

Por "fragmento de un antígeno que se presenta" o expresiones similares se entiende que el fragmento puede ser presentado por el MHC de clase I o clase II, preferiblemente MHC de clase I, por ejemplo, cuando se agrega directamente a las células presentadoras de antígeno. En una realización, el fragmento es un fragmento que es presentado naturalmente por células que expresan un antígeno.

El término "inmunológicamente equivalente" significa que la molécula inmunológicamente equivalente tal como la secuencia de aminoácidos inmunológicamente equivalente exhibe las mismas o esencialmente las mismas propiedades inmunológicas y/o ejerce los mismos o esencialmente los mismos efectos inmunológicos, por ejemplo, con respecto a el tipo de efecto inmunológico, tal como la inducción de una respuesta inmune humoral y/o celular, la fuerza y/o duración de la reacción inmune inducida, o la especificidad de la reacción inmune inducida. En el contexto de la presente enseñanza, el término "inmunológicamente equivalente" se usa preferiblemente con respecto a los efectos o propiedades inmunológicos de un péptido usado para inmunización. Por ejemplo, una secuencia de aminoácidos es inmunológicamente equivalente a una secuencia de aminoácidos de referencia si dicha secuencia de aminoácidos cuando se expone al sistema inmunológico de un sujeto induce una reacción inmunitaria que tiene la especificidad de reaccionar con la secuencia de aminoácidos de referencia.

El término "funciones efectoras inmunes" en el contexto de la presente enseñanza incluye cualquier función mediada por componentes del sistema inmunológico que dan como resultado, por ejemplo, la destrucción de células tumorales o la inhibición del crecimiento tumoral y/o inhibición del desarrollo de tumores, incluida la inhibición de la diseminación y metástasis tumorales. Preferiblemente, las funciones efectoras inmunes en el contexto de la presente enseñanza son funciones efectoras mediadas por células T. Tales funciones comprenden, en el caso de una célula T colaboradora (célula T CD4+), el reconocimiento de un antígeno o un fragmento de antígeno en el contexto de moléculas del MHC de clase II por receptores de células T, la liberación de citocinas y/o la activación de linfocitos CD8+ (CTL) y/o células B, y en el caso de CTL, el reconocimiento de un antígeno o un fragmento de antígeno en el contexto de moléculas del MHC de clase I por receptores de células T, la eliminación de células presentadas en el contexto de moléculas del MHC de clase I, es decir, células caracterizadas por la presentación de un antígeno con MHC de clase I, por ejemplo, mediante apoptosis o lisis celular mediada por perforina, producción de citocinas yal como IFN- γ y TNF- α , y citolíticos específicos que destruyen las células diana que expresan antígenos.

De acuerdo con la enseñanza, el término "puntuación" se refiere a un resultado, normalmente expresado numéricamente, de una prueba o examen. Términos como "mejor puntuación" o "la mejor puntuación" se relacionan con un mejor resultado o el mejor resultado de una prueba o examen.

Términos tales como "predecir", "que predice" o "predicción" se refieren a la determinación de una probabilidad.

De acuerdo con la enseñanza, determinar una puntuación para la unión de un péptido a una o más moléculas del MHC incluye determinar la probabilidad de unión de un péptido a una o más moléculas del MHC.

Puede determinarse una puntuación para la unión de un péptido a una o más moléculas del MHC usando cualquier péptido: herramientas de predicción de la unión del MHC. Por ejemplo, se puede utilizar el recurso de análisis de la base de datos de epítotos inmunes (IEDB-AR: <http://tools.iedb.org>).

Las predicciones se hacen generalmente contra un conjunto de moléculas del MHC tal como un conjunto de diferentes alelos del MHC tal como todos los alelos posibles del MHC o un conjunto o subconjunto de alelos del MHC encontrados en un paciente que preferiblemente tiene la modificación o modificaciones) cuya inmunogenicidad se determinará de acuerdo con la enseñanza.

De acuerdo con la enseñanza, determinar una puntuación para la unión de un péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T incluye determinar la probabilidad de unión de un péptido cuando está presente en un complejo con una molécula del MHC a los receptores de células T.

Se pueden hacer predicciones contra un receptor de células T tal como un receptor de células T que se encuentra en un paciente o preferiblemente contra un conjunto de receptores de células T tal como un conjunto desconocido de diferentes receptores de células T o un conjunto o subconjunto de receptores de células T encontrado en un paciente que tiene preferiblemente la modificación o modificaciones cuya inmunogenicidad se determinará de acuerdo con la enseñanza.

Además, las predicciones se hacen generalmente contra un conjunto de moléculas del MHC, tal como un conjunto de diferentes alelos del MHC, como todos los alelos del MHC posibles o un conjunto o subconjunto de alelos del MHC que se encuentran en un paciente que preferiblemente tiene la modificación o modificaciones cuya inmunogenicidad se determinará de acuerdo con la enseñanza.

Se puede determinar una puntuación para la unión de un péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T estimando el efecto de la modificación sobre la unión de un complejo péptido-MHC del receptor de células T, dado un repertorio de receptores de células T (desconocido). La puntuación para la unión de un péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T puede definirse generalmente como una especie de proxy para el reconocimiento de una molécula de péptido-MHC dada a un receptor de células T coincidente.

La puntuación para la unión de un péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T puede determinarse determinando las diferencias fisicoquímicas entre el aminoácido modificado y el no modificado. Por ejemplo, pueden usarse matrices de sustitución. Estas matrices describen la velocidad a la que un aminoácido de una secuencia cambia a otros estados de aminoácidos a lo largo del tiempo.

Por ejemplo, se pueden usar matrices logarítmicas de probabilidades tales como matrices logarítmicas de probabilidades basadas en la evolución: una sustitución con una puntuación baja del logaritmo de probabilidades tiene una mayor probabilidad de encontrar un receptor de células T coincidente del grupo de receptores de células T (desconocidas) que una sustitución con una alta puntuación logarítmica de probabilidad (debido a la selección negativa de péptidos no modificados que coinciden con el receptor de células T). Sin embargo, existen otras formas de determinar esta puntuación. Por ejemplo, considerando la posición de la mutación en el péptido (algunas posiciones pueden tener un menor impacto en la unión que otras), teniendo en cuenta los vecinos más cercanos del aminoácido sustituido (lo que podría afectar la estructura secundaria del aminoácido sustituido), teniendo en cuenta la secuencia completa del péptido, teniendo en cuenta la información estructural completa del péptido en la molécula del MHC, etc. La determinación de la puntuación también podría implicar la determinación de un repertorio de receptores de células T (tal como el repertorio de receptores de células T de un paciente o un subconjunto del mismo), por ejemplo, mediante NGS y realizando simulaciones de acoplamiento de complejos de péptido-MHC del receptor de células T.

La presente enseñanza también puede comprender realizar el método de la enseñanza en diferentes péptidos que comprenden la misma modificación o modificaciones y/o diferentes modificaciones.

El término "péptidos diferentes que comprenden la misma modificación o modificaciones" en una realización se refiere a péptidos que comprenden o consisten en diferentes fragmentos de una proteína modificada, comprendiendo dichos diferentes fragmentos las mismas modificaciones presentes en la proteína pero que difieren en longitud y/o posición de la modificación o modificaciones. Si una proteína tiene una modificación en la posición x, dos o más fragmentos de dicha proteína que comprenden cada uno una ventana de secuencia diferente de dicha proteína que cubre dicha posición x se consideran péptidos diferentes que comprenden las mismas modificaciones.

El término "péptidos diferentes que comprenden modificaciones diferentes" en una realización se refiere a péptidos de la misma y/o diferentes longitudes que comprenden diferentes modificaciones de la misma y/o proteínas diferentes. Si una proteína tiene modificaciones en las posiciones x e y, dos fragmentos de dicha proteína que comprenden cada uno una ventana de secuencia de dicha proteína que cubre la posición x o la posición y se consideran péptidos diferentes que comprenden modificaciones diferentes.

La presente enseñanza también puede comprender la ruptura de secuencias de proteínas que tienen modificaciones cuya inmunogenicidad se determinará de acuerdo con la enseñanza en longitudes de péptido apropiadas para la unión del MHC y determinar las puntuaciones para la unión a una o más moléculas del MHC de diferentes péptidos modificados que comprenden las mismas y/o diferentes modificaciones de las mismas y/o diferentes proteínas. Las salidas pueden clasificarse y pueden consistir en una lista de péptidos y sus puntuaciones previstas, lo que indica su probabilidad de unión.

La etapa de determinar una puntuación para la unión del péptido no modificado a una o más moléculas del MHC y/o la etapa de determinar una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T pueden realizarse posteriormente con todos los péptidos modificados diferentes que comprenden las mismas y/o diferentes modificaciones, un subconjunto de las mismas, por ejemplo, aquellos péptidos modificados que comprenden las mismas y/o diferentes modificaciones que puntúan mejor para unirse a una o más moléculas del MHC, o solo con el péptido modificado que puntúa mejor para unirse a una o más moléculas del MHC.

Después de dichas etapas adicionales, los resultados pueden clasificarse y pueden consistir en una lista de péptidos y sus puntuaciones previstas, lo que indica su probabilidad de ser inmunogénicos.

5 Preferiblemente, en tal clasificación, una puntuación para la unión del péptido modificado a una o más moléculas del MHC se pondera más alta que una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T, preferiblemente una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados y una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido a uno o más receptores de células T, preferiblemente una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados se ponderan más que una puntuación para la unión del péptido no modificado a una o más moléculas del MHC.

15 Las modificaciones de aminoácidos cuya inmunogenicidad se va a determinar de acuerdo con la presente enseñanza pueden resultar de mutaciones en el ácido nucleico de una célula. Tales mutaciones pueden identificarse mediante técnicas de secuenciación conocidas.

En una realización, las mutaciones son mutaciones somáticas específicas de cáncer en una muestra de tumor de un paciente con cáncer que pueden determinarse identificando diferencias de secuencia entre el genoma, exoma y/o transcriptoma de una muestra de tumor y el genoma, exoma y/o transcriptoma de una muestra no tumorigena.

20 De acuerdo con la enseñanza, una muestra de tumor se refiere a cualquier muestra tal como una muestra corporal derivada de un paciente que contiene o se espera que contenga células tumorales o cancerosas. La muestra corporal puede ser cualquier muestra de tejido tal como sangre, una muestra de tejido obtenida del tumor primario o de metástasis tumorales o cualquier otra muestra que contenga células tumorales o cancerosas. Preferiblemente, una muestra corporal es sangre y se determinan mutaciones somáticas o diferencias de secuencia específicas del cáncer en una o más células tumorales circulantes (CTC) contenidas en la sangre. En otra realización, una muestra de tumor se refiere a una o más células tumorales o cancerosas aisladas tales como células tumorales circulantes (CTC) o una muestra que contiene una o más células tumorales o cancerosas aisladas tales como células tumorales circulantes (CTC).

30 Una muestra no tumoral se refiere a cualquier muestra tal como una muestra corporal derivada de un paciente u otro individuo que preferiblemente es de la misma especie que el paciente, preferiblemente un individuo sano que no contiene o no se espera que contenga células tumorales o cancerosas. La muestra corporal puede ser cualquier muestra de tejido tal como sangre o una muestra de un tejido no tumorigena.

35 La enseñanza puede implicar la determinación de la firma de mutación del cáncer de un paciente. El término "firma de mutación cancerosa" puede referirse a todas las mutaciones cancerosas presentes en una o más células cancerosas de un paciente o puede referirse solo a una parte de las mutaciones cancerosas presentes en una o más células cancerosas de un paciente. Por consiguiente, la presente enseñanza puede implicar la identificación de todas las mutaciones específicas de cáncer presentes en una o más células cancerosas de un paciente o puede implicar la identificación de solo una parte de las mutaciones específicas de cáncer presentes en una o más células cancerosas de un paciente. Generalmente, los métodos de la enseñanza proporcionan la identificación de una serie de mutaciones que proporcionan un número suficiente de modificaciones o péptidos modificados para ser incluidos en los métodos de la enseñanza.

45 Preferiblemente, las mutaciones identificadas de acuerdo con la presente enseñanza son mutaciones no sinónimas, preferiblemente mutaciones no sinónimas de proteínas expresadas en una célula tumoral o cancerosa.

50 En una realización, las mutaciones somáticas específicas de cáncer o las diferencias de secuencia se determinan en el genoma, preferiblemente el genoma completo, de una muestra de tumor. Por lo tanto, la enseñanza puede comprender identificar la firma de la mutación del cáncer del genoma, preferiblemente el genoma completo de una o más células cancerosas. En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer en una muestra de tumor de un paciente con cáncer comprende identificar el perfil de mutación de cáncer de todo el genoma.

55 En una realización, las mutaciones somáticas específicas de cáncer o las diferencias de secuencia se determinan en el exoma, preferiblemente en todo el exoma, de una muestra de tumor. Por lo tanto, la enseñanza puede comprender identificar la firma de mutación del cáncer del exoma, preferiblemente el exoma completo de una o más células cancerosas. En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer en una muestra de tumor de un paciente con cáncer comprende identificar el perfil de mutación de cáncer en todo el exoma.

60 En una realización, las mutaciones somáticas específicas de cáncer o las diferencias de secuencia se determinan en el transcriptoma, preferiblemente el transcriptoma completo, de una muestra de tumor. Por lo tanto, la enseñanza puede comprender identificar la firma de mutación del cáncer del transcriptoma, preferiblemente el transcriptoma completo de una o más células cancerosas. En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer en una muestra de tumor de un paciente con cáncer comprende identificar el perfil de mutación de cáncer en todo el transcriptoma.

65

- 5 En una realización, la etapa de identificar mutaciones somáticas específicas del cáncer o identificar diferencias de secuencia comprende la secuenciación de una sola célula de una o más, preferiblemente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 o incluso más células cancerosas. Por lo tanto, la enseñanza puede comprender identificar una firma de mutación cancerosa de dichas una o más células cancerosas. En una realización, las células cancerosas son células tumorales circulantes. Las células cancerosas, tal como las células tumorales circulantes, pueden aislarse antes de la secuenciación de células individuales.
- 10 En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer o identificar diferencias de secuencia implica el uso de secuenciación de próxima generación (NGS).
- 15 En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer o identificar diferencias de secuencia comprende secuenciar el ADN y/o el ARN genómico de la muestra de tumor.
- 20 Para revelar mutaciones somáticas específicas de cáncer o diferencias de secuencia, la información de secuencia obtenida de la muestra de tumor se compara preferiblemente con una referencia, tal como la información de secuencia obtenida de la secuenciación de ácidos nucleicos, tal como ADN o ARN de células normales no cancerosas, tales como células de línea germinal que pueden obtenerse del paciente o de un individuo diferente. En una realización, el ADN de la línea germinal genómica normal se obtiene a partir de células mononucleares de sangre periférica (PBMC)
- 25 El término "genoma" se refiere a la cantidad total de información genética en los cromosomas de un organismo o una célula.
- El término "exoma" se refiere a parte del genoma de un organismo formado por exones, que codifican porciones de genes expresados. El exoma proporciona el modelo genético utilizado en la síntesis de proteínas y otros productos génicos funcionales. Es la parte más funcionalmente relevante del genoma y, por lo tanto, es más probable que contribuya al fenotipo de un organismo. Se estima que el exoma del genoma humano comprende el 1,5% del genoma total (Ng, PC et al., PLoS Gen., 4 (8): 1-15, 2008).
- 30 El término "transcriptoma" se refiere al conjunto de todas las moléculas de ARN, incluyendo ARNm, ARNr, ARNt y otro ARN no codificante producido en una célula o una población de células. En el contexto de la presente enseñanza, el transcriptoma significa el conjunto de todas las moléculas de ARN producidas en una célula, una población de células, preferiblemente una población de células cancerosas, o todas las células de un individuo dado en un momento determinado.
- 35 Un "ácido nucleico" es según la enseñanza preferiblemente ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN), más preferiblemente ARN, lo más preferiblemente ARN transcrito *in vitro* (ARN IVT) o ARN sintético. Los ácidos nucleicos incluyen, según la enseñanza, ADN genómico, ADNc, ARNm, moléculas producidas de forma recombinante y sintetizadas químicamente. De acuerdo con la enseñanza, un ácido nucleico puede estar presente como una molécula monocatenaria o bicatenaria y lineal o covalentemente cerrada circularmente. Un ácido nucleico puede, según la enseñanza, aislarse. El término "ácido nucleico aislado" significa, de acuerdo con la enseñanza, que el ácido nucleico (i) se amplificó *in vitro*, por ejemplo mediante la reacción en cadena de la polimerasa (PCR), (ii) se produjo de forma recombinante mediante clonación, (iii) se purificó, por ejemplo, mediante escisión y separación mediante electroforesis en gel, o (iv) se sintetizó, por ejemplo, mediante síntesis química. Puede emplearse un ácido nucleico para la introducción, es decir, transfección de células, en particular, en forma de ARN que puede prepararse mediante transcripción *in vitro* a partir de una plantilla de ADN. Además, el ARN se puede modificar antes de la aplicación mediante la estabilización de las secuencias, la protección y la poliadenilación.
- 40 El término "material genético" se refiere a un ácido nucleico aislado, ya sea ADN o ARN, una sección de una doble hélice, una sección de un cromosoma o el genoma completo de un organismo o célula, en particular su exoma o transcriptoma.
- 45 El término "mutación" se refiere a un cambio o diferencia en la secuencia de ácido nucleico (sustitución, adición o eliminación de nucleótidos) en comparación con una referencia. Una "mutación somática" puede ocurrir en cualquiera de las células del cuerpo, excepto en las células germinales (espermatozoides y óvulos) y, por lo tanto, no se transmiten a los niños. Estas alteraciones pueden (pero no siempre) causar cáncer u otras enfermedades. Preferiblemente, una mutación es una mutación no sinónima. El término "mutación no sinónima" se refiere a una mutación, preferiblemente una sustitución de nucleótidos, que da como resultado un cambio de aminoácido tal como una sustitución de aminoácido en el producto de traducción.
- 50 De acuerdo con la enseñanza, el término "mutación" incluye mutaciones puntuales, Indels, fusiones, cromotripsis y ediciones de ARN.
- 55 De acuerdo con la enseñanza, el término "Indel" describe una clase de mutación especial, definida como una mutación que da como resultado una inserción y eliminación colocalizada y una ganancia o pérdida neta de nucleótidos. En las regiones codificantes del genoma, a menos que la longitud de un indel sea un múltiplo de 3, producen una mutación por desplazamiento del marco de lectura. Los indeles se pueden contrastar con una mutación puntual; en la que un
- 60
- 65

Indel inserta y elimina nucleótidos de una secuencia, una mutación puntual es una forma de sustitución que reemplaza uno de los nucleótidos.

5 Las fusiones pueden generar genes híbridos formados a partir de dos genes previamente separados. Puede ocurrir como resultado de una translocación, eliminación intersticial o inversión cromosómica. A menudo, los genes de fusión son oncogenes. Los genes de fusión oncogénicos pueden conducir a un producto génico con una función nueva o diferente a la de los dos compañeros de fusión. Alternativamente, un protooncogén se fusiona con un promotor fuerte y, por lo tanto, la función oncogénica se establece para que funcione mediante una sobreexpresión causada por el promotor fuerte del compañero de fusión secuencia arriba. Los transcritos de fusión oncogénica también pueden ser
10 causadas por eventos de empalme trans o de lectura.

De acuerdo con la enseñanza, el término "cromotripsis" se refiere a un fenómeno genético por el cual regiones específicas del genoma se rompen y luego se unen mediante un solo evento devastador.

15 De acuerdo con la enseñanza, el término "edición de ARN" o "editar ARN" se refiere a procesos moleculares en los que el contenido de información en una molécula de ARN se altera mediante un cambio químico en la composición de la base. La edición de ARN incluye modificaciones de nucleósidos tales como desaminaciones de citidina (C) a uridina (U) y adenosina (A) a inosina (I), así como adiciones e inserciones de nucleótidos sin plantilla. La edición de ARN en ARNm altera eficazmente la secuencia de aminoácidos de la proteína codificada para que difiera de la predicha por la
20 secuencia de ADN genómico.

El término "firma de mutación del cáncer" se refiere a un conjunto de mutaciones que están presentes en las células cancerosas en comparación con las células de referencia no cancerosas.

25 De acuerdo con la enseñanza, se puede usar una "referencia" para correlacionar y comparar los resultados obtenidos en los métodos de enseñanza a partir de una muestra de tumor. Normalmente, la "referencia" puede obtenerse sobre la base de una o más muestras normales, en particular muestras que no se ven afectadas por una enfermedad cancerosa, obtenidas de un paciente o de uno o más individuos diferentes, preferiblemente individuos sanos, en particular individuos de la misma especie. Una "referencia" puede determinarse empíricamente analizando un número
30 suficientemente grande de muestras normales.

Se puede usar cualquier método de secuenciación adecuado de acuerdo con la enseñanza para determinar mutaciones, prefiriéndose las tecnologías de secuenciación de próxima generación (NGS). Los métodos de
35 secuenciación de tercera generación podrían sustituir a la tecnología NGS en el futuro para acelerar la etapa de secuenciación del método. Para fines de aclaración: los términos "Secuenciación de próxima generación" o "NGS" en el contexto de la presente enseñanza se refieren a todas las nuevas tecnologías de secuenciación de alto rendimiento que, en contraste con la metodología de secuenciación "convencional" conocida como química de Sanger, leen plantillas de ácido nucleico al azar en paralelo a lo largo de todo el genoma rompiendo todo el genoma en pequeños trozos. Dichas tecnologías NGS (también conocidas como tecnologías de secuenciación masivamente paralela) son
40 capaces de entregar información de secuencia de ácido nucleico de un genoma completo, exoma, transcriptoma (todas las secuencias transcritas de un genoma) o metiloma (todas las secuencias metiladas de un genoma) en períodos de muy poco tiempo, por ejemplo dentro de 1-2 semanas, preferiblemente dentro de 1-7 días o más preferiblemente dentro de menos de 24 horas y permitir, en principio, enfoques de secuenciación de células individuales. Se pueden usar múltiples plataformas NGS que están disponibles comercialmente o que se mencionan en la literatura en el
45 contexto de la presente enseñanza, por ejemplo, las descritas en detalle en Zhang et al., 2011: The impact of next-generation sequencing on genomics. J. Genet Genomics 38 (3), 95-109; o en Voelkerding et al., 2009: Next generation sequencing: From basic research to diagnostics. Clinical chemistry 55, 641-658. Ejemplos no limitantes de tales tecnologías/plataformas NGS son

50 1) La tecnología de secuenciación por síntesis conocida como pirosecuenciación implementada, por ejemplo en el GS-FLX 454 Genome Sequencer^{MR} de la empresa asociada a Roche 454 Life Sciences (Branford, Connecticut), descrito por primera vez en Ronaghi et al., 1998: A sequencing method based on real-time pyrophosphate". Science 281. (5375), 363-365. Esta tecnología utiliza una PCR en emulsión en la que las perlas de unión a ADN monocatenarias se encapsulan mediante agitación vigorosa en micelas acuosas que contienen reactivos de PCR rodeados de aceite para amplificación por PCR en emulsión. Durante el proceso de pirosecuenciación, la luz emitida por las moléculas de fosfato durante la incorporación de nucleótidos se registra a medida que la polimerasa sintetiza la cadena de ADN.

2) Los enfoques de secuenciación por síntesis desarrollados por Solexa (ahora parte de Illumina Inc., San Diego, California) que se basan en terminadores de colorantes reversibles e implementados por ejemplo en Illumina/Solexa Genome Analyzer^{MR} y en Illumina HiSeq 2000 Genome Analyzer^{MR}. En esta tecnología, los cuatro nucleótidos se agregan simultáneamente en fragmentos de racimo cebados con oligo en canales de celdas de flujo junto con la ADN polimerasa. La amplificación de puente extiende las cadenas de grupos con los cuatro nucleótidos marcados con fluorescencia para la secuenciación.

3) Enfoques de secuenciación por ligadura, por ejemplo implementados en la plataforma SOLid^{MR} de Applied Biosystems (ahora Life Technologies Corporation, Carlsbad, California). En esta tecnología, se marca un conjunto de todos los posibles oligonucleótidos de una longitud fija de acuerdo con la posición secuenciada. Los oligonucleótidos
65

se hibridan y se ligan; la ligadura preferencial por ADN ligasa para emparejar secuencias da como resultado una señal informativa del nucleótido en esa posición. Antes de la secuenciación, el ADN se amplifica mediante PCR en emulsión. La perla resultante, cada una de las cuales contiene solo copias de la misma molécula de ADN, se deposita en un portaobjetos de vidrio. Como segundo ejemplo, la plataforma Polonator^{MR} G.007 de Dover Systems (Salem, New Hampshire) también emplea un enfoque de secuenciación por ligación mediante el uso de una PCR de emulsión basada en perlas y ordenada al azar para amplificar fragmentos de ADN para la secuenciación paralela.

4) Tecnologías de secuenciación de una sola molécula, tales como por ejemplo las implementadas en el sistema PacBio RS de Pacific Biosciences (Menlo Park, California) o en la plataforma HeliScope^{MR} de Helicos Biosciences (Cambridge, Massachusetts). La característica distintiva de esta tecnología es su capacidad para secuenciar moléculas simples de ADN o ARN sin amplificación, definida como secuenciación de ADN de una sola molécula en tiempo real (SMRT). Por ejemplo, HeliScope utiliza un sistema de detección de fluorescencia de alta sensibilidad para detectar directamente cada nucleótido a medida que se sintetiza. Un enfoque similar basado en la transferencia de energía por resonancia de fluorescencia (FRET) se ha desarrollado en Visigen Biotechnology (Houston, Texas). Otras técnicas de molécula única basadas en fluorescencia son de U.S. Genomics (GeneEngine^{MR}) y Genovox (AnyGene^{MR}).

5) Nanotecnologías para la secuenciación de una sola molécula en las que se utilizan varias nanoestructuras que se disponen, por ejemplo en un chip para monitorear el movimiento de una molécula de polimerasa en una sola cadena durante la replicación. Ejemplos no limitativos de enfoques basados en nanotecnologías son la plataforma GridON^{MR} de Oxford Nanopore Technologies (Oxford, Reino Unido), las plataformas de secuenciación de nanoporos asistida por hibridación (HANS^{MR}) desarrolladas por Nabsys (Providence, Rhode Island), y la plataforma patentada de secuenciación de ADN basada en ligasa con tecnología de nanoesferas de ADN (DNB) denominada ligadura combinatoria de sonda-anclaje (cPAL^{MR}).

6) Tecnologías basadas en microscopía electrónica para secuenciación de una sola molécula, por ejemplo las desarrolladas por LightSpeed Genomics (Sunnyvale, California) y Halcyon Molecular (Redwood City, California).

7) Secuenciación de semiconductores de iones que se basa en la detección de iones de hidrógeno que se liberan durante la polimerización del ADN. Por ejemplo, Ion Torrent Systems (San Francisco, California) utiliza una matriz de pozos micromecanizados de alta densidad para realizar este proceso bioquímico de forma masivamente paralela. Cada pocillo contiene una plantilla de ADN diferente. Debajo de los pozos hay una capa sensible a los iones y debajo de ella un sensor de iones patentado.

Preferiblemente, las preparaciones de ADN y ARN sirven como material de partida para NGS. Dichos ácidos nucleicos se pueden obtener fácilmente a partir de muestras tales como material biológico, por ejemplo, de tejidos tumorales incrustados en parafina (FFPE) recién congelados o fijados en formalina o de células recién aisladas o de CTC que están presentes en la sangre periférica de los pacientes. El ADN o ARN genómico normal no mutado se puede extraer de tejido somático normal, sin embargo, se prefieren las células de la línea germinal en el contexto de la presente enseñanza. El ADN o ARN de la línea germinal se puede extraer de las células mononucleares de sangre periférica (PBMC) en pacientes con neoplasias malignas no hematológicas. Aunque los ácidos nucleicos extraídos de tejidos FFPE o células individuales recién aisladas están muy fragmentados, son adecuados para aplicaciones de NGS.

En la bibliografía se describen varios métodos de NGS dirigidos para la secuenciación del exoma (para una revisión, véase, por ejemplo, Teer y Mullikin 2010: Human Mol Genet 19 (2), R145-51), todos los cuales pueden usarse junto con la presente enseñanza. Muchos de estos métodos (descritos, por ejemplo, como captura del genoma, partición del genoma, enriquecimiento del genoma, etc.) utilizan técnicas de hibridación e incluyen enfoques de hibridación basados en matrices (por ejemplo, Hodges et al. 2007: Nat. Genet. 39, 1522-1527) y basados en líquidos (por ejemplo, Choi et al., 2009: Proc. Natl. Acad. Sci USA 106, 19096-19101). También se encuentran disponibles kits comerciales para la preparación de muestras de ADN y la posterior captura del exoma: por ejemplo, Illumina Inc. (San Diego, California) ofrece el kit de preparación de muestras de ADN TruSeq^{MR} y el kit de enriquecimiento del exoma TruSeq^{MR}.

Con el fin de reducir el número de hallazgos falsos positivos en la detección de mutaciones somáticas específicas de cáncer o diferencias de secuencia cuando se comparan, por ejemplo, la secuencia de una muestra de tumor con la secuencia de una muestra de referencia, tal como la secuencia de una muestra de línea germinal, se prefiere determinar la secuencia en réplicas de uno o ambos tipos de muestras. Por lo tanto, se prefiere que la secuencia de una muestra de referencia, tal como la secuencia de una muestra de la línea germinal, se determine dos veces, tres veces o más. Alternativa o adicionalmente, la secuencia de una muestra de tumor se determina dos veces, tres veces o más. También puede ser posible determinar la secuencia de una muestra de referencia, tal como la secuencia de una muestra de línea germinal y/o la secuencia de una muestra de tumor más de una vez, determinando al menos una vez la secuencia en el ADN genómico y determinando al menos una vez la secuencia en ARN de dicha muestra de referencia y/o de dicha muestra tumoral. Por ejemplo, al determinar las variaciones entre réplicas de una muestra de referencia, tal como una muestra de línea germinal, se puede estimar la tasa esperada de mutaciones somáticas de falso positivo (FDR) como una cantidad estadística. Las repeticiones técnicas de una muestra deben generar resultados idénticos y cualquier mutación detectada en esta "comparación igual frente a la misma" es un falso positivo. En particular, para determinar la tasa de descubrimiento falso para la detección de mutaciones somáticas en una muestra de tumor en relación con una muestra de referencia, se puede usar una repetición técnica de la muestra de referencia como referencia para estimar el número de falsos positivos. Además, varias métricas relacionadas con la calidad (por ejemplo, cobertura o calidad de SNP) se pueden combinar en una sola puntuación de calidad utilizando un enfoque de aprendizaje automático. Para una variación somática dada, se pueden contar todas las demás

variaciones con un puntaje de calidad superior, lo que permite una clasificación de todas las variaciones en un conjunto de datos.

En el contexto de la presente enseñanza, el término "ARN" se refiere a una molécula que comprende al menos un residuo de ribonucleótido y que preferiblemente está compuesta total o sustancialmente por residuos de ribonucleótido. "Ribonucleótido" se refiere a un nucleótido con un grupo hidroxilo en la posición 2' de un grupo β -D-ribofuranosilo. El término "ARN" comprende ARN bicatenario, ARN monocatenario, ARN aislado, tal como ARN parcial o completamente purificado, ARN esencialmente puro, ARN sintético y ARN generado de forma recombinante, tal como ARN modificado que se diferencia del ARN natural por adición, eliminación, sustitución y/o alteración de uno o más nucleótidos. Tales alteraciones pueden incluir la adición de material no nucleotídico, tal como en el extremo o extremos de un ARN o internamente, por ejemplo en uno o más nucleótidos del ARN. Los nucleótidos en las moléculas de ARN también pueden comprender nucleótidos no estándar, tales como nucleótidos de origen no natural o nucleótidos o desoxinucleótidos sintetizados químicamente. Estos ARN alterados pueden denominarse análogos o análogos de ARN de origen natural.

De acuerdo con la presente enseñanza, el término "ARN" incluye y preferiblemente se refiere a "ARNm". El término "ARNm" significa "ARN mensajero" y se refiere a un "transcripto" que se genera usando una plantilla de ADN y codifica un péptido o polipéptido. Normalmente, un ARNm comprende una 5'-UTR, una región codificante de proteína y una 3'-UTR. El ARNm solo posee una vida media limitada en las células e *in vitro*. En el contexto de la presente enseñanza, el ARNm puede generarse mediante transcripción *in vitro* a partir de una plantilla de ADN. El experto en la materia conoce la metodología de transcripción *in vitro*. Por ejemplo, existe una variedad de kits de transcripción *in vitro* disponibles comercialmente.

De acuerdo con la enseñanza, la estabilidad y la eficacia de traducción del ARN pueden modificarse según se requiera. Por ejemplo, el ARN puede estabilizarse y su traducción aumentada mediante una o más modificaciones que tienen efectos estabilizadores y/o un aumento de la eficiencia de traducción del ARN. Tales modificaciones se describen, por ejemplo, en el documento PCT/EP2006/009448. Para aumentar la expresión del ARN utilizado de acuerdo con la presente enseñanza, se puede modificar dentro de la región codificante, es decir, la secuencia que codifica el péptido o proteína expresados, preferiblemente sin alterar la secuencia del péptido o proteína expresados, para aumentar el contenido de GC para aumentar la estabilidad del ARNm y realizar una optimización de codones y, por lo tanto, mejorar la traducción en las células.

El término "modificación" en el contexto del ARN usado en la presente enseñanza incluye cualquier modificación de un ARN que no está presente de forma natural en dicho ARN.

En una realización de la enseñanza, el ARN utilizado de acuerdo con la enseñanza no tiene 5'-trifosfatos desprotegidos. La eliminación de dichos 5'-trifosfatos desprotegidos se puede lograr tratando el ARN con una fosfatasa.

El ARN de acuerdo con la enseñanza puede tener ribonucleótidos modificados para aumentar su estabilidad y/o disminuir la citotoxicidad. Por ejemplo, en una realización, en el ARN usado de acuerdo con la enseñanza, la 5-metilcitosina se sustituye parcial o completamente, preferiblemente completamente, por citidina. Alternativa o adicionalmente, en una realización, en el ARN usado de acuerdo con la enseñanza, la pseudouridina se sustituye parcial o completamente, preferiblemente completamente, por uridina.

En una realización, el término "modificación" se refiere a proporcionar un ARN con una protección de 5' o análogo de protección de 5'. El término "protección de 5'" se refiere a una estructura de protección que se encuentra en el extremo 5' de una molécula de ARNm y generalmente consiste en un nucleótido de guanosina conectado al ARNm a través de un enlace inusual de 5' a 5' trifosfato. En una realización, esta guanosina está metilada en la posición 7. El término "protección convencional de 5'" se refiere a una protección 5' de ARN natural, preferiblemente a la protección de 7-metilguanosina (m⁷G). En el contexto de la presente enseñanza, el término "protección de 5'" incluye un análogo de protección de 5' que se asemeja a la estructura de protección del ARN y se modifica para poseer la capacidad de estabilizar el ARN y/o mejorar la traducción del ARN si está unido al mismo, preferiblemente *in vivo* y/o en una célula.

Proporcionar un ARN con una protección de 5' o un análogo de protección de 5' se puede lograr mediante la transcripción *in vitro* de una plantilla de ADN en presencia de dicha protección de 5' o análogo de protección de 5', en la que dicha protección de 5' se incorpora co-transcripcionalmente en la cadena de ARN generada, o el ARN puede generarse, por ejemplo, mediante transcripción *in vitro*, y la protección de 5' puede unirse al ARN post-transcripcionalmente usando enzimas de protección, por ejemplo, enzimas de protección del virus vaccinia.

El ARN puede comprender modificaciones adicionales. Por ejemplo, una modificación adicional del ARN usado en la presente enseñanza puede ser una extensión o truncamiento de la cola poli(A) de origen natural o una alteración de las regiones 5' o 3' no traducidas (UTR) como la introducción de una UTR que no está relacionada con la región codificante de dicho ARN, por ejemplo, el intercambio de la 3'-UTR existente con o la inserción de una o más, preferiblemente dos copias de una 3'-UTR derivada de un gen de globina, tales como alfa2-globina, alfa1-globina, beta-globina, preferiblemente beta-globina, más preferiblemente beta-globina humana.

El ARN que tiene una secuencia poli-A sin enmascarar se traduce de manera más eficaz que el ARN que tiene una secuencia poli-A enmascarada. El término "cola de poli(A)" o "secuencia poli(A)" se refiere a una secuencia de residuos de adenilo (A) que normalmente se encuentra en el extremo 3' de una molécula de ARN y "secuencia de poli-A desenmascarada" significa que la secuencia de poli-A en el extremo 3' de una molécula de ARN termina con una A de la secuencia poli-A y no es seguida por nucleótidos distintos de A ubicado en el extremo 3', es decir, secuencia abajo, de la secuencia poli-A. Además, una secuencia poli-A larga de aproximadamente 120 pares de bases da como resultado una estabilidad de transcripción y una eficiencia de traducción óptimas del ARN.

Por lo tanto, para aumentar la estabilidad y/o la expresión del ARN usado de acuerdo con la presente enseñanza, se puede modificar para que esté presente junto con una secuencia de poli-A, preferiblemente con una longitud de 10 a 500, más preferiblemente de 30 a 300, incluso más preferiblemente de 65 a 200 y especialmente de 100 a 150 residuos de adenosina. En una realización especialmente preferida, la secuencia de poli-A tiene una longitud de aproximadamente 120 residuos de adenosina. Para aumentar aún más la estabilidad y/o la expresión del ARN usado de acuerdo con la enseñanza, se puede desenmascarar la secuencia de poli-A.

Además, la incorporación de una región no traducida 3' (UTR) en la región no traducida 3' de una molécula de ARN puede dar como resultado una mejora en la eficacia de la traducción. Puede conseguirse un efecto sinérgico incorporando dos o más de dichas regiones 3' no traducidas. Las regiones 3' no traducidas pueden ser autólogas o heterólogas al ARN en el que se introducen. En una realización particular, la región 3' no traducida se deriva del gen de la β -globina humana.

Una combinación de las modificaciones descritas anteriormente, es decir, la incorporación de una secuencia de poli-A, el desenmascaramiento de una secuencia de poli-A y la incorporación de una o más regiones 3' no traducidas, tiene una influencia sinérgica sobre la estabilidad del ARN y aumento de la eficiencia de la traducción.

El término "estabilidad" del ARN se refiere a la "vida media" del ARN. La "vida media" se refiere al período de tiempo que se necesita para eliminar la mitad de la actividad, cantidad o número de moléculas. En el contexto de la presente enseñanza, la vida media de un ARN es indicativa de la estabilidad de dicho ARN. La vida media del ARN puede influir en la "duración de la expresión" del ARN. Se puede esperar que el ARN que tiene una vida media larga se exprese durante un período de tiempo prolongado.

Por supuesto, si de acuerdo con la presente enseñanza se desea disminuir la estabilidad y/o la eficiencia de traducción del ARN, es posible modificar el ARN para interferir con la función de los elementos descritos anteriormente aumentando la estabilidad y/o eficiencia de traducción del ARN.

El término "expresión" se usa de acuerdo con la enseñanza en su significado más general y comprende la producción de ARN y/o péptidos, polipéptidos o proteínas, por ejemplo, por transcripción y/o traducción. Con respecto al ARN, el término "expresión" o "traducción" se refiere en particular a la producción de péptidos, polipéptidos o proteínas. También comprende la expresión parcial de ácidos nucleicos. Además, la expresión puede ser transitoria o estable.

De acuerdo con la enseñanza, el término expresión también incluye una "expresión aberrante" o "expresión anormal". "Expresión aberrante" o "expresión anormal" significa de acuerdo con la enseñanza que la expresión se altera, preferiblemente aumenta, en comparación con una referencia, por ejemplo, un estado en un sujeto que no tiene una enfermedad asociada con la expresión aberrante o anormal de una determinada proteína, por ejemplo, un antígeno tumoral. Un aumento de expresión se refiere a un aumento de al menos un 10%, en particular al menos un 20%, al menos un 50% o al menos un 100% o más. En una realización, la expresión solo se encuentra en un tejido enfermo, mientras que la expresión en un tejido sano está reprimida.

El término "expresado específicamente" significa que una proteína se expresa esencialmente solo en un tejido u órgano específico. Por ejemplo, un antígeno tumoral expresado específicamente en la mucosa gástrica significa que dicha proteína se expresa principalmente en la mucosa gástrica y no se expresa en otros tejidos o no se expresa en un grado significativo en otros tipos de tejidos u órganos. Así, una proteína que se expresa exclusivamente en células de la mucosa gástrica y en un grado significativamente menor en cualquier otro tejido, tal como los testículos, se expresa específicamente en células de la mucosa gástrica. En algunas realizaciones, un antígeno tumoral también puede expresarse específicamente en condiciones normales en más de un tipo de tejido u órgano, tal como en 2 o 3 tipos de tejidos u órganos, pero preferiblemente en no más de 3 tipos de tejidos u órganos diferentes. En este caso, el antígeno tumoral se expresa específicamente en estos órganos. Por ejemplo, si un antígeno tumoral se expresa en condiciones normales preferiblemente en una extensión aproximadamente igual en pulmón y estómago, dicho antígeno tumoral se expresa específicamente en pulmón y estómago.

En el contexto de la presente enseñanza, el término "transcripción" se refiere a un proceso, en el que el código genético en una secuencia de ADN se transcribe en ARN. Posteriormente, el ARN puede traducirse en proteína. De acuerdo con la presente enseñanza, el término "transcripción" comprende "transcripción *in vitro*", en el que el término "transcripción *in vitro*" se refiere a un proceso en el que el ARN, en particular el ARNm, se sintetiza *in vitro* en un sistema libre de células, preferiblemente usando extractos celulares apropiados. Preferiblemente, los vectores de clonación se aplican para la generación de transcripciones. Estos vectores de clonación se designan generalmente

como vectores de transcripción y, de acuerdo con la presente enseñanza, están abarcados por el término "vector". De acuerdo con la presente enseñanza, el ARN usado en la presente enseñanza es preferiblemente ARN transcrito *in vitro* (ARN-IVT) y puede obtenerse mediante transcripción *in vitro* de una plantilla de ADN apropiada. El promotor para controlar la transcripción puede ser cualquier promotor de cualquier ARN polimerasa. Ejemplos particulares de ARN polimerasas son las ARN polimerasas T7, T3 y SP6. Preferiblemente, la transcripción *in vitro* de acuerdo con la enseñanza está controlada por un promotor T7 o SP6. Puede obtenerse una plantilla de ADN para la transcripción *in vitro* clonando un ácido nucleico, en particular ADNc, e introduciéndolo en un vector apropiado para la transcripción *in vitro*. El ADNc puede obtenerse mediante transcripción inversa de ARN.

El término "traducción" de acuerdo con la enseñanza se refiere al proceso en los ribosomas de una célula mediante el cual una cadena de ARN mensajero dirige el ensamblaje de una secuencia de aminoácidos para producir un péptido, polipéptido o proteína.

Las secuencias de control de la expresión o secuencias reguladoras, que de acuerdo con la enseñanza pueden estar unidas funcionalmente con un ácido nucleico, pueden ser homólogas o heterólogas con respecto al ácido nucleico. Una secuencia codificante y una secuencia reguladora están unidas "funcionalmente" si están unidas covalentemente, de modo que la transcripción o traducción de la secuencia codificante está bajo el control o bajo la influencia de la secuencia reguladora. Si la secuencia codificante se va a traducir en una proteína funcional, con un enlace funcional de una secuencia reguladora con la secuencia codificante, la inducción de la secuencia reguladora conduce a una transcripción de la secuencia codificante, sin provocar un cambio del marco de lectura en la secuencia codificante o incapacidad de la secuencia codificante para traducirse en la proteína o péptido deseado.

El término "secuencia de control de la expresión" o "secuencia reguladora" comprende, de acuerdo con la enseñanza, promotores, secuencias de unión a ribosomas y otros elementos de control, que controlan la transcripción de un ácido nucleico o la traducción del ARN derivado. En determinadas formas de realización de la enseñanza, se pueden controlar las secuencias reguladoras. La estructura precisa de las secuencias reguladoras puede variar dependiendo de la especie o dependiendo del tipo de célula, pero generalmente comprende secuencias 5' no transcritas y 5' y 3' no traducidas, que están involucradas en el inicio de la transcripción o traducción, tales como la Caja TATA, la secuencia de protección, secuencia CAAT y similares. En particular, las secuencias reguladoras no transcritas 5' comprenden una región promotora que incluye una secuencia promotora para el control transcripcional del gen unido funcionalmente. Las secuencias reguladoras también pueden comprender secuencias potenciadoras o secuencias activadoras cadena arriba.

Preferiblemente, de acuerdo con la enseñanza, el ARN que se va a expresar en una célula se introduce en dicha célula. En una realización de los métodos de acuerdo con la enseñanza, el ARN que se va a introducir en una célula se obtiene mediante la transcripción *in vitro* de una plantilla de ADN apropiada.

De acuerdo con la enseñanza, términos tales como "ARN capaz de expresarse" y "ARN que codifica" se usan indistintamente en este documento y con respecto a un péptido o polipéptido particular significa que el ARN, si está presente en el entorno apropiado, preferiblemente dentro de una célula, puede expresarse para producir dicho péptido o polipéptido. Preferiblemente, el ARN de acuerdo con la enseñanza puede interactuar con la maquinaria de traducción celular para proporcionar el péptido o polipéptido que es capaz de expresar.

Los términos tales como "transferir", "introducir" o "transfectar" se usan indistintamente en el presente documento y se refieren a la introducción de ácidos nucleicos, en particular ácidos nucleicos exógenos o heterólogos, en particular ARN en una célula. De acuerdo con la presente enseñanza, la célula puede formar parte de un órgano, un tejido y/o un organismo. De acuerdo con la presente enseñanza, la administración de un ácido nucleico se logra como ácido nucleico desnudo o en combinación con un reactivo de administración. Preferiblemente, la administración de ácidos nucleicos se realiza en forma de ácidos nucleicos desnudos. Preferiblemente, el ARN se administra en combinación con sustancias estabilizantes tales como inhibidores de RNasa. La presente enseñanza también prevé la introducción repetida de ácidos nucleicos en las células para permitir una expresión sostenida durante periodos de tiempo prolongados.

Las células se pueden transfectar con cualquier portador con el que se pueda asociar ARN, por ejemplo, formando complejos con el ARN o formando vesículas en las que el ARN está encerrado o encapsulado, lo que da como resultado una mayor estabilidad del ARN en comparación con el ARN desnudo. Los vehículos útiles de acuerdo con la enseñanza incluyen, por ejemplo, vehículos que contienen lípidos tales como lípidos catiónicos, liposomas, en particular liposomas catiónicos y micelas y nanopartículas. Los lípidos catiónicos pueden formar complejos con ácidos nucleicos cargados negativamente. Se puede usar cualquier lípido catiónico de acuerdo con la enseñanza.

Preferiblemente, la introducción de ARN que codifica un péptido o polipéptido en una célula, en particular en una célula presente *in vivo*, da como resultado la expresión de dicho péptido o polipéptido en la célula. En realizaciones particulares, se prefiere el direccionamiento de los ácidos nucleicos a células particulares. En tales realizaciones, un portador que se aplica para la administración del ácido nucleico a una célula (por ejemplo, un retrovirus o un liposoma) exhibe una molécula de direccionamiento. Por ejemplo, una molécula tal como un anticuerpo que es específico para una proteína de membrana de superficie en la célula diana o un ligando para un receptor en la célula diana puede

incorporarse al portador de ácido nucleico o puede unirse al mismo. En caso de que el ácido nucleico se administre mediante liposomas, las proteínas que se unen a una proteína de la membrana de la superficie que está asociada con la endocitosis pueden incorporarse en la formulación de liposomas para permitir el direccionamiento y/o la absorción. Dichas proteínas abarcan proteínas de la cápside de fragmentos de las mismas que son específicas para un tipo celular particular, anticuerpos contra proteínas que están internalizadas, proteínas que se dirigen a una ubicación intracelular, etc.

El término "célula" o "célula huésped" es preferiblemente una célula intacta, es decir, una célula con una membrana intacta que no ha liberado sus componentes intracelulares normales tales como enzimas, orgánulos o material genético. Una célula intacta es preferiblemente una célula viable, es decir, una célula viva capaz de llevar a cabo sus funciones metabólicas normales. Preferiblemente, dicho término se refiere de acuerdo con la enseñanza a cualquier célula que pueda transformarse o transfectarse con un ácido nucleico exógeno. El término "célula" incluye, de acuerdo con la enseñanza, células procariotas (por ejemplo, E. coli) o células eucariotas (por ejemplo, células dendríticas, células B, células CHO, células COS, células K562, células HEK293, células HELA, células de levadura y células de insectos). El ácido nucleico exógeno se puede encontrar dentro de la célula (i) libremente disperso como tal, (ii) incorporado en un vector recombinante, o (iii) integrado en el genoma de la célula huésped o ADN mitocondrial. Se prefieren particularmente las células de mamíferos, tales como células de humanos, ratones, hámsteres, cerdos, cabras y primates. Las células pueden derivar de un gran número de tipos de tejidos e incluyen células primarias y líneas celulares. Los ejemplos específicos incluyen queratinocitos, leucocitos de sangre periférica, células madre de la médula ósea y células madre embrionarias. En realizaciones adicionales, la célula es una célula presentadora de antígeno, en particular una célula dendrítica, un monocito o macrófago.

Una célula que comprende una molécula de ácido nucleico expresa preferiblemente el péptido o polipéptido codificado por el ácido nucleico.

El término "expansión clonal" se refiere a un proceso en el que se multiplica una entidad específica. En el contexto de la presente enseñanza, el término se usa preferiblemente en el contexto de una respuesta inmunológica en la que los linfocitos son estimulados por un antígeno, proliferan y se amplifica el linfocito específico que reconoce dicho antígeno. Preferiblemente, la expansión clonal conduce a la diferenciación de los linfocitos.

Los términos tales como "reducir" o "inhibir" se refieren a la capacidad de causar una disminución general, preferiblemente del 5% o más, del 10% o más, del 20% o más, más preferiblemente del 50% o más, y más preferiblemente del 75% o más, en el nivel. El término "inhibir" o frases similares incluye una inhibición completa o esencialmente completa, es decir, una reducción hasta cero o esencialmente hasta cero.

Los términos tales como "aumentar", "mejorar", "promover" o "prolongar" se refieren preferiblemente a un aumento, mejora, promoción o prolongación en aproximadamente al menos un 10%, preferiblemente al menos un 20%, preferiblemente al menos un 30%, preferiblemente al menos 40%, preferiblemente al menos 50%, preferiblemente al menos 80%, preferiblemente al menos 100%, preferiblemente al menos 200% y en particular al menos 300%. Estos términos también pueden referirse a un aumento, mejora, promoción o prolongación desde cero o un nivel no medible o no detectable hasta un nivel de más de cero o un nivel que es medible o detectable.

La presente enseñanza proporciona vacunas tales como vacunas contra el cáncer diseñadas sobre la base de modificaciones de aminoácidos o péptidos modificados que se predice que son inmunogénicos por los métodos de la presente enseñanza.

De acuerdo con la enseñanza, el término "vacuna" se refiere a una preparación farmacéutica (composición farmacéutica) o producto que tras la administración induce una respuesta inmune, en particular una respuesta inmune celular, que reconoce y ataca a un patógeno o una célula enferma tal como una célula cancerosa. Se puede usar una vacuna para la prevención o el tratamiento de una enfermedad. El término "vacuna contra el cáncer personalizada" o "vacuna contra el cáncer individualizada" se refiere a un paciente de cáncer en particular y significa que una vacuna contra el cáncer se adapta a las necesidades o circunstancias especiales de un paciente de cáncer individual.

En una realización, una vacuna proporcionada de acuerdo con la enseñanza puede comprender un péptido o polipéptido que comprende una o más modificaciones de aminoácidos o uno o más péptidos modificados que se predice que son inmunogénicos por los métodos de la enseñanza o un ácido nucleico, preferiblemente ARN, que codifica dicho péptido o polipéptido.

Las vacunas contra el cáncer proporcionadas de acuerdo con la enseñanza cuando se administran a un paciente proporcionan uno o más epítomos de células T adecuados para estimular, cebar y/o expandir células T específicas para el tumor del paciente. Las células T se dirigen preferiblemente contra las células que expresan antígenos de los que se derivan los epítomos de las células T. Por lo tanto, las vacunas descritas en el presente documento son preferiblemente capaces de inducir o promover una respuesta celular, preferiblemente actividad de células T citotóxicas, contra una enfermedad cancerosa caracterizada por la presentación de uno o más neoantígenos asociados a tumores con MHC de clase I. Dado que una vacuna proporcionada de acuerdo con la presente enseñanza se dirigirá a mutaciones específicas del cáncer, será específica para el tumor del paciente.

Una vacuna proporcionada de acuerdo con la enseñanza se refiere a una vacuna que cuando se administra a un paciente proporciona preferiblemente uno o más epítomos de células T, tales como 2 o más, 5 o más, 10 o más, 15 o más, 20 o más, 25 o más, 30 o más y preferiblemente hasta 60, hasta 55, hasta 50, hasta 45, hasta 40, hasta 35 o hasta 30 epítomos de células T, incorporando modificaciones de aminoácidos o péptidos modificados predichos como inmunogénicos por los métodos de la enseñanza. Dichos epítomos de células T también se denominan "neopítomos" en el presente documento. La presentación de estos epítomos por las células de un paciente, en particular las células presentadoras de antígenos, preferiblemente da como resultado células T dirigidas a los epítomos cuando se unen al MHC y, por lo tanto, al tumor del paciente, preferiblemente al tumor primario, así como a las metástasis tumorales, que expresan antígenos a partir de los cuales los epítomos de células T se derivan y presentan los mismos epítomos en la superficie de las células tumorales.

Los métodos de la enseñanza pueden comprender la etapa adicional de determinar la utilidad de las modificaciones de aminoácidos identificadas o péptidos modificados para la vacunación contra el cáncer. Por lo tanto, las etapas adicionales pueden involucrar uno o más de las siguientes: (i) evaluar si las modificaciones están ubicadas en epítomos presentados por MHC conocidos o predichos, (ii) pruebas *in vitro* y/o *in silico* si las modificaciones están ubicadas en epítomos presentados por MHC, por ejemplo, probar si las modificaciones son parte de secuencias de péptidos que se procesan y/o presentan como epítomos presentados por MHC, y (iii) probar *in vitro* si los epítomos modificados previstos, en particular cuando están presentes en su contexto de secuencia natural, por ejemplo, cuando están flanqueadas por secuencias de aminoácidos que también flanquean dichos epítomos en la proteína de origen natural, y cuando se expresan en células presentadoras de antígenos son capaces de estimular células T tales como células T del paciente que tienen la especificidad deseada. Cada una de estas secuencias flanqueantes puede comprender 3 o más, 5 o más, 10 o más, 15 o más, 20 o más y preferiblemente hasta 50, hasta 45, hasta 40, hasta 35 o hasta 30 aminoácidos y pueden flanquear la secuencia del epítomo de terminal N y/o el terminal C.

Los péptidos modificados determinados de acuerdo con la enseñanza pueden clasificarse por su utilidad como epítomos para la vacunación contra el cáncer. Por lo tanto, en un aspecto, el método de la enseñanza comprende un proceso analítico manual o informático en el que los péptidos modificados identificados se analizan y seleccionan por su utilidad en la vacuna respectiva que se va a proporcionar. En una realización preferida, dicho proceso analítico es un proceso basado en algoritmos computacionales. Preferiblemente, dicho proceso analítico comprende determinar y/o clasificar epítomos de acuerdo con una predicción de su capacidad de ser inmunogénicos.

Los neoepítomos identificados de acuerdo con la enseñanza y proporcionados por una vacuna de la enseñanza están presentes preferiblemente en forma de un polipéptido que comprende dichos neoepítomos tales como un polipéptido poliepitópico o un ácido nucleico, en particular ARN, que codifica dicho polipéptido. Además, los neoepítomos pueden estar presentes en el polipéptido en forma de una secuencia de vacuna, es decir, presentes en su contexto de secuencia natural por ejemplo, flanqueados por secuencias de aminoácidos que también flanquean dichos epítomos en la proteína de origen natural. Cada una de estas secuencias flanqueantes puede comprender 5 o más, 10 o más, 15 o más, 20 o más y preferiblemente hasta 50, hasta 45, hasta 40, hasta 35 o hasta 30 aminoácidos y pueden flanquear la secuencia del epítomo del terminal N y/o terminal C. Por lo tanto, una secuencia de vacuna puede comprender 20 o más, 25 o más, 30 o más, 35 o más, 40 o más y preferiblemente hasta 50, hasta 45, hasta 40, hasta 35 o hasta 30 aminoácidos. En una realización, los neoepítomos y/o las secuencias de la vacuna se alinean en el polipéptido de cabeza a cola.

En una realización, los neoepítomos y/o secuencias de vacuna están espaciados por enlazadores, en particular enlazadores neutros. El término "enlazador" de acuerdo con la enseñanza se refiere a un péptido añadido entre dos dominios peptídicos, tales como epítomos o secuencias de vacunas, para conectar dichos dominios peptídicos. No existe una limitación particular con respecto a la secuencia del enlazador. Sin embargo, se prefiere que la secuencia enlazadora reduzca el impedimento estérico entre los dos dominios peptídicos, esté bien traducida y soporte o permita el procesamiento de los epítomos. Además, el enlazador debe tener pocos elementos de secuencia inmunogénica o ninguno. Los enlazadores preferiblemente no deberían crear neoepítomos no endógenos como los generados a partir de la sutura de unión entre neoepítomos adyacentes, que podrían generar reacciones inmunes no deseadas. Por lo tanto, la vacuna poliepitópica debería contener preferiblemente secuencias enlazadoras que sean capaces de reducir el número de epítomos de unión al MHC no deseados. Hoyt et al., (EMBO J. 25 (8), 1720-9, 2006) y Zhang et al. (J. Biol. Chem., 279 (10), 8635-41, 2004) han demostrado que las secuencias ricas en glicina perjudican el procesamiento proteasómico y, por lo tanto, el uso de secuencias enlazadoras ricas en glicina actúa para minimizar el número de péptidos que contienen enlazador que pueden ser procesados por el proteasoma. Además, se observó que la glicina inhibía una fuerte unión en las posiciones de los surcos de unión del MHC (Abastado et al., J. Immunol. 151 (7), 3569-75, 1993). Schlessinger et al., (Proteins, 61 (1), 115-26, 2005) habían descubierto que los aminoácidos glicina y serina incluidos en una secuencia de aminoácidos dan como resultado una proteína más flexible que es más eficientemente traducida y procesada por el proteasoma, lo que permite un mejor acceso a la neoepítomos codificados. Cada enlazador puede comprender 3 o más, 6 o más, 9 o más, 10 o más, 15 o más, 20 o más y preferiblemente hasta 50, hasta 45, hasta 40, hasta 35 o hasta 30 aminoácidos. Preferiblemente, el enlazador está enriquecido en aminoácidos glicina y/o serina. Preferiblemente, al menos el 50%, al menos el 60%, al menos el 70%, al menos el 80%, al menos el 90% o al menos el 95% de los aminoácidos del enlazador son glicina y/o serina. En una realización preferida, un enlazador está compuesto sustancialmente por los aminoácidos glicina y serina. En una realización, el enlazador comprende la secuencia de aminoácidos (GGG)_a(GSS)_b(GGG)_c(SSG)_d(GSG)_e en la que a, b, c, d y e es

independientemente un número seleccionado entre 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 y en la que $a + b + c + d + e$ son diferentes de 0 y preferiblemente son 2 o más, 3 o más, 4 o más o 5 o más. En una realización, el enlazador comprende una secuencia como se describe en el presente documento que incluye las secuencias del enlazador descritas en los ejemplos tales como la secuencia GGSGGGGSG.

5 En una realización particularmente preferida, un polipéptido que incorpora uno o más neoepítomos tales como un polipéptido poliepitópico de acuerdo con la presente enseñanza se administra a un paciente en forma de ácido nucleico, preferiblemente ARN tal como ARN *in vitro* transcrito o sintético, que puede expresarse en células de un paciente, tal como células presentadoras de antígenos, para producir el polipéptido. La presente enseñanza también
10 prevé la administración de uno o más polipéptidos multiepitópicos que para el propósito de la presente enseñanza están comprendidos por el término "polipéptido poliepitópico", preferiblemente en forma de un ácido nucleico, preferiblemente ARN tal como ARN *in vitro* transcrito o sintético, que puede expresarse en células de un paciente, tales como células presentadoras de antígenos, para producir uno o más polipéptidos. En el caso de una
15 administración de más de un polipéptido multiepitópico, los neoepítomos proporcionados por los diferentes polipéptidos multiepitópicos pueden ser diferentes o superponerse parcialmente. Una vez presente en las células de un paciente, tal como las células presentadoras de antígenos, el polipéptido de acuerdo con la enseñanza se procesa para producir los neoepítomos identificados de acuerdo con la enseñanza. La administración de una vacuna proporcionada de acuerdo con la enseñanza puede proporcionar epítomos presentados por el MHC de clase II que son capaces de
20 provocar una respuesta de células T colaboradoras CD4+ contra células que expresan antígenos de los que se derivan los epítomos presentados por el MHC. Alternativa o adicionalmente, la administración de una vacuna proporcionada de acuerdo con la enseñanza puede proporcionar epítomos presentados por MHC de clase I que son capaces de provocar una respuesta de células T CD8+ contra células que expresan antígenos de los que se derivan los epítomos presentados por el MHC. Además, la administración de una vacuna proporcionada de acuerdo con la enseñanza puede proporcionar uno o más neoepítomos (incluidos neoepítomos y neoepítomos conocidos identificados de acuerdo
25 con la enseñanza) así como uno o más epítomos que no contienen mutaciones somáticas específicas del cáncer pero que son expresados por células cancerosas y preferiblemente induciendo una respuesta inmune contra las células cancerosas, preferiblemente una respuesta inmune específica del cáncer. En una realización, la administración de una vacuna proporcionada de acuerdo con la enseñanza proporciona neoepítomos que son epítomos presentados por el MHC de clase II y/o son capaces de provocar una respuesta de células T colaboradoras CD4+ contra células que expresan antígenos de los que se derivan los epítomos presentados por el MHC así como epítomos que no contienen mutaciones somáticas específicas de cáncer que son epítomos presentados por el MHC de clase I y/o son capaces de
30 provocar una respuesta de células T CD8+ contra células que expresan antígenos de los que se derivan los epítomos presentados por el MHC. En una realización, los epítomos que no contienen mutaciones somáticas específicas de cáncer se derivan de un antígeno tumoral. En una realización, los neoepítomos y epítomos que no contienen mutaciones somáticas específicas del cáncer tienen un efecto sinérgico en el tratamiento del cáncer. Preferiblemente, una vacuna proporcionada de acuerdo con la enseñanza es útil para la estimulación poliepitópica de respuestas de células T citotóxicas y/o colaboradoras.

La vacuna proporcionada de acuerdo con la enseñanza puede ser una vacuna recombinante.

40 El término "recombinante" en el contexto de la presente enseñanza significa "fabricado mediante ingeniería genética". Preferiblemente, una "entidad recombinante" tal como un polipéptido recombinante en el contexto de la presente enseñanza no se produce de forma natural, y preferiblemente es el resultado de una combinación de entidades tales como secuencias de aminoácidos o ácidos nucleicos que no se combinan en la naturaleza. Por ejemplo, un polipéptido
45 recombinante en el contexto de la presente enseñanza puede contener varias secuencias de aminoácidos tales como neoepítomos o secuencias de vacuna derivadas de diferentes proteínas o diferentes porciones de la misma proteína fusionadas juntas, por ejemplo, mediante enlaces peptídicos o enlazadores apropiados.

El término "de origen natural", como se usa en este documento, se refiere al hecho de que un objeto se puede encontrar en la naturaleza. Por ejemplo, un péptido o ácido nucleico que está presente en un organismo (incluidos los virus) y que puede aislarse de una fuente en la naturaleza y que no ha sido modificado intencionalmente por el hombre en el laboratorio es de origen natural.

Los agentes, composiciones y métodos descritos en el presente documento se pueden usar para tratar a un sujeto con una enfermedad, por ejemplo, una enfermedad caracterizada por la presencia de células enfermas que expresan un antígeno y presentan un fragmento del mismo. Las enfermedades particularmente preferidas son las enfermedades cancerosas. Los agentes, composiciones y métodos descritos en este documento también pueden usarse para inmunización o vacunación para prevenir una enfermedad descrita en este documento.

60 De acuerdo con la enseñanza, el término "enfermedad" se refiere a cualquier estado patológico, incluidas las enfermedades cancerosas, en particular las formas de enfermedades cancerosas descritas en el presente documento.

El término "normal" se refiere al estado sano o las condiciones en un sujeto o tejido sano, es decir, condiciones no patológicas, en las que "saludable" significa preferiblemente no canceroso.

65

- 5 "Enfermedad que implica células que expresan un antígeno" significa de acuerdo con la enseñanza que se detecta la expresión del antígeno en células de un tejido u órgano enfermo. La expresión en las células de un tejido u órgano enfermo puede aumentar en comparación con el estado en un tejido u órgano sano. Un aumento se refiere a un aumento de al menos un 10%, en particular al menos un 20%, al menos un 50%, al menos un 100%, al menos un 200%, al menos un 500%, al menos un 1.000%, al menos un 10.000% o incluso más. En una realización, la expresión solo se encuentra en un tejido enfermo, mientras que la expresión en un tejido sano está reprimida. De acuerdo con la enseñanza, las enfermedades que involucran o están asociadas con células que expresan un antígeno incluyen enfermedades cancerosas.
- 10 De acuerdo con la enseñanza, el término "tumor" o "enfermedad tumoral" se refiere a un crecimiento anormal de células (llamadas células neoplásicas, células tumorigénicas o células tumorales) que preferiblemente forman una hinchazón o una lesión. Por "célula tumoral" se entiende una célula anormal que crece mediante una proliferación celular rápida y descontrolada y continúa creciendo después de que cesan los estímulos que iniciaron el nuevo crecimiento. Los tumores muestran una falta parcial o total de organización estructural y coordinación funcional con el
- 15 tejido normal y, por lo general, forman una masa distinta de tejido, que puede ser benigno, premaligno o maligno.
- El cáncer (término médico: neoplasia maligna) es una clase de enfermedades en las que un grupo de células muestra un crecimiento descontrolado (división más allá de los límites normales), invasión (intrusión y destrucción de tejidos adyacentes) y, a veces, metástasis (diseminación a otras ubicaciones del cuerpo a través de la linfa o la sangre). Estas
- 20 tres propiedades malignas de los cánceres los diferencian de los tumores benignos, que son autolimitados y no invaden ni hacen metástasis. La mayoría de los cánceres forman un tumor, pero algunos, como la leucemia, no lo hacen. Malignidad, neoplasia maligna y tumor maligno son esencialmente sinónimos de cáncer.
- La neoplasia es una masa anormal de tejido como resultado de una neoplasia. La neoplasia (nuevo crecimiento en griego) es la proliferación anormal de células. El crecimiento de las células excede y no está coordinado con el de los
- 25 tejidos normales que lo rodean. El crecimiento persiste de la misma manera excesiva incluso después de la cesación de los estímulos. Por lo general, causa un bulto o un tumor. Las neoplasias pueden ser benignas, premalignas o malignas.
- 30 El "crecimiento de un tumor" o "crecimiento tumoral" de acuerdo con la enseñanza se refiere a la tendencia de un tumor a aumentar su tamaño y/o a la tendencia de las células tumorales a proliferar.
- Para los propósitos de la presente enseñanza, los términos "cáncer" y "enfermedad cancerosa" se usan de manera intercambiable con los términos "tumor" y "enfermedad tumoral".
- 35 Los cánceres se clasifican por el tipo de célula que se asemeja al tumor y, por lo tanto, el tejido que se presume es el origen del tumor. Estos son la histología y la ubicación, respectivamente.
- El término "cáncer" de acuerdo con la enseñanza comprende carcinomas, adenocarcinomas, blastomas, leucemias, seminomas, melanomas, teratomas, linfomas, neuroblastomas, gliomas, cáncer de recto, cáncer de endometrio, cáncer de riñón, cáncer suprarrenal, cáncer de tiroides, cáncer de sangre, cáncer de piel, cáncer de cerebro, 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, cáncer de ganglio linfático, cáncer de esófago, cáncer colorrectal, cáncer de páncreas, cáncer de oído, nariz y garganta (ENT), cáncer de mama, cáncer de próstata, cáncer de útero,
- 40 cáncer de ovario y cáncer de pulmón y sus metástasis. Ejemplos de los mismos son carcinomas de pulmón, carcinomas de mama, carcinomas de próstata, carcinomas de colon, carcinomas de células renales, carcinomas de cuello uterino o metástasis de los tipos de cáncer o tumores descritos anteriormente. El término cáncer de acuerdo con la enseñanza también comprende metástasis de cáncer y recidiva de cáncer.
- 50 Por "metástasis" se entiende la diseminación de células cancerosas desde su sitio original a otra parte del cuerpo. La formación de metástasis es un proceso muy complejo y depende del desprendimiento de células malignas del tumor primario, la invasión de la matriz extracelular, la penetración de las membranas basales endoteliales para ingresar a la cavidad y vasos corporales, y luego, después de ser transportadas por la sangre, infiltración de órganos diana. Finalmente, el crecimiento de un nuevo tumor, es decir, un tumor secundario o un tumor metastásico, en el sitio diana
- 55 depende de la angiogénesis. La metástasis tumoral a menudo ocurre incluso después de la extirpación del tumor primario porque las células o componentes tumorales pueden permanecer y desarrollar potencial metastásico. En una realización, el término "metástasis" de acuerdo con la enseñanza se refiere a "metástasis distante" que se refiere a una metástasis que está alejada del tumor primario y del sistema de ganglios linfáticos regionales.
- 60 Las células de un tumor secundario o metastásico son como las del tumor original. Esto significa, por ejemplo, que, si el cáncer de ovario hace metástasis en el hígado, el tumor secundario está formado por células ováricas anormales, no por células hepáticas anormales. El tumor en el hígado se denomina cáncer de ovario metastásico, no cáncer de hígado.
- 65 El término "células tumorales circulantes" o "CTC" se refiere a células que se han desprendido de un tumor primario o metástasis tumorales y circulan en el torrente sanguíneo. Las CTC pueden constituir semillas para el crecimiento

posterior de tumores adicionales (metástasis) en diferentes tejidos. Las células tumorales circulantes se encuentran en frecuencias del orden de 1-10 CTC por ml de sangre total en pacientes con enfermedad metastásica. Se han desarrollado métodos de investigación para aislar las CTC. Se han descrito varios métodos de investigación en la técnica para aislar las CTC, por ejemplo, técnicas que utilizan el hecho de que las células epiteliales expresan comúnmente la proteína de adhesión celular EpCAM, que está ausente en las células sanguíneas normales. La captura basada en perlas inmunomagnéticas implica el tratamiento de muestras de sangre con anticuerpos contra EpCAM que se han conjugado con partículas magnéticas, seguido de la separación de las células marcadas en un campo magnético. A continuación, las células aisladas se tiñen con anticuerpo contra otro marcador epitelial, citoqueratina, así como un marcador de leucocitos común CD45, para distinguir las CTC raras de los glóbulos blancos contaminantes. Este enfoque robusto y semiautomático identifica las CTC con un rendimiento promedio de aproximadamente 1 CTC/ml y una pureza del 0,1% (Allard et al., 2004: Clin Cancer Res 10, 6897-6904). Un segundo método para aislar las CTC utiliza un dispositivo de captura de CTC basado en microfluidos que implica hacer fluir sangre completa a través de una cámara incrustada con 80.000 micropostes que se han vuelto funcionales al recubrir con anticuerpos contra EpCAM. Luego, las CTC se tiñen con anticuerpos secundarios contra citoqueratina o marcadores específicos de tejido, tales como PSA en el cáncer de próstata o HER2 en el cáncer de mama, y se visualizan mediante escaneo automático de micropostes en múltiples planos a lo largo de coordenadas tridimensionales. Los chips de CTC son capaces de identificar células tumorales circulantes positivas para citoqueratina en pacientes con un rendimiento medio de 50 células/ml y una pureza que oscila entre el 1 y el 80% (Nagrath et al., 2007: Nature 450, 1235-1239). Otra posibilidad para aislar las CTC es utilizar la prueba de células tumorales circulantes (CTC) CellSearch^{MR} de Veridex, LLC (Raritan, NJ) que captura, identifica y cuenta las CTC en un tubo de sangre. El sistema CellSearch^{MR} es una metodología aprobada por la Administración de Alimentos y Fármacos de los Estados Unidos (FDA) para el recuento de CTC en sangre completa que se basa en una combinación de etiquetado inmunomagnético y microscopía digital automatizada. Hay otros métodos para aislar CTC descritos en la literatura, todos los cuales pueden usarse junto con la presente enseñanza.

Una recaída o recurrencia ocurre cuando una persona se ve afectada nuevamente por una condición que la afectó en el pasado. Por ejemplo, si un paciente ha sufrido una enfermedad tumoral, ha recibido un tratamiento satisfactorio de dicha enfermedad y vuelve a desarrollar dicha enfermedad, dicha enfermedad recién desarrollada puede considerarse como recaída o recurrencia. Sin embargo, de acuerdo con la enseñanza, una recaída o recurrencia de una enfermedad tumoral puede ocurrir, pero no necesariamente, en el sitio de la enfermedad tumoral original. Así, por ejemplo, si una paciente ha sufrido un tumor de ovario y ha recibido un tratamiento exitoso, una recaída o recurrencia puede ser la aparición de un tumor de ovario o la aparición de un tumor en un sitio diferente al ovario. Una recaída o recurrencia de un tumor también incluye situaciones en las que un tumor ocurre en un sitio diferente al sitio del tumor original así como en el sitio del tumor original. Preferiblemente, el tumor original para el que el paciente ha recibido un tratamiento es un tumor primario y el tumor en un sitio diferente al sitio del tumor original es un tumor secundario o metastásico.

Por "tratar" se entiende administrar un compuesto o composición como se describe en el presente documento a un sujeto para prevenir o eliminar una enfermedad, incluida la reducción del tamaño de un tumor o el número de tumores en un sujeto; detener o retrasar una enfermedad en un sujeto; inhibir o retrasar el desarrollo de una nueva enfermedad en un sujeto; disminuir la frecuencia o gravedad de los síntomas y/o las recurrencias en un sujeto que actualmente tiene o que ha tenido una enfermedad anteriormente; y/o prolongar, es decir, aumentar la vida útil del sujeto. En particular, el término "tratamiento de una enfermedad" incluye curar, acortar la duración, mejorar, prevenir, ralentizar o inhibir la progresión o empeorar, o prevenir o retrasar la aparición de una enfermedad o los síntomas de la misma.

Por "estar en riesgo" se entiende un sujeto, es decir, un paciente, que se identifica con una probabilidad mayor de lo normal de desarrollar una enfermedad, en particular cáncer, en comparación con la población general. Además, un sujeto que ha tenido, o que tiene actualmente, una enfermedad, en particular cáncer, es un sujeto que tiene un mayor riesgo de desarrollar una enfermedad, ya que dicho sujeto puede continuar desarrollando una enfermedad. Los sujetos que actualmente tienen, o que han tenido, un cáncer también tienen un mayor riesgo de metástasis del cáncer.

El término "inmunoterapia" se refiere a un tratamiento que implica la activación de una reacción inmunitaria específica. En el contexto de la presente enseñanza, términos como "proteger", "prevenir", "profiláctico", "preventivo" o "protector" se relacionan con la prevención o el tratamiento, o ambos, de la aparición y/o propagación de una enfermedad en un sujeto y, en particular, para minimizar la posibilidad de que un sujeto desarrolle una enfermedad o para retrasar el desarrollo de una enfermedad. Por ejemplo, una persona en riesgo de tener un tumor, como se describió anteriormente, sería un candidato para la terapia para prevenir un tumor.

Una administración profiláctica de una inmunoterapia, por ejemplo, una administración profiláctica de una vacuna de la enseñanza, protege preferiblemente al receptor del desarrollo de una enfermedad. Una administración terapéutica de una inmunoterapia, por ejemplo, una administración terapéutica de una vacuna de la enseñanza, puede conducir a la inhibición del progreso/crecimiento de la enfermedad. Esto comprende la desaceleración del progreso/crecimiento de la enfermedad, en particular una interrupción del progreso de la enfermedad, que preferiblemente conduce a la eliminación de la enfermedad.

La inmunoterapia se puede realizar usando cualquiera de una variedad de técnicas, en las que los agentes proporcionados en el presente documento funcionan para eliminar las células enfermas de un paciente. Tal eliminación

puede tener lugar como resultado de potenciar o inducir una respuesta inmune en un paciente específico para un antígeno o una célula que expresa un antígeno.

5 En determinadas realizaciones, la inmunoterapia puede ser inmunoterapia activa, en la que el tratamiento se basa en la estimulación *in vivo* del sistema inmunitario endógeno del huésped para reaccionar contra las células enfermas con la administración de agentes modificadores de la respuesta inmunitaria (tales como polipéptidos y ácidos nucleicos como los proporcionados en este documento).

10 Los agentes y composiciones proporcionados en este documento pueden usarse solos o en combinación con regímenes terapéuticos convencionales tales como cirugía, irradiación, quimioterapia y/o trasplante de médula ósea (autólogo, singénico, alogénico o no relacionado).

15 El término "inmunización" o "vacunación" describe el proceso de tratar a un sujeto con el propósito de inducir una respuesta inmune por razones terapéuticas o profilácticas.

El término "*in vivo*" se refiere a la situación en un sujeto.

20 Los términos "sujeto", "individuo", "organismo" o "paciente" se usan indistintamente y se refieren a vertebrados, preferiblemente mamíferos. Por ejemplo, los mamíferos en el contexto de la presente enseñanza son humanos, primates no humanos, animales domesticados tales como perros, gatos, ovejas, vacas, cabras, cerdos, caballos, etc., animales de laboratorio tales como ratones, ratas, conejos, conejillos de Indias, etc., así como animales en cautiverio tales como animales de zoológicos. El término "animal" como se usa en este documento también incluye humanos. El término "sujeto" también puede incluir un paciente, es decir, un animal, preferiblemente un ser humano que tiene una enfermedad, preferiblemente una enfermedad como se describe en el presente documento.

25 El término "autólogo" se usa para describir cualquier cosa que se derive del mismo sujeto. Por ejemplo, "trasplante autólogo" se refiere a un trasplante de tejido u órganos derivados del mismo sujeto. Tales procedimientos son ventajosos porque superan la barrera inmunológica que de otro modo da como resultado el rechazo.

30 El término "heterólogo" se usa para describir algo que consta de múltiples elementos diferentes. Por ejemplo, la transferencia de la médula ósea de un individuo a otro individuo constituye un trasplante heterólogo. Un gen heterólogo es un gen derivado de una fuente distinta al sujeto.

35 Como parte de la composición para una inmunización o una vacunación, preferiblemente uno o más agentes como se describen en este documento se administran junto con uno o más adyuvantes para inducir una respuesta inmune o para aumentar una respuesta inmune. El término "adyuvante" se refiere a compuestos que prolongan o potencian o aceleran una respuesta inmunitaria. La composición de la presente enseñanza ejerce preferiblemente su efecto sin adición de adyuvantes. Además, la composición de la presente solicitud puede contener cualquier adyuvante conocido. Los adyuvantes comprenden un grupo heterogéneo de compuestos tales como emulsiones oleosas (por ejemplo, adyuvantes de Freund), compuestos minerales (tales como alumbre), productos bacterianos (tales como la toxina de *Bordetella pertussis*), liposomas y complejos inmunoestimulantes. Ejemplos de adyuvantes son 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 (So et al., 1997, Mol. Cells 7: 178-186), adyuvantes incompletos de Freund, adyuvantes completos de Freund, vitamina E, montánido, alumbre, oligonucleótidos CpG (Krieg et al., 45 1995, Nature 374: 546-549), y diversas emulsiones de agua en aceite que se preparan a partir de aceites biológicamente degradables tales como escualeno y/o tocoferol.

50 También se pueden administrar otras sustancias que estimulan una respuesta inmune del paciente. Es posible, por ejemplo, usar citocinas en una vacunación, debido a sus propiedades reguladoras sobre los linfocitos. Dichas citocinas comprenden, por ejemplo, interleucina-12 (IL-12) que se demostró que aumenta las acciones protectoras de las vacunas (véase Science 268: 1432-1434, 1995), GM-CSF e IL-18.

55 Hay una serie de compuestos que potencian una respuesta inmune y que, por lo tanto, pueden usarse en una vacunación. Dichos compuestos comprenden moléculas coestimulantes proporcionadas en forma de proteínas o ácidos nucleicos tales como B7-1 y B7-2 (CD80 y CD86, respectivamente).

60 De acuerdo con la enseñanza, una muestra corporal puede ser una muestra de tejido, que incluye fluidos corporales y/o una muestra celular. Tales muestras corporales pueden obtenerse de manera convencional, tal como mediante biopsia de tejido, incluida la biopsia por punción, y extrayendo sangre, aspirado bronquial, esputo, orina, heces u otros fluidos corporales. De acuerdo con la enseñanza, el término "muestra" también incluye muestras procesadas tales como fracciones o aislados de muestras biológicas, por ejemplo, ácidos nucleicos o aislados de células.

65 Los agentes tales como vacunas y composiciones descritas en el presente documento pueden administrarse mediante cualquier ruta convencional, incluso mediante inyección o infusión. La administración puede realizarse, por ejemplo, por vía oral, intravenosa, intraperitoneal, intramuscular, subcutánea o transdérmica. En una realización, la administración se lleva a cabo por vía intranodal, por ejemplo mediante inyección en un ganglio linfático. Otras formas

de administración prevén la transfección *in vitro* de células presentadoras de antígeno tales como células dendríticas con los ácidos nucleicos descritos en el presente documento, seguida de la administración de las células presentadoras de antígeno.

5 Los agentes descritos en este documento se administran en cantidades eficaces. Una "cantidad eficaz" se refiere a la cantidad que logra una reacción deseada o un efecto deseado solo o junto con dosis adicionales. En el caso del tratamiento de una enfermedad particular o de una afección particular, la reacción deseada se refiere preferiblemente a la inhibición del curso de la enfermedad. Esto comprende ralentizar el progreso de la enfermedad y, en particular, interrumpir o revertir el progreso de la enfermedad. La reacción deseada en un tratamiento de una enfermedad o de una afección también puede ser un retraso del inicio o una prevención del inicio de dicha enfermedad o dicha condición.

15 Una cantidad eficaz de un agente descrito en el presente documento dependerá de la afección a tratar, la gravedad de la enfermedad, los parámetros individuales del paciente, incluida la edad, la afección fisiológica, el tamaño y el peso, la duración del tratamiento, el tipo de una terapia acompañante (si está presente), la vía de administración específica y factores similares. Por consiguiente, las dosis administradas de los agentes descritos en este documento pueden depender de varios de dichos parámetros. En el caso de que una reacción en un paciente sea insuficiente con una dosis inicial, se pueden usar dosis más altas (o dosis efectivamente más altas logradas por una vía de administración diferente, más localizada).

20 Las composiciones farmacéuticas descritas en el presente documento son preferiblemente estériles y contienen una cantidad eficaz de la sustancia terapéuticamente activa para generar la reacción deseada o el efecto deseado.

25 Las composiciones farmacéuticas descritas en el presente documento se administran generalmente en cantidades farmacéuticamente compatibles y en una preparación farmacéuticamente compatible. El término "farmacéuticamente compatible" se refiere a un material no tóxico que no interactúa con la acción del componente activo de la composición farmacéutica. Las preparaciones de este tipo pueden contener habitualmente sales, sustancias tampón, conservantes, vehículos, sustancias complementarias que mejoran la inmunidad tales como adyuvantes, por ejemplo, oligonucleótidos CpG, citocinas, quimiocinas, saponina, GM-CSF y/o ARN y, en su caso, otros compuestos terapéuticamente activos. Cuando se usan en medicina, las sales deben ser farmacéuticamente compatibles. Sin embargo, las sales que no son farmacéuticamente compatibles pueden usarse para preparar sales farmacéuticamente compatibles y están incluidas en la enseñanza. Las sales farmacológica y farmacéuticamente compatibles de este tipo comprenden de forma no limitativa las preparadas a partir de los siguientes ácidos: ácido clorhídrico, bromhídrico, sulfúrico, nítrico, fosfórico, maleico, acético, salicílico, cítrico, fórmico, malónico, succínico y similares. También se pueden preparar sales farmacéuticamente compatibles como sales de metales alcalinos o sales de metales alcalinotérreos, tales como sales de sodio, sales de potasio o sales de calcio.

40 Una composición farmacéutica descrita en el presente documento puede comprender un vehículo farmacéuticamente compatible. El término "vehículo" se refiere a un componente orgánico o inorgánico, de naturaleza natural o sintética, en el que el componente activo se combina para facilitar la aplicación. De acuerdo con la enseñanza, el término "vehículo farmacéuticamente compatible" incluye uno o más rellenos sólidos o líquidos compatibles, diluyentes o sustancias encapsulantes que son adecuados para la administración a un paciente. Los componentes de la composición farmacéutica de la enseñanza son normalmente tales que no se produce interacción que perjudique sustancialmente la eficacia farmacéutica deseada.

45 Las composiciones farmacéuticas descritas en este documento pueden contener sustancias tampón adecuadas tales como ácido acético en una sal, ácido cítrico en una sal, ácido bórico en una sal y ácido fosfórico en una sal.

50 Las composiciones farmacéuticas pueden contener, cuando sea apropiado, también conservantes adecuados tales como cloruro de benzalconio, clorobutanol, parabeno y timerosal.

Las composiciones farmacéuticas se proporcionan normalmente en una forma de dosificación uniforme y se pueden preparar de una manera ya conocida. Las composiciones farmacéuticas de la enseñanza pueden estar en forma de cápsulas, comprimidos, grageas, soluciones, suspensiones, jarabes, elixires o en forma de emulsión, por ejemplo.

55 Las composiciones adecuadas para la administración parenteral normalmente comprenden una preparación acuosa o no acuosa estéril del compuesto activo, que preferiblemente es isotónica a la sangre del receptor. Ejemplos de vehículos compatibles y disolventes son la solución de Ringer y la solución isotónica de cloruro de sodio. Además, se utilizan aceites fijos normalmente estériles como medio de solución o suspensión.

60 La presente enseñanza se describe en detalle mediante las figuras y ejemplos siguientes, que se utilizan sólo con fines ilustrativos y no pretenden ser limitantes. Debido a la descripción y los ejemplos, el experto en la materia puede acceder a otras formas de realización que también se incluyen en la enseñanza.

Figuras

65 Figura 1. Descripción general de la predicción de unión al MHC

Figura 2. Análisis de inmunogenicidad en función de la puntuación de M_{mut} para 50 mutaciones B16F10 priorizadas y 82 mutaciones CT26.WT priorizadas (132 mutaciones en total), de las cuales 30 eran inmunogénicas. Todas las vacunaciones se realizaron con ARN. Para B 16F10, se ensayó la inmunogenicidad desafiando las BMDC con ARN y midiendo la respuesta inmune de los esplenocitos con ELISPOT y FACS. Para CT26.WT se ensayó la inmunogenicidad desafiando las BMDC con ARN y péptidos por separado y midiendo la respuesta inmune de los esplenocitos con ELISPOT; una mutación se consideraba inmunogénica si el péptido o el ARN registraban una respuesta inmune. A. Una distribución acumulativa de mutaciones inmunogénicas en función de la puntuación de M_{mut} . El gráfico muestra el número total de mutaciones por debajo de una puntuación determinada de M_{mut} (rojo), de éstos, el número de mutaciones que fueron inmunogénicas (azul) y el porcentaje de mutaciones inmunogénicas del total (negro). B. Histograma del porcentaje de mutaciones inmunogénicas por compartimiento de M_{mut} para los siguientes intervalos: $\leq 0,3$, $(0,3; 1)$, > 1 . Los errores que se muestran son errores estándar.

Figura 3. Análisis de la inmunogenicidad de B16F10 y CT26.WT en función de M_{mut} . Distribución acumulativa de mutaciones inmunogénicas en función de la puntuación de M_{mut} para B16 (A) y CT26 (C). Histograma del porcentaje de mutaciones inmunogénicas por compartimiento de M_{mut} para los siguientes intervalos: $(0; 1)$, $(0,3; 1)$, $(1, \infty)$ para B16 (B) y CT26 (D). Las Figuras A y B se basan en los análisis de 50 mutaciones priorizadas de B16F10, de las cuales 12 eran inmunogénicas. Las Figuras C y D se basan en el análisis de 82 mutaciones priorizadas de B16F10, de las cuales 30 eran inmunogénicas. Para más detalles, véase la leyenda de la Figura 2. Los errores son errores estándar.

Figura 4. Modelos de inmunogenicidad e hipótesis de control. La inmunogenicidad de clase I, denotada por H_A , asume que tanto los epítomos WT como MUT son presentados por células, y que la suficiencia de mutación alteró las propiedades fisicoquímicas del aminoácido para que el sistema inmunológico registre este cambio y genere una respuesta inmune (denotada por el rayo). La hipótesis de H_n , que sirve como control para H_A , es simplemente la hipótesis de H_A invertida, es decir, que la mutación no alteró significativamente las propiedades fisicoquímicas del aminoácido y, por lo tanto, tiene una menor probabilidad de ser "detectada" por el sistema inmunológico y generar una respuesta inmune. En la inmunogenicidad de clase II (H_B U H_C) no se presenta el epítomo WT pero sí el epítomo MUT. H_B y H_C se distinguen por puntuaciones T altas ($T > \tau$) frente a bajas ($T \leq \tau$), respectivamente. Nótese que para $\alpha^* = \alpha$, el modelo H_{BC1} de inmunogenicidad ($M_{mut} < \beta$) es una combinación de los cuatro grupos: $H_{BC1} = U[H_A, H_B, H_C, H_n]$.

Figura 5. Relación hipotética del puntaje T con la inmunogenicidad. De acuerdo con el modelo de inmunogenicidad de clase I, durante el desarrollo de las células T, se eliminaron los TCR que se unían fuertemente al epítomo de tipo silvestre. Los TCR existentes deberían exhibir solo una afinidad de unión débil o nula al epítomo de tipo silvestre (A). Los epítomos que contienen una sustitución de aminoácidos que tiene una puntuación T alta tienen propiedades fisicoquímicas similares a las del aminoácido de tipo silvestre y, por lo tanto, probablemente tendrán poco impacto en la afinidad de unión a los TCR existentes (B). Los epítomos que contienen una sustitución de aminoácidos con una puntuación T tienen una mayor probabilidad de aumentar la afinidad de unión que los TCR exactos y, por lo tanto, una mayor probabilidad de ser inmunogénicos (C). En esta ilustración esquemática, la codificación de colores se usa para emparejar células T con un péptido coincidente. Las mutaciones naranja/amarillo representan mutaciones con puntuaciones T altas (similares a las WT), mientras que las mutaciones azul/púrpura representan mutaciones con puntuaciones T bajas (diferencia fisicoquímica significativa en comparación con la WT).

Figura 6. Distribución acumulativa de mutaciones inmunogénicas en función de M_{mut} . A. Comparación del porcentaje de mutaciones inmunogénicas que satisfacen la hipótesis de control inicial H_{BC1} : $\{M_{mut} \leq \beta\}$, con el porcentaje de mutaciones inmunogénicas que satisfacen la hipótesis parcial H_A : $H_{BC1} \cap \{T \leq \tau\}$, la hipótesis parcial H_{BC2} : $H_{BC1} \cap \{M_{mut} \leq \alpha\}$ y la hipótesis completa H_A : $H_{BC1} \cap \{M_{mut} \leq \alpha\} \cap \{T \leq \tau\}$, para $\alpha = 1$, $\tau = 1$. B. Comparación del porcentaje de mutaciones inmunogénicas que satisfacen las hipótesis de control iniciales H_{BC1} : $\{M_{mut} \leq \beta\}$ con el porcentaje de mutaciones inmunogénicas que satisfacen las hipótesis parciales inversas: $H_{BC1} \cap \{T > \tau\}$ y $H_{BC1} \cap \{M_{mut} > \alpha\}$. El análisis en A y B se basa en los conjuntos de datos agrupados de B16F10 y CT26.WT, que comprenden 132 mutaciones, de las cuales 30 eran inmunogénicas. Cada punto de datos en los gráficos se basa en ≥ 4 mutaciones.

Figura 7. Distribución acumulativa de mutaciones inmunogénicas en función de la puntuación de M_{mut} . Comparación del porcentaje de mutaciones inmunogénicas que satisfacen la hipótesis de control inicial H_{BC1} : $\{M_{mut} \leq \beta\}$, con el porcentaje de mutaciones inmunogénicas que satisfacen la hipótesis parcial H_A : $H_{BC1} \cap \{T \leq \tau\}$, la hipótesis parcial H_{BC2} : $H_{BC1} \cap \{M_{mut} \leq \alpha\}$ y la hipótesis completa H_A : $H_{BC1} \cap \{M_{mut} \leq \alpha\} \cap \{T \leq \tau\}$, dado $\alpha = 1$, $\tau = 1$ para B16 (A) y CT26 (C). Comparación del porcentaje de mutaciones inmunogénicas que satisfacen las hipótesis de control inicial H_{BC1} : $\{M_{mut} \leq \beta\}$ con el porcentaje de mutaciones inmunogénicas que satisfacen las hipótesis parciales inversas: $H_{BC1} \cap \{T > \tau\}$ y $H_{BC1} \cap \{M_{mut} > \alpha\}$ para B16 (B) y CT26 (D). Las Figuras A y B se basan en el análisis de las 50 mutaciones priorizadas de B16F10, de las cuales 12 eran inmunogénicas. Las Figuras C y D se basan en las 82 mutaciones priorizadas de B16F10, de las cuales 30 eran inmunogénicas. Cada punto de datos en los gráficos se basa en ≥ 4 mutaciones.

Figura 8. Control de la inmunogenicidad de WT. Para comprobar si la omisión de las soluciones MUT+/WT+ tuvo un impacto en estos hallazgos, se excluyó del conjunto de datos 9 mutaciones MUT+/WT+ y 2 mutaciones para las que no se ha medido el WT, dejando un total de 121 mutaciones (43 B16 y 78 CT26) de las cuales 19 fueron MUT+/WT- (5 para B16 y 14 para CT26). Nuevamente se encontraron las mismas tendencias que en el conjunto de datos

completo, es decir, respuesta altamente no lineal en función de la puntuación de M_{mut} , superioridad de la hipótesis de H_A sobre la hipótesis parcial e inferioridad de las hipótesis invertidas en comparación con el control de referencia H_{BC1} . A. Distribución acumulada de la inmunogenicidad en función de la puntuación de M_{mut} . B. Histograma del porcentaje de mutaciones inmunogénicas por compartimento de M_{mut} . C. Comparación del porcentaje de mutaciones inmunogénicas que satisfacen las hipótesis de control iniciales H_{BC1} con H_A , H_{BC2} y H_A . D. Comparación del porcentaje de mutaciones inmunogénicas que satisfacen la hipótesis de control inicial H_{BC1} con la hipótesis inversa. Véase la leyenda de la Figura 5 para obtener detalles adicionales.

Figura 9. Fracción de mutaciones inmunogénicas en función de RPKM. Rojo: todas las 50 mutaciones B16 y las 82 mutaciones CT26 sin filtro (132 mutaciones en total). Azul: mutaciones que superan la hipótesis de H_A con $\alpha = 1$, $\beta = 0,5$, $\tau = 1$. B. Porcentaje de mutaciones inmunogénicas para diferentes intervalos de RPKM sin filtro. Los contenedores de RPKM son: 1 = (0, 1), 2 = (1, 5), 3 = (5, 50), 4 = (50, ∞). C. Porcentaje de mutaciones inmunogénicas para diferentes intervalos de RPKM bajo la hipótesis de H_A con $\alpha = 1$, $\beta = 0,5$, $\tau = 1$. Los contenedores de RPKM son: 1 = (0, 1), 2 = (1, ∞). Los errores son SE.

Figura 10. Epítomos inmunogénicos de clase II mutados en posición de anclaje y no anclaje. Los motivos de la posición de anclaje se analizaron usando SYFPEITHI.

Figura 11. Modelos propuestos para epítomos inmunogénicos asociados a tumores.

Figura 12. Ejemplo de un método para ponderar la posición del rango de mutaciones. Para cada mutación, la posición del rango en la lista de mutaciones clasificadas se puede ponderar aún más por el número de soluciones para las cuales la combinación de tipos de HLA para el paciente, las posibles longitudes de ventana para el tipo de HLA y la posición de la mutación dentro del epítomo dieron como resultado una solución con M_{mut} bajo o resultó en una clasificación H_A y/o H_B U H_C . Dado que todas las soluciones por mutación pueden presentarse potencialmente en paralelo, este factor de ponderación puede ser un contribuyente importante a la posición de rango de la mutación.

Figura 13. Ejemplo de gráfico de dispersión de todas las soluciones de epítomos para la mutación chr14_52837882 de CT26 contra M_{mut} y $\Delta M = M_{mut} - M_{wt}$.

Ejemplos

Las técnicas y métodos usados en este documento se describen en este documento o se llevan a cabo de una manera ya conocida y como se describe, por ejemplo, en Sambrook et al., Molecular Cloning: A Laboratory Manual, 2ª edición (1989) Cold Spring Harbor Laboratory. Press, Cold Spring Harbor, Nueva York. Todos los métodos, incluido el uso de kits y reactivos, se llevan a cabo de acuerdo con la información del fabricante, a menos que se indique específicamente lo contrario.

Ejemplo 1: Establecimiento de un modelo para predecir la inmunogenicidad de los epítomos de células T

Previamente se exploró la inmunogenicidad de 50 mutaciones somáticas identificadas en la línea celular de melanoma murino B16F10 (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). Estas 50 mutaciones se seleccionaron de un conjunto de 563 mutaciones somáticas no sinónimas expresadas principalmente para maximizar la expresión del MHC de clase I (J.C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)) (véase también el Ejemplo 2). Para cada mutación, se predijo el epítomo mínimo, es decir, el epítomo con la puntuación de consenso del MHC de clase I más baja (Y. Kim et al., Nucleic Acids Research 40, W525 (2012)) (definido aquí como M_{mut}) al buscar en el espacio de todos los posibles alelos del MHC de clase I, posibles longitudes de epítomo y ventanas de secuencia (en la que colocar la mutación) (J.C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)). La medición de la inmunogenicidad de estas mutaciones usando vacunación de ARN seguida de lectura de péptidos (véase el Ejemplo 2) confirmó los hallazgos anteriores usando vacunación de péptidos (J.C. Castle et al., Exploiting the mutanome for tumor vaccination. Cancer Research 72, 1081 (2012)), y mostró que solo 12 de 50 mutaciones (24%) eran inmunogénicas (Tabla 1), con secuencias MUT+/WT- que comprenden solo el 10% de todas las mutaciones probadas.

Tabla 1. Número de mutaciones inmunogénicas después de la vacunación con ARN de las cepas murinas B19F10 y CT26.WT

Cepa	Inmunogenicidad de MUT/WT				Total	% inmunogénico MUT+/total
	+/+	+/-	+/-	-/-		
B16F10	7	5	0	38	50	24%
CT26	2*	14*	1	63	80	20%
Total	9	19	1	101	130	21,50%

*Se excluyeron de esta tabla dos mutaciones CT26 MUT+ porque aún no se ha medido su reactividad de WT. En total, hubo 18 mutaciones MUT+ de las 82 mutaciones CT26 medidas hasta el momento, lo que resultó en una tasa de éxito del 22%.

Los resultados del caso de prueba murino B16F10 demuestran que la selección sin alteración de mutaciones no sinónimas expresadas con puntuaciones de M_{mut} bajas ($\leq 3,9$) produce tasas de éxito bastante bajas para predecir la inmunogenicidad. Por lo tanto, se requiere una mejor comprensión de los mecanismos que impulsan la inmunogenicidad para que las vacunas personalizadas dirigidas a neoantígenos específicos de tumores se conviertan en terapias efectivas. En un esfuerzo por descubrir variables adicionales que contribuyan a la inmunogenicidad, se exploró la inmunogenicidad de mutaciones somáticas no sinónimas expresadas identificadas en una línea celular colorrectal murina CT26.WT. En total, se seleccionaron 96 mutaciones en función de sus puntuaciones de M_{mut} (baja frente a alta), RPKM media (baja frente a alta) y localización celular (intracelular frente a extracelular), y se evaluó la inmunogenicidad mediante vacunación con ARN con ambos lectura de péptidos y ARN (véase el Ejemplo 2 para obtener más detalles). Junto con la línea celular B16F10, este conjunto de datos consta de 132 epítomos, cuya inmunogenicidad se midió *ex vivo* en esplenocitos murinos.

La puntuación de consenso del MHC. Para investigar la dependencia de la inmunogenicidad en M_{mut} , se graficó el porcentaje acumulado de mutaciones inmunogénicas en función de M_{mut} , es decir, el porcentaje de mutaciones con una puntuación de M_{mut} menor que un umbral dado (denotado por β) que eran inmunogénicas. Un análisis de los conjuntos de datos combinados de B16 y CT26 que abarcan un total de 132 mutaciones revela una dependencia altamente no lineal de la tasa de éxito de la inmunogenicidad en M_{mut} (Fig. 2A). La Fig. 2A muestra que las mutaciones inmunogénicas se enriquecen con puntuaciones de M_{mut} extremadamente bajas ($\leq \sim 0,2$). Para $M_{mut} \leq 0,1$, el porcentaje de mutaciones inmunogénicas alcanza un máximo de $\sim 60\%$ y decae rápidamente a medida que aumenta M_{mut} , cayendo por debajo de $\sim 25\%$ para $M_{mut} \geq 2$. El porcentaje de mutaciones inmunogénicas con $M_{mut} \leq 0,3$ frente a $> 0,3$ fue del 44,4% en comparación con el 17,1%, una diferencia estadísticamente significativa (valor $p = 0,004$, prueba exacta de Fisher, de una cola). Un histograma del porcentaje de mutaciones inmunogénicas para tres contenedores de M_{mut} : $\leq 0,3$, $(0,3; 1)$ y > 1 muestra que el porcentaje de mutaciones inmunogénicas disminuye a medida que aumenta M_{mut} (Fig. 2B). Las diferencias entre la tasa de éxito de los compartimientos más bajos ($M_{mut} \leq 0,3$), 44,4%, y tanto el compartimiento central, 20,7%, como el compartimiento más alto ($M_{mut} > 1$), 15,8%, en la Fig.2B fueron estadísticamente significativos (valores $P = 0,05$ y $0,004$, respectivamente, prueba exacta de Fisher, una cola), lo que indica que para $M_{mut} > \sim 0,3$ la tasa de éxito desciende de manera estadísticamente significativa. También se observa una tendencia similar en la tasa de éxito al analizar los mutanomas B16 y CT26 por separado (Fig. 3).

Hasta ahora, los criterios para seleccionar mutaciones se centraron en la presentación, y se ha observado que restringir la puntuación de unión a MHC del epítomo mutado permite la predicción de epítomos inmunogénicos con hasta un 60% de precisión. La presentación, sin embargo, es una condición necesaria pero no suficiente para inducir inmunogenicidad. Al identificar criterios adicionales para el reconocimiento de TCR, se puede mejorar aún más la precisión de esta predicción. Se plantea la hipótesis de dos mecanismos mutuamente excluyentes para impulsar la inmunogenicidad, a los que se hace referencia como modelos de inmunogenicidad de clase I y clase II.

Inmunogenicidad de clase I. Para que el repertorio de TCR reconozca un epítomo mutado y genere una respuesta inmune, se plantea la hipótesis de que deben cumplirse tres condiciones (HA Fig.4): (i) el epítomo de tipo silvestre, en algún momento durante el desarrollo del organismo, fue presentado al sistema inmunológico que conduce a la eliminación de los TCR coincidentes a través de una fuerte unión de TCR/pMHC, (ii) se presenta el epítomo mutado, y (iii) las propiedades fisicoquímicas del aminoácido mutado son suficientemente "diferentes" de las del aminoácido de tipo silvestre (por alguna métrica que se define a continuación) para que el repertorio de TCR sea capaz de "detectar" o "registrar" esta sustitución. Las condiciones (i) y (ii) aseguran que el sistema inmunológico esté realmente expuesto al cambio, es decir, la mutación. La condición (iii) requiere que la mutación cambie significativamente el carácter fisicoquímico del aminoácido de tipo silvestre de modo que la afinidad de unión del epítomo mutado al TCR existente (no eliminado) aumente potencialmente, activando de esta manera la cascada de señalización que conduce a una respuesta inmune (Fig. 5).

La puntuación de reconocimiento de TCR. Los modelos de inmunogenicidad de clase I requieren una métrica para estimar la diferencia fisicoquímica entre dos aminoácidos. Es bien sabido en la evolución molecular que es probable que los aminoácidos que se intercambian con frecuencia tengan similitudes químicas y físicas, mientras que los aminoácidos que se intercambian rara vez tienen diferentes propiedades fisicoquímicas. La probabilidad de que ocurra una sustitución dada en la naturaleza en comparación con la probabilidad de que esta sustitución ocurra por casualidad se mide mediante matrices logarítmicas de probabilidades. Los patrones observados en matrices logarítmicas de probabilidades impuestas por la selección natural "reflejan la similitud de las funciones de los residuos de aminoácidos en sus interacciones débiles entre sí en la conformación tridimensional de las proteínas" (M.O. Dayhoff, RM Schwartz, B.C. Orcutt, A model for evolutionary change. M.O. Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978)). Por lo tanto, se utilizaron matrices logarítmicas de probabilidades basadas en la evolución, a las que se hace referencia en el presente documento como "puntuaciones T" para reflejar el reconocimiento de TCR, como matrices de puntuación eficaces para sustituciones de aminoácidos asociados al cáncer. Es probable que se produzcan sustituciones con puntuaciones T positivas (es decir, logarítmicas de probabilidades) en la naturaleza y, por lo tanto, correspondan a dos aminoácidos que tienen propiedades fisicoquímicas similares. El modelo de clase I predice que las sustituciones con puntuaciones T positivas tendrían una menor probabilidad de ser inmunogénicas. Por el contrario, las sustituciones con puntuaciones T negativas reflejan sustituciones que es poco probable que ocurran en la naturaleza y, por lo tanto, corresponden a dos aminoácidos que tienen propiedades fisicoquímicas significativamente diferentes. De acuerdo con este modelo, tales sustituciones tendrían una mayor probabilidad de ser inmunogénicas.

Se compararon diferentes métodos de estimación de matrices logarítmicas de probabilidades y se encontró que los resultados son en gran medida robustos para el método exacto elegido. El enfoque de estimación basado en la máxima probabilidad (ML) conocido como WAG (S. Whelan, N. Goldman, Molecular Biology and Evolution 18, 691 (2001)), utilizando una distancia PAM (mutación puntual aceptada) de 250, pareció separar mejor las mutaciones inmunogénicas predichas de las no inmunogénicas y, por lo tanto, se presentan los resultados con esta matriz (véase el Ejemplo 2 para obtener más detalles).

Immunogenicidad de clase II. En el modelo de inmunogenicidad de clase II, se plantea la hipótesis de que es probable que una mutación sea inmunogénica si el sistema inmunológico nunca antes ha visto el epítipo de tipo silvestre y, por lo tanto, es desafiado por el epítipo mutado. Por lo tanto, para que una mutación sea inmunogénica en este modelo, se plantea la hipótesis de que deben cumplirse dos condiciones: (i) el epítipo de tipo silvestre nunca se presentó al sistema inmunológico, (ii) se presenta el péptido mutado. Estas condiciones pueden satisfacerse conjuntamente si, por ejemplo, la mutación golpea una posición de anclaje, cambiando así un epítipo "no ligante" en un "ligante". Formalmente, la inmunogenicidad de clase II se puede separar en dos subhipótesis: puntuaciones T altas (H_B en la Figura 4) y puntuaciones T bajas (H_C en la Figura 4). Sin embargo, dado que se supone que no se presenta el epítipo de tipo silvestre, no se espera que la naturaleza de la sustitución de aminoácidos tenga un impacto en el reconocimiento de TCR y, por lo tanto, se equiparára la inmunogenicidad de clase II con la hipótesis unificada: $H_B \cup H_C$.

Ensayo de inmunogenicidad de clase I. Los supuestos de inmunogenicidad de clase I (H_A en la Fig.4) se pueden reformular matemáticamente de la siguiente manera: se requiere que se presente el epítipo de tipo silvestre ($M_{wt} \leq \alpha$), se presente el epítipo mutado ($M_{mut} \leq \beta$) y la sustitución del aminoácido no sea trivial ($T \leq \tau$), en el que M_{wt} se define como la puntuación de consenso del MHC del epítipo mutado (mismo alelo de HLA y longitud de ventana) reemplazando el aminoácido mutado con el aminoácido de tipo silvestre, y T denota la puntuación de T. Dado que las tres condiciones son necesarias, se espera que la precisión del clasificador de H_A sea mayor en comparación con un clasificador basado solo en M_{mut} (H_{BC1} en la Figura 4) o en comparación con las hipótesis parciales: $H_{BC1} \cap \{M_{wt} \leq \alpha\}$ y $H_{BC1} \cap \{T \leq \tau\}$. Por lo tanto, se calcula el porcentaje de mutaciones inmunogénicas (número de verdaderos positivos dividido por la suma de verdaderos positivos y falsos positivos) en función de β para H_{BC1} , para la hipótesis parcial H_A : $H_{BC1} \cap \{T \leq \tau\}$ y para la hipótesis parcial H_{BC2} : $H_{BC1} \cap \{M_{wt} \leq \alpha\}$. Se encontró que un umbral conservador para τ en el intervalo de $\approx 0,5$ a 1 se desempeñó mejor (el intervalo de la matriz WAG250 es de -5,1 (sustitución de F \leftrightarrow G) a +5,4 (sustitución F \leftrightarrow Y)). También se encontró que α se puede restringir de forma conservadora en comparación con β , estableciendo $\alpha \approx 1$. La Figura 6A muestra de hecho que, al considerar el mutanoma agrupado de B16 y CT26, los clasificadores basados en H_{BC2} y H_A alcanzaron una mayor precisión que la hipótesis de control inicial H_{BC1} . Además un clasificador con base en la hipótesis completa H_A alcanzó mayor precisión que las hipótesis parciales H_{BC1} y H_{BC2} , demostrando así un efecto aditivo. Las mismas conclusiones se mantienen al analizar los conjuntos de datos B16 y CT26 por separado (Fig. 7).

Dado que se postula que las condiciones $M_{wt} \leq \alpha$ y $T \leq \tau$ son condiciones necesarias para la inmunogenicidad, se esperaría que un clasificador con base en la condición $H_{BC1} \cap \{T > \tau\}$ o la condición $H_{BC1} \cap \{M_{wt} > \alpha\}$ (es decir, negando la condición secundaria) funcionaría peor que H_{BC1} . De hecho, se encontró que este es el caso de B16 y CT26 cuando se analizan juntos (Fig. 6B) o por separado (Fig. 7). Por lo tanto, se llegó a la conclusión de que los conjuntos de datos de B16 y CT26 apoyan tanto juntos como por separado la hipótesis de H_A . La omisión de mutaciones en las que el ARN WT también mostró reactividad no afectó estas conclusiones (Fig. 8).

Control de la hipótesis de H_A . Aunque las mutaciones con puntuaciones T altas pueden seguir siendo inmunogénicas, una hipótesis que enriquezca tales mutaciones debería enriquecer estadísticamente mutaciones no inmunogénicas. Por lo tanto, si se compara la hipótesis de H_A ($H_{BC2} \cap \{T \leq \tau\}$) con su inversa, ($H_{BC2} \cap \{T > \tau\}$) (H_n en la Fig. 4), se debería observar un agotamiento estadísticamente significativo de mutaciones inmunogénicas. De hecho, la Tabla 2 muestra que para $M_{mut} \leq \beta = 0,5$, $M_{wt} \leq \alpha = 1$ y $T \leq \tau = 1$, H_A supera a H_n , con una tasa de éxito del 52,5% ($n = 21$) en comparación con 21,4% ($n = 14$; $P = 0,068$, prueba exacta de Fisher de una cola).

Tabla 2. Porcentaje de mutaciones inmunogénicas bajo diversas hipótesis con base en los conjuntos de datos agrupados de B16 y CT26 que comprenden 133 mutaciones

		Parámetros de la hipótesis			
		Umbral (β) de M_{mut}	Umbral (α) de M_{wt}	Umbral (τ) de Puntuación de T	% de mutaciones inmunogénicas
Hipótesis	H_{BC1}	≤ 2			22/83 (26,5%)
		≤ 1			18/56 (32,1%)
		$\leq 0,1$			6/10 (60%)
	H_A	$\leq 0,1$	≤ 1	≤ 1	6/9 (66,7%)
	H_{BC2}			≤ 1	6/9 (66,7%)
	H_A			≤ 1	6/8 (75%)
H_A	$\leq 0,2$		≤ 1	8/14 (57,1%)	

	H _{BC2}				9/17 (52,9%)
	H _n		≤ 1	> 1	1/6 (16,7%)
	H _A			≤ 1	8/12 (66,7%)
	H _{BC2}				14/34 (41,2%)
	H _n	≤ 0,5	≤ 1	> 1	3/14 (21,4%)
	H _A			≤ 1	11/21 (52,4%)
	H _B U _H C	≤ 0,5	> 1		1/6 (16,7%)

H_A también se comporta mejor que H_{BC2} de control inicial, que alcanza el 41,2% (n = 35). A medida que disminuye β, la diferencia entre las tasas de éxito de H_A y H_n aumenta, ya que cuanto más estricta es la condición en β, más falsos positivos se eliminan del grupo H_A. Por ejemplo, para β = 0,25, la tasa de éxito del grupo H_A fue del 67% (n = 14) en comparación con una tasa de éxito del 17% (n = 6) para el grupo H_n (P = 0,066, prueba exacta de Fisher de una cola), véase la Tabla 3.

Tabla 3. Lista clasificada de 133 mutaciones B16F10/CT26.WT medidas que satisfacen la hipótesis de control básica H_{BC1} (M_{mut} ≤ 0,25) desglosada en las tres clases de hipótesis desarticuladas: hipótesis de H_A para mutaciones inmunogénicas, (M_{wt} ≤ 0,8, T ≤ 0,5), hipótesis de H_n/H_A inversa enriquecedora para mutaciones no inmunogénicas (M_{wt} ≤ 0,8, T > 0,5) e hipótesis H_BU_HC para mutaciones inmunogénicas (M_{wt} > 0,8). Se propone que los candidatos de H_A Y H_BU_HC se clasifiquen en función de la importancia relativa de variables distintivas. Para H_A, el orden propuesto es: M_{mut} (descendente) → Puntuación T (descendente) → M_{WT} descendente. Para H_BU_HC, el orden propuesto es M_{mut} (descendente) → M_{WT} (ascendente). Los errores son errores estándar

Muestra	Mut	Respuesta	Respuesta de ARN (MUT)	Respuesta de ARN (WT)	Símbolo (Ingenuidad)	Alelo de MHC I	Epitopo (MUT)	Epitopo (WT)	Expresión media	M _{mut}	M _{wt}	Puntuación T (WAG250)
---------	-----	-----------	------------------------	-----------------------	----------------------	----------------	---------------	--------------	-----------------	------------------	-----------------	-----------------------

Inmunogenicidad Clase I (H_A): 67 ± 14% de tasa de éxito

B16	33	CD8	sí	no	PBK	H-2-Db	AAVILRDALHM	AAVILRVALHM	19,6	0,1	0,1	-2,7
CT26	14		no	no	Nphp3	H-2-Dd	GGPGSEKSL	GGPGSGKSL	0,6	0,1	0,1	-0,5
B16	37		no	no	DPF2	H-2-Db	LALPNNYCDV	LALPNNYCDF	20,0	0,1	0,1	-0,1
B16	25	CD4	sí	sí	PLOD2	H-2-Db	SHLNNDVWQI	SHLNNDVWQI	21,7	0,1	0,1	-0,1
CT26	37	CD4	sí	no	Dhx35	H-2-Kd	YYMRDVIAI	YYMRDVTAI	5,5	0,1	0,1	-0,1
CT26	26	CD8	sí	no	E2f8	H-2-Kd	TYLQPAQAQM	IYLQPAQAQM	29,5	0,1	0,1	-0,1
B16	21	CD4	sí	no	ATP11A	H-2-Db	QSLGFTYL	QRLGFTYL	42,7	0,1	0,8	-0,1
CT26	13		sí	no	Nphp3	H-2-Kd	EYWASRALDS	EYWASRALGS	0,5	0,1	0,1	0,1
CT26	16		no	no	H2-Q8	H-2-Kd	GYLQFAYEGC	GYLQFAYEGR	5,1	0,2	0,2	-1,7
B16	46	CD4	sí	sí	ACTN4	H-2-Kb	VTFQAFIDVMS	VTFQAFIDFMS	125,6	0,2	0,1	-0,1
CT26	15		no	no	Slc41a2	H-2-Kd	PYLTALDILL	PYLTALGDLL	2,8	0,2	0,2	0,1
CT26	27		sí		Agxt2l2	H-2-Dd	AGGLFVADAI	AGGLFVADEI	29,7	0,2	0,3	0,1

Inmunogenicidad Clase II (H_BU_HC): 0% de tasa de éxito

CT26	25		no	no	Pcdhga8	H-2-Dd	VGINFLQSYQ	VGINSLQSYQ	26,2	0,1	2,7	-2,1
CT26	1		no	no	Gm8909	H-2-Dd	TRPARDGTF	TRPAGDGTF	11,0	0,2	1,6	-1,0
CT26	40		no	no	Zfp449	H-2-Ld	EPQIAMDDM	EPQIDMDDM	0,7	0,25	1,25	-0,2

H_n: 17 ± 15% de tasa de éxito

B16	49		no	no	FAT1	H-2-Db	IAIQNTTQL	IAIQNTTQL	18,8	0,1	0,1	2,6
B16	36		no	no	TM9SF3	H-2-Kb	AIYHHSRAI	AIYHHSRAI	51,1	0,2	0,3	2,7
CT26	39	CD8	sí	no	ALS2	H-2-Kd	SYIALVDKNI	SYLALVDKNI	4,7	0,2	0,1	2,7
CT26	2		no	no	SNAP47	H-2-Dd	VIPILEMQF	VIPILEVQF	15,5	0,2	0,2	1,8
CT26	17		no	no	H2-Q8	H-2-Kd	GYLQFAYDGR	GYLQFAYEGR	5,1	0,2	0,2	2,9
CT26	38		no		Cspp1	H-2-Kd	VYLNLLKFT	VYLNLFKFT	0,7	0,2	0,1	2,0

En el Ejemplo 3 se da un ejemplo de factores de ponderación adicionales que pueden mejorar aún más la clasificación de inmunogenicidad.

De manera más general, la lista de mutaciones que satisfacen H_{BC1} (M_{mut} ≤ β) se puede clasificar en las tres categorías: H_A, H_n y H_BU_HC (Tabla 3), en las que H_A enriquece las mutaciones inmunogénicas, H_n enriquece las mutaciones no inmunogénicas. En el caso de B16 y CT26, los tres candidatos en el grupo H_BU_HC no eran inmunogénicos, contrariamente a nuestras expectativas. Sin embargo, si se elige un umbral α* más realista para M_{wt} tal que α* >> α, entonces no habría predicciones que pudieran probarse para H_BU_HC.

Precisión máxima de clasificadores de inmunogenicidad. De acuerdo con la Tabla 1, la tasa de éxito promedio para la inmunogenicidad de predicción en los conjuntos de datos combinados de B16 y CT26 fue del 22,7% (= 30/132). Al

aplicar el umbral más estricto en la puntuación de M_{mut} ($\beta = 0,1$), la precisión de un clasificador de inmunogenicidad aumenta al 60% (= 6/10; H_{BC1} en la Tabla 2). Combinando H_{BC1} ya sea con la condición $M_{wt} \leq \alpha$ o la condición $T \leq \tau$ ($\alpha = 1, \tau = 1$), la precisión aumenta al 66,7% (= 6/9). El clasificador basado en H_A , que combina ambos criterios, da como resultado una respuesta aditiva, que aumenta la precisión al 75% (= 6/8) (Tabla 2).

Epítipo B16 MUT33. El epítipo de clase H_A que fue clasificado como el más alto por todos los modelos evolutivos (excepto la matriz PAM) en el conjunto de datos B16/CT26 agrupado fue MUT33 de B16 (véase Tabla 3). Un análisis adicional reveló que MUT33 de hecho invocaba una respuesta CD8+ restringida al MHC de clase I y exhibía inmunogenicidad *ex vivo* contra el epítipo mínimo predicho (datos no mostrados).

Papel de la expresión génica. El gráfico de la fracción de mutaciones inmunogénicas (número de mutaciones inmunogénicas con valores de RPKM por debajo de un umbral determinado sobre el número total de mutaciones inmunogénicas) en función de los valores de RPKM para B16 y CT26 indica que esta relación se estanca algo en valores de RPKM muy bajos (Figura 9A). Este efecto se observa tanto si se aplica el criterio de H_A como si no. Graficar el porcentaje de mutaciones inmunogénicas para diferentes contenedores de RPKM (Fig. 9B y C) sugiere que los valores de RPKM $\leq \sim 1$ tienen una tasa de éxito algo menor (con o sin la aplicación de la hipótesis de filtrado de H_A), aunque sugestivo, debe tenerse en cuenta que estos resultados están dentro del intervalo de error.

Estudio de epítipos CD8+ publicados. A continuación, nos interesó observar si los antígenos tumorales definidos por células T publicados con sustituciones de un solo aminoácido que provocan una respuesta restringida de CD8+ cumplieran con los modelos de inmunogenicidad. De los 17 epítipos que se publicaron (P. Van der Bruggen, V. Stroobant, N. Vigneron, B. Van den Eynde. (Cancer Immun. <http://www.cancerimmunity.org/peptide/>, 2013)) (Tabla 4), cinco cumplieron los criterios de H_A ($\alpha = 0,7, \beta = 0,2, \tau = 0,5$), cuatro cumplieron los criterios para H_CUH_B ($\alpha = 2,2, \beta = 0,4$) y dos cumplieron el criterio de H_n ($\alpha = 0,6, \beta = 0,3, \tau = 1,7$).

Tabla 4. Epítipos publicados con sustitución de un solo aminoácido generando respuestas CD8+. Véase el Ejemplo 2 para el listado de referencias. Las mutaciones de la posición de anclaje en el grupo de H_BUH_C están resaltadas en rojo

Hipótesis	Umbral de M_{mut}	Umbral de M_{wt}	Umbral de puntuación T	M_{mut}	M_{wt}	Puntuación T^*	Gen	Epítipo MUT	Epítipo MW	Fuente
H_A	$\leq 0,2$	$\leq 0,7$	$\leq 0,5$	0,10	0,10	-2,1	SIRT2	KIFSEVTLK	KIFSEVTPK	1
				0,10	0,10	-0,1	SNRPD	SHETVHIEL	SHEVTVIEL	1
				0,14	0,22	0,5	ME1	FLDEFMEGV	FLDEFMEAV	2
				0,20	0,70	-1,0	RBAF600	RPHVPESAF	GPHVPESAF	1
H_n	$\leq 0,3$	$\leq 0,6$	$\geq 1,7$	0,15	0,20	-2,1	PRDX5	LLDDLLVSI	LLDDSLVSI	3
				0,25	0,20	1,7	N-ras	ILDTAGREEY	ILDTAGQEEY	4
				0,30	0,60	1,9	EEF2	ETVSEQSNV	ETVSEESNV	10
H_B	$\leq 0,25$	$\geq 2,25$	$> 0,5$	0,15	10,20	1,2	EFTUD2	KILDVAVAQK	KILDVAVAQE	1
				0,25	2,25	1,2	MYO1B	KINKNPKYK	EINKNPKYK	5
				0,20	10,95	-2,3	FLT3	YVDFREYEEY	YVDFREYED	6
H_C	$< 0,4$	> 10	$\leq 0,5$	0,38	10,58	-2,1	CTNNB1	SYLDSGIHF	SYLDSGIHS	7
				0,65	2,70	-2,1	OS-9	KELEGILL	KELEGILLP	8
				1,20	1,00	0,1	GPNMB	TLDWLLQTPK	TLGWLLQTPK	1
Otro	$\geq 0,65$	≥ 1		2,68	3,58	1,0	ACTN4	FIASNGVKLV	FIASKGVKLV	9
				4,20	4,60	-1,2	P53	VVPCEPPEV	VVPYEPPEV	11
				26,40	29,75	-1,7	CDK4	ACDPHSGHFV	ARDPHSGHFV	12
				33,30	29,80	-2,9	K-ras	VVGAVGVG	VVGAGGVG	13

* con base en la Matrix del logaritmo de probabilidades de WAG250, leyenda en color: $T \leq 0,5$; $0,5 < T \leq 1$; $T > 1$

Por lo tanto, las hipótesis de H_A y H_CUH_B juntas representaron aproximadamente el 50% de los epítipos publicados. Curiosamente, 3 de los 4 epítipos publicados que cumplían la condición de H_CUH_B (recuadros rojos en la Tabla 4) tenían una puntuación de M_{wt} mayor que 10 debido a mutaciones en la posición de anclaje (Fig. 10). Dado que el requisito para la hipótesis de H_CUH_B es que la probabilidad de que cualquier célula presente el epítipo de tipo silvestre durante el desarrollo del organismo se mantenga insignificamente pequeña, se espera que el umbral de M_{wt} se mantenga alto, es decir, $\alpha^* \gg \alpha$. De hecho, al aumentar α de 0,8 a > 3 , los falsos positivos para B16/CT26 en la Tabla 3 desaparecen. Por lo tanto, un umbral más realista para M_{wt} bajo la hipótesis de H_CUH_B puede estar entre 3 y 10.

La línea celular MZ7-MEL. Para probar la capacidad estos modelos de inmunogenicidad para predecir epítipos inmunogénicos en un modelo de tumor humano, se exploró la línea celular MZ7-MEL, establecida en 1988 a partir de una metástasis esplénica de un paciente con melanoma maligno (V. Lennerz et al., Proceedings of the National Academy of Sciences of the United States of America 102, 16013 (2005)). El cribado de una biblioteca de ADNc de células MZ7-MEL con células T autólogas reactivas a tumores reveló al menos cinco neoantígenos capaces de generar respuestas CD8+ (V. Lennerz et al., Proceedings of the National Academy of Sciences of the United States of America 102, 16013 (2005)). Este constituye el mayor conjunto de neoantígenos CD8+ derivados de un paciente hasta la fecha. Aplicando estos modelos de inmunogenicidad a estos epítipos, se encontró que tres neoantígenos se clasificaron como epítipos H_A , y un neoantígeno, una mutación de posición de anclaje, se clasificó como un epítipo H_BUH_C (flechas en la Tabla 4 y Fig. 10). Por lo tanto, cuatro de los cinco epítipos podrían explicarse por nuestros modelos de inmunogenicidad.

Para probar nuestra capacidad para predecir estos epítipos de novo en la línea celular MZ7-MEL, se secuenció el exoma de la línea celular MZ7-MEL (véase los Métodos). En total, se identificaron 743 mutaciones no sinónimas expresadas. Las cinco mutaciones identificadas previamente por Lennerz et al., (V. Lennerz et al., Proceedings of the

National Academy of Sciences of the United States of America 102, 16013 (2005)). Luego se calculó para cada mutación la puntuación T, M_{mut} y M_{wt} , informando también el alelo HLA y el epítipo que resultó en la puntuación de consenso mínima del MHC para la mutación dada. Las mutaciones se clasificaron en uno de tres grupos: H_A , $H_B UH_C$ y H_n usando los umbrales $\alpha = 0,8$, $\beta = 0,2$, $\tau = 0,5$ (más la condición $RPKM > 0,2$), y luego se clasificaron con base en su potencial para ser inmunogénicos, como se explica en la Tabla 3. Se encontró que de 743 mutaciones, 32 mutaciones cumplieron los criterios de H_A (Tabla 5), 12 cumplieron el criterio $H_B UH_C$ (Tabla 6) y 15 cumplieron el criterio de H_n .

5
10
Tabla 5. Mutaciones de células MZ-MEL clasificadas por H_A . 32 de las 743 mutaciones no sinónimas expresadas en MZ7-MEL se clasificaron como inmunogénicas para H_A usando los umbrales: $\alpha = 0,8$; $\beta = 0,2$; $\tau = 0,5$. La clasificación se basa en un esquema de clasificación de M_{mut} (descendente) \rightarrow puntuación T (descendente). Los neoantígenos inmunogénicos identificados por Lennerz et al., están resaltados en amarillo. Además se requería que RPKM excediera 0,2

Clasificación *	Gen	M_{mut}	Puntuación T**	M_{wt}	Media exp.	
1	DPH2	0,1	-2,9	0,1	10,7	
	ADHFE1	0,1	-2,9	0,1	2,7	
2	DDX41	0,1	-2,1	0,1	24,8	← MZ7-MEL
	SIRT2	0,1	-2,1	0,1	15,7	
3	PRIC285	0,1	-1,0	0,4	5,4	
4	CSTF3	0,1	-0,5	0,1	11,2	
	ETFDH	0,1	-0,5	0,1	10,8	
5	MED12	0,1	-0,1	0,1	21,9	← MZ7-MEL
	SNRPD1	0,1	-0,1	0,1	18,0	
	MLLT6	0,1	-0,1	0,2	14,7	
	AFAP1	0,1	-0,1	0,2	5,1	
6	MAP3K11	0,1	0,1	0,1	41,9	
7	DHX30	0,1	0,3	0,1	34,4	
	ALK	0,1	0,3	0,1	0,4	
	CHMP4B	0,1	0,3	0,7	52,8	
	HADHB	0,1	0,5	0,1	60,6	
8	SUPT6H	0,1	0,5	0,1	24,4	
	C12orf35	0,1	0,5	0,1	3,1	
	ZDHHC5	0,1	0,5	0,4	27,6	
9	WIPF1	0,15	-2,1	0,15	37,2	
	ZNF740	0,15	-2,1	0,5	9,7	
10	MLL	0,15	0,5	0,3	3,7	
11	KIAA1715	0,2	-2,1	0,1	6,3	
	CHD8	0,2	-2,1	0,2	10,3	
12	DDX28	0,2	-1,7	0,2	7,2	
13	MAPK11P1L	0,2	-1,0	0,2	12,1	← MZ7-MEL
	UBR4 (RBAF600)	0,2	-1,0	0,7	11,8	
14	TRAK2	0,2	-0,1	0,5	21,8	
15	MAEA	0,2	0,0	0,25	25,3	
16	KLHL13	0,2	0,1	0,2	3,4	
17	FOSL2	0,2	0,3	0,2	9,4	
18	UTRN	0,2	0,5	0,15	6,3	

15 *La clasificación se basa en M_{mut} y la puntuación T

** La puntuación T se basa en la Matrix logarítmica de probabilidades WAG250

20
Tabla 6. Mutaciones de células MZ7-MEL clasificadas como $H_B UH_C$. 12 de las 743 mutaciones no sinónimas expresadas en MZ7-MEL se clasificaron como inmunogénicas para $H_B UH_C$ usando los umbrales: $\alpha^* = 0,8$; $\beta = 0,2$; $RPKM > 2$. La clasificación se basa en un esquema de clasificación de M_{mut} (descendente) \rightarrow M_{wt} (ascendente). Los neoantígenos inmunogénicos identificados por Lennerz et al., están resaltados en amarillo.

Clasificación*	Gen	M_{mut}	M_{wt}	Media exp.	
1	NF1	0,1	10,4	6,8	
2	MESP2	0,1	1,3	0,3	
3	EFTUD2 (SNRP116)	0,15	10,2	22,0	← MZ7-MEL
4	SEC31A	0,15	2,55	33,3	
5	ZNF335	0,2	18,35	8,3	
6	CPEB1	0,2	4,2	6,0	
7	UBAC2	0,2	2,8	7,5	

8	FANCF	0,2	2,4	3,9
	RAPH1	0,2	2,4	3,0
9	ZNF557	0,2	1,4	1,5
10	TLK2	0,2	1,35	2,9
11	ST7	0,2	1,15	10,9

* La clasificación se basa en M_{mut} y M_{wt}

De las 32 mutaciones clasificadas como HA, las tres mutaciones de clase H_A identificadas por Lennerz et al. (SIRT2, SNRPD1 y RBAF600) se clasificaron en las posiciones 2^a, 4^a y 13^{ava} de 18 clases de clasificación, utilizando un esquema de clasificación de puntuación de $M_{mut} \rightarrow T$ (véase la Tabla 3). De las 12 mutaciones clasificadas como H_BUHC, la cuarta mutación de Lennerz et al. (SNRP116) se clasifica en la 3^{era} posición. Además, si se empleó un umbral más alto (más realista) para M_{wt} (por ejemplo, $\alpha^* \sim 5$), entonces la cuarta mutación de Lennerz et al., se clasifica en la 1^a posición (junto con una única mutación de posición de anclaje adicional - Tabla 7). Finalmente, se predijo que las cuatro mutaciones de Lennerz et al. tienen el alelo correcto de HLA, la longitud del epítipo y la posición de la mutación de acuerdo con lo informado por los autores.

Tabla 7. Mutaciones de células MZ7-MEL clasificadas como H_BUHC. Dos de las 743 expresaron mutaciones no sinónimas expresadas en MZ7-MEL se clasificaron como inmunogénicas para H_BUHC utilizando los umbrales: $\alpha^* = 5$, $\beta = 0,2$ y RPKM > 2. La clasificación se basa en un esquema de clasificación de M_{mut} (descendente) $\rightarrow M_{wt}$ (ascendente). Los neoantígenos inmunogénicos identificados por Lennerz et al. están resaltados en amarillo.

Clasificación*	Gen	M_{mut}	M_{wt}	Media exp	← MZ7-MEL
1	EFTUD2 (SNRP116)	0,15	10,2	22,0	
2	ZNF335	0,2	18,35	8,3	

* La clasificación se basa en M_{mut} y M_{wt}

Conclusiones

El análisis de los conjuntos de datos de B16 y CT26 respalda un modelo en el que se confiere inmunogenicidad si se cumplen tres condiciones: se presenta el péptido de tipo silvestre, se presenta el péptido mutado y la sustitución de aminoácidos tiene una puntuación logarítmica de probabilidades suficientemente baja. (Figura 11A). Este modelo de inmunogenicidad, al que se hace referencia como inmunogenicidad de clase I, está respaldado además en el modelo de línea celular de melanoma humano, MZ7-MEL. El modelo MZ7-MEL y los neoantígenos restringidos CD8+ publicados respaldan un segundo modelo, al que se hace referencia como inmunogenicidad de clase II, en el que no se presenta el epítipo de tipo silvestre, pero una sustitución (por ejemplo, en una posición de anclaje) conduce a un aumento significativo en la puntuación de consenso del MHC (> 5 a 10), lo que da como resultado un epítipo nuevo, nunca antes visto (Fig. 11B). Este marco para definir la inmunogenicidad se captura con un esquema de clasificación de tres variables (M_{mut} , M_{wt} , puntuación T). Usando este esquema de clasificación se pudieron reducir 743 mutaciones de MZ7-MEL a una listado de 34 mutaciones, con 3 de los 5 epítipos de Lennerz et al., que se clasifican en las 5 clases principales.

La Tabla 7 demuestra que las mutaciones inmunogénicas de clase II son raras. De 743 mutaciones, solo 2 se clasificaron como inmunogénicas de clase II (utilizando un umbral realista para M_{wt}) en comparación con aproximadamente 30 mutaciones inmunogénicas de clase I. También se observó una escasez de mutaciones de la clase H_BUHC en los modelos de melanoma de ratón (Tabla 8). Esta observación subraya la importancia de las mutaciones inmunogénicas de clase I para personalizar las vacunas, que se espera sean las mutaciones de tipo dominante que se encuentran en las muestras de pacientes que pueden utilizarse para la vacunación. Al mismo tiempo, el hecho de que uno de los cinco epítipos encontrados por Lennerz et al., era inmunogénico de clase II puede indicar que las mutaciones inmunogénicas de clase II son más potentes o de alguna manera las selecciona el sistema inmunológico.

Tabla 8. Número de mutaciones candidatas de H_A y H_BUHC en diferentes modelos de tumores ($\alpha = 0,8$, $\alpha^* = 5$, $\beta = 0,2$, $\tau = 0,5$).

Cepa	Hipótesis	
	H _A	H _B UHC
B16	27	0
CT26	13	0
MZ7-MEL	35	2

Ejemplo 2: Materiales y métodos

Los materiales y métodos usados en el Ejemplo 1 se describen a continuación:

50

Animales

Los ratones C57BL/6J y Balb/cJ (CRL) se mantuvieron de acuerdo con las políticas federales y estatales sobre investigación con animales en la Universidad de Mainz.

Células para melanoma y modelo de tumor murino colorrectal

La línea celular de melanoma B16F10 (Producto: ATCC CRL-6475, Número de lote: 58078645) y la línea celular de carcinoma de colon CT26.WT (Producto: ATCC CRL-2638, Número de lote: 58494154) se adquirieron en 2010 a través de la American Type Culture Collection. Se usaron pasajes tempranos (tercero, cuarto) de células para experimentos de secuenciación. Las células se analizaron de forma rutinaria para detectar Mycoplasma. No se han autenticado nuevamente las células desde su recepción. La línea celular MZ7-MEL (establecida en enero de 1988) y una línea celular B autóloga transformada por el virus de Epstein-Barr se obtuvieron del Dr. Thomas Wölfel (Departamento de Medicina, Oncología Hematológica, Universidad Johannes Gutenberg).

Péptidos sintéticos

Los péptidos se adquirieron a través de Jerini Peptide Technologies (Berlín, Alemania) o se sintetizaron en la instalación de péptidos TRON. Los péptidos sintéticos tenían 27 aminoácidos de longitud con el aminoácido mutado (MUT) o de tipo silvestre (WT) en la posición 14.

Inmunización de ratones

Se inyectaron por vía intravenosa ratones hembra C57BL/6 o Balb/c de la misma edad con 20 µg de ARNm transcrito *in vitro* formulado con 20 µl de Lipofectamine^{MR} RNAiMAX (Invitrogen) en PBS en un volumen de inyección total de 200 µl (3 ratones por grupo). Los ratones se inmunizaron los días 0, 3, 7, 14 y 18. Veintitrés días después de la inyección inicial, se sacrificaron los ratones y se aislaron los esplenocitos para pruebas inmunológicas (véase el ensayo ELISPOT). Las secuencias de ADN que representan una (monoepítipo) o dos mutaciones (biepítipo) se construyeron utilizando la secuencia de 27 aminoácidos (aa) con la mutación en la posición 14 y se clonaron en la cadena de pST1-2BgUTR-A120 (S. Holtkamp et al., Blood 108, 4009 (2006)). La transcripción *in vitro* de esta plantilla y la purificación se describieron previamente (S. Kreiter et al., Cancer Immunology, Immunotherapy 56, 1577 (2007)).

Ensayo ImmunoSpot ligado a enzimas

Se describieron previamente el ensayo ImmunoSpot ligado a enzima (ELISPOT) (S. Kreiter et al., Cancer Research 70, 9031 (2010)) y la generación de células dendríticas derivadas de médula ósea singénica (BMDC) como estimuladores (L. MB et al., J. Immunol. Methods 223, 77 (1999)). Para el modelo B16F10, las BMDC se pulsaron con péptidos (6 µg/ml), con la mutación indicada, el péptido de tipo silvestre o de control correspondiente (VSV-NP). Para el modelo CT26, además de la reestimulación con péptidos, las BMDC se transfectaron con el ARNm transcrito *in vitro* correspondiente y también se utilizaron para la reestimulación. Para el ensayo, se incubaron 5×10^4 BMDC con 5×10^5 esplenocitos recién aislados en una placa de microtitulación recubierta con anticuerpo anti-IFN γ (10 µg/ml, clon AN18; Mabtech). Después de 18 horas a 37 °C, se detectó la secreción de citocinas con un anticuerpo anti-IFN- γ (clon R4-6A2; Mabtech). Los números de puntos se contaron y analizaron con el analizador ELISPOT ImmunoSpot® S5 Versa, el software de adquisición de imágenes ImmunoCapture^{MR} y el software de análisis ImmunoSpot® Versión 5. El análisis estadístico se realizó mediante la prueba t de Student y la prueba de Mann-Whitney (prueba no paramétrica). Las respuestas se consideraron significativas con un valor $p < 0,05$.

Ensayo de citocinas intracelulares

Se sometieron alícuotas de los esplenocitos preparados para el ensayo ELISPOT a análisis de producción de citocinas mediante citometría de flujo intracelular. Con este fin, se sembraron 2×10^6 esplenocitos por muestra en medio de cultivo (RPMI + FCS al 10%) suplementado con el inhibidor de Golgi Brefeldin A (10 µg/ml) en una placa de 96 pocillos. Las células de cada animal se reestimularon durante 5 h a 37 °C con 2×10^5 BMDC pulsadas con péptidos o transfectadas con ARN. Después de la incubación, las células se lavaron con PBS, se resuspendieron en 50 µl de PBS y se tiñeron extracelularmente con los siguientes anticuerpos anti-ratón durante 20 min a 4 °C: FITC anti-CD4, APC-Cy7 anti-CD8 (BD Pharmingen). Después de la incubación, las células se lavaron con PBS y posteriormente se resuspendieron en 100 µl de solución de Cytotfix/Cytoperm (BD Bioscience) durante 20 min a 4 °C para la permeabilización de la membrana externa. Después de la permeabilización, las células se lavaron con Perm/Tampón de Lavado (BD Bioscience), se resuspendieron en 50 µl/muestra en Perm/Tampón de Lavado y se tiñeron intracelularmente con los siguientes anticuerpos anti-ratón durante 30 min a 4 °C: PE anti-IFN γ , PE-Cy7 anti-TNF- α , APC anti-IL2 (BD Pharmingen). Después de lavar con Perm/Tampón de Lavado, las células se resuspendieron en PBS que contenía paraformaldehído al 1% para análisis de citometría de flujo. Las muestras se analizaron utilizando un citómetro BD FACSCanto^{MR} II y FlowJo (Versión 7.6.3).

Secuenciación de próxima generación

Extracción de ácido nucleico: se extrajeron ADN y ARN de células a granel y ADN de tejidos de ratón usando el kit de sangre y tejido Qiagen DNeasy (para ADN) y el kit Qiagen RNeasy Micro (para ARN).

Secuenciación del exoma de ADN: la captura del exoma para B16F10, C57BL/6J y CT26.WT y la nueva secuenciación del ADN para Balb/cJ se realizaron por triplicado como se describió anteriormente (JC Castle et al., Exploiting the mutanome for tumor vaccination. *Cancer Research* 72, 1081 (2012)). La captura del exoma para la nueva secuenciación del ADN de MZ7-MEL/EBV-B se realizó por duplicado utilizando el ensayo de captura basado en una solución Agilent XT Human all Exon 50 Mb, diseñado para capturar todas las regiones codificantes de proteínas. Se fragmentaron 3 µg de ADN genómico purificado (ADNg) a 150-200 pb utilizando un dispositivo de ultrasonido Covaris S2. Los fragmentos se repararon en los extremos, se fosforilaron en 5' y se adenilaron en 3' de acuerdo con las instrucciones del fabricante. Se ligaron adaptadores de extremos emparejados específicos de indexación Agilent a los fragmentos de ADNg usando una relación molar de adaptador a ADNg de 10:1. La amplificación previa a la captura de 4 ciclos se realizó utilizando los cebadores de PCR de precaptura de indexación InPE 1.0 y SureSelect de Agilent y la polimerasa Herculasa II. Se hibridaron 500 ng de fragmentos de ADNg enriquecidos con PCR, ligados con adaptador, con cebos de captura del exoma de Agilent durante 24 horas a 65 °C. Los complejos de cebo de ADNg/ARN hibridados se eliminaron utilizando perlas magnéticas recubiertas de estreptavidina, se lavaron y los cebos de ARN se escindieron durante la elución en tampón de elución SureSelect. Los fragmentos de ADNg eluidos se amplificaron por PCR después de la captura durante 10 ciclos usando PCR post-captura de indexación SureSelect y cebadores de PCR de índice y polimerasa Herculasa II. Todas las limpiezas se realizaron con 1,8 veces el volumen de perlas magnéticas Agencourt AMPure XP. Todos los controles de calidad se realizaron con el ensayo Qubit HS de Invitrogen y el tamaño del fragmento se determinó mediante el ensayo de ADN 2100 Bioanalyzer HS de Agilent. Las bibliotecas de ADNg enriquecidas del exoma se agruparon en el cBot usando el kit de agrupación Truseq SR v2.5 usando una biblioteca de 7 pM y se secuenciaron 1x100 pb en HiSeq2000 Illumina usando los kits Truseq SBS.

Perfiles de expresión génica de ARN (ARN-Seq): Se prepararon por duplicado bibliotecas de ADNc de ARNm-seq con código de barras, a partir de 5 µg de ARN total (protocolo de ARNm-seq de Illumina modificado utilizando reactivos NEB). Se aisló ARNm utilizando perlas magnéticas de Oligo (dT) Seramag (Thermo Scientific) y se fragmentaron utilizando cationes divalentes y calor. Los fragmentos resultantes (160-220 pb) se convirtieron en ADNc usando cebadores aleatorios y SuperScriptIII (Invitrogen) seguido de síntesis de la segunda cadena usando ADN polimerasa I y RNasaH. El ADNc se reparó en los extremos, se fosforiló en 5' y se adeniló en 3' de acuerdo con las instrucciones del kit de la biblioteca de ARN NEB. Se ligaron adaptadores específicos multiplex de Illumina con una sola proyección en T 3' con ADN ligasa T4 usando una relación molar de 10:1 de adaptador con respecto al inserto de ADNc. Las bibliotecas de ADNc se purificaron y se seleccionaron por tamaño a 300 pb (gel SizeSelect, al 2% E-Gel, Invitrogen). El enriquecimiento, la adición del índice de seis bases de Illumina y las secuencias específicas de la celda de flujo se realizó mediante PCR utilizando la ADN polimerasa Phusion y los cebadores de PCR específicos de Illumina. Todas las limpiezas hasta esta etapa se realizaron con 1,8 veces el volumen de perlas magnéticas Agencourt AMPure XP. Todos los controles de calidad se realizaron con el ensayo Qubit HS de Invitrogen y el tamaño del fragmento se determinó mediante el ensayo de ADN 2100 Bioanalyzer HS de Agilent. Las bibliotecas de ARN-Seq con código de barras se agruparon y se secuenciaron 50 pb como se describió anteriormente.

Análisis de datos de NGS, expresión génica: las lecturas de la secuencia de salida de las muestras de ARN se procesaron previamente de acuerdo con el protocolo estándar de Illumina, incluido el filtrado para lecturas de baja calidad. Las lecturas de secuencia se alinearon con la secuencia genómica de referencia mm9 (A.T. Chinwalla et al., *Nature* 420, 520 (2002)) o hg18 (F. Collins, E. Lander, J. Rogers, R. Waterston, I. Conso, *Nature* 431, 931 (2004)) con Bowtie (versión 0.12.5) (B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, *Genome Biol* 10, R25 (2009)). Para las alineaciones del genoma se permitieron dos desajustes y solo se informó la mejor alineación ("-v2-mejor"), para las alineaciones de transcripción se usaron los parámetros predeterminados. Las lecturas no alineables con la secuencia genómica se alinearon con una base de datos de todas las posibles secuencias de unión exón-exón de los genes conocidos de UCSC (F. Hsu et al., *Bioinformatics* 22, 1036 (2006)). Los valores de expresión se determinaron mediante la intersección de las coordenadas de lectura con las de los transcritos de RefSeq, contando las lecturas de exón y unión superpuestas y normalizando a las unidades de expresión de RPKM (Lecturas que mapean por Kilobase del modelo de exón por millón de lecturas mapeadas) (A. Mortazavi, B.A. Williams, K McCue, L. Schaeffer, B. Wold, *Nature methods* 5, 621 (2008)).

Análisis de datos de NGS, descubrimiento de mutaciones somáticas: Se identificaron mutaciones somáticas como se describió previamente (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. *Cancer Research* 72, 1081 (2012)). La secuencia se lee alineada con el genoma de referencia mm9 o hg18 usando bwa (opciones predeterminadas, versión 0.5.8c) (H. Li, R. Durbin, *Bioinformatics* 25, 1754 (2009)). Se eliminaron las lecturas ambiguas que mapean en múltiples ubicaciones del genoma. Las mutaciones se identificaron mediante un consenso de dos programas de software: samtools (versión 0.1.8) (H. Li, *Bioinformatics* 27, 1157 (2011)) y SomaticSniper (A. McKenna et al., *Genome Research* 20, 1297 (2010)). Para B16F10 y C57BL/6J, también se incluyó GATK (A. McKenna et al., *Genome Research* 20, 1297 (2010)). A las posibles variaciones somáticas identificadas en todas las réplicas respectivas se les asignó un valor de confianza de "tasa de descubrimiento falso" (FDR) (M. Löwer et al., *PLoS computational biology* 8, e1002714 (2012)) (solo CT26 y MZ7-MEL).

Selección y validación de mutaciones

Los criterios para seleccionar las 50 mutaciones de B16F10 para las pruebas de inmunogenicidad se describieron previamente (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. *Cancer Research* 72, 1081 (2012)). Estos criterios para las mutaciones incluyeron: (i) presencia en las tres réplicas de B16F10 y ausencia de todos los triplicados de C57BL/6, (ii) ocurrencia en una transcritpo de RefSeq, (iii) causar un cambio no sinónimo, (iv) ocurrencia en genes expresados en B16F10 (RPKM mediana en las réplicas > 10, expresión del exón > 0) y (v) para cada mutación, la puntuación de M_{mut} (véase más abajo) debía ser < 5. De las 59 mutaciones restantes, el producto de los intervalos de cuantiles de la puntuación de MHC de clase I, la puntuación del MHC de clase II y se formó la expresión del transcritpo, y se seleccionaron las primeras 50 mutaciones ($0,1 \leq M_{mut} \leq 3,9$) para su confirmación mediante PCR (véase (J. C. Castle et al., Exploiting the mutanome for tumor vaccination. *Cancer Research* 72, 1081 (2012)) para más detalles). Los criterios para las 96 mutaciones CT26.WT seleccionadas para las pruebas de inmunogenicidad se refinaron aún más e incluyeron lo siguiente: (i) presencia en todas las tres réplicas CT26.WT y ausencia de todas las tres réplicas de Balb/cJ, (ii) $FDR \leq 0,05$, (iii) ocurrencia en un transcritpo génico conocido de UCSC, (iv) causan cambios no sinónimos, (v) no están presentes en la base de datos de dbSNP (vi) no están en una región de repetición genómica. De las 493 mutaciones restantes, se definieron ocho grupos de 12 miembros de acuerdo con tres características: puntuación de M_{mut} (la más baja - [0,1; 1,9] frente a la más alta - [3,9 - 20,3]), compartimiento de la proteína (extracelular, intracelular) y expresión génica (por debajo frente a por encima de la media de 7,1 RPKM), seleccionando mutaciones de acuerdo con un algoritmo codicioso y ajustando los umbrales en consecuencia. 94 de las 96 mutaciones resultantes se confirmaron mediante PCR seguida de secuenciación de Sanger.

Los criterios para seleccionar mutaciones de MZ7-ML para el análisis incluyeron: (i) presencia en dos réplicas de MZ7-MEL y ausencia de dos réplicas de EBV-B autólogas, seguidas de las etapas (ii) a (vi) descritas anteriormente para CT26.WT. La aplicación de las etapas (i) - (vi) redujo la lista inicial de ~8000 mutaciones a 743.

Predicción de la unión al MHC y cálculo de la puntuación de M_{mut}

Las predicciones de unión al MHC se realizan utilizando la herramienta de consenso de recursos de análisis de IEDB (http://tools.immuneepitope.org/analyze/html/mhc_binding.html) (Y. Kim et al., *Nucleic Acids Research* 40, W525 (2012)), que combina los mejores métodos de predicción de desempeño basados en estudios de evaluación comparativa (HH Lin, S. Ray, S. Tongchusak, EL Reinherz, V. Brusic, *BMC Immunology* 9, 8 (2008); B. Peters et al., *PLoS computational biology* 2, e65 (2006)) de ANN (C. Lundegaard et al., *Nucleic Acids Research* 36, W509 (2008); M. Nielsen et al., *Protein Science* 12, 1007 (2009)), SMM (B. Peters, A. Sette, *BMC bioinformatics* 6, 132 (2005)) y para algunos modelos de alelos también combilib (J. Sidney et al., *Immunome Research* 4, 2 (2008)). El enfoque de consenso combina las puntuaciones de predicción de todas las herramientas generando un intervalo percentil, que refleja las puntuaciones de predicción de unión del péptido dado contra las puntuaciones de péptidos de cinco millones de péptidos aleatorios del SWISSPROT.

Para cada mutación se calcularon las puntuaciones de consenso predichas del MHC para todas las posibles (i) ventanas de secuencia (en las que colocar la mutación), (ii) longitudes de epítipo y (iii) posibles alelos del MHC clase I murino. El mínimo de todas las puntuaciones de consenso del MHC se definió como la puntuación de M_{mut} .

Cálculo de matrices logarítmicas de probabilidades y puntuación T

Las matrices logarítmicas de probabilidades se pueden estimar a partir de comparaciones de alineación de secuencias de grandes bases de datos de proteínas. Las primeras matrices logarítmicas de probabilidades se basaron en la comparación por pares de secuencias (BLOSUM62 (S. Kreiter et al., *Cancer Immunology, Immunotherapy* 56, 1577 (2007))) y el método de estimación de máxima parsimonia (MP) (por ejemplo, PAM250 (MO Dayhoff, R.M. Schwartz, BC Orcutt, *A model for evolutionary change*. MO Dayhoff, ed. *Atlas of protein sequence and structure* Vol.5, 345 (1978)), JTT250 (S.Q. Le, O. Gascuel, *Molecular biology and evolution* 25, 1307 (2008)) y la matriz de Gonnet (C.C .Dang, V. Lefort, V.S. Le, Q.S. Le, O. Gascuel, *Bioinformatics* 27, 2758 (2011))). Más recientemente, se desarrollaron métodos basados en la máxima probabilidad (ML) (por ejemplo, VT160 (P.G. Higgs, TK Attwood, *Bioinformatics and molecular evolution*. (Wiley-Blackwell, 2009)), WAG (S. Whelan, N. Goldman, *Molecular biology and evolution* 18, 691 (2001)) y LG (V. Lennerz et al., *Proceedings of the National Academy of Sciences of the United States of America* 102, 16013 (2005))). Dado que ML no se limita a la comparación de secuencias estrechamente relacionadas, como es el caso de los enfoques basados en MP, se espera que este enfoque de estimación sea el más preciso.

El cálculo de la matriz logarítmica de probabilidad se ha descrito en detalle en otra parte (C. C. Dang, V. Lefort, V. S. Le, Q. S. Le, O. Gascuel, *Bioinformatics* 27, 2758 (2011)). Brevemente, el modelo estándar para la sustitución de aminoácidos asume un modelo Markoviano, continuo en el tiempo y reversible en el tiempo, representado por una matriz de tasa de 20×20 Q_{ij} , en la que q_{ij} ($i \neq j$) es el número de sustituciones del aminoácido i por j por unidad de

tiempo, y en la que los elementos diagonales se eligen para satisfacer $Q_{ii} = -\sum_{j \neq i} Q_{ij}$. Q se puede descomponer de manera que $Q_{ij} = S_{ij} \cdot \pi_j$ para $i \neq j$, en la que S_{ij} es una matriz de intercambiabilidad simétrica, π_i es la probabilidad de observar el aminoácido i (C.C. Dang, V. Lefort, V.S. Le, Q.S. Le, O. Gascuel, *Bioinformatics* 27, 2758 (2011)).

Finalmente, Q se normaliza de manera que $1 = -\sum_i \pi_i Q_{ii}$, una unidad de tiempo $t = 1,0$ corresponde a la sustitución

esperada 1,0 por sitio, o una "mutación puntual aceptada" por sitio, denotado por una distancia PAM de 100 (MO Dayhoff, R.M. Schwartz, B.C. Orcutt, A model for evolutionary change. M.O. Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978); S.Q. Le, O. Gascuel, Molecular biology and evolution 25, 1307 (2008); C.C. Dang, V. Lefort, V.S. Le, Q.S. Le, O. Gascuel, Bioinformatics 27, 2758 (2011)). La probabilidad de que el aminoácido i sea reemplazado por el aminoácido j después del tiempo t , $\Pr(i \rightarrow j | t) = P_{ij}(t)$, está dada por la matriz de probabilidad de 20×20 $P(t) = e^{tQ}$ (con notación que denota exponenciación matricial). La matriz logarítmica de probabilidades

$$T_{i,j} = 10 \log_{10} \left(\frac{\pi_i P_{ij}(t)}{\pi_i \pi_j} \right)$$

calculada para el tiempo t viene dada por la matriz logarítmica de probabilidades 20×20 (M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt, A model for evolutionary change. M.O. Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978, 1978)). Una media reversible en el tiempo que $\pi_i P_{ij}(t) = \pi_j P_{ji}(t)$, y por lo tanto $T_{i,j}$ es simétrica (P. G. Higgs, T. K. Attwood, Bioinformatics and molecular evolution. (Wiley-Blackwell, 2009)).

La puntuación T para la sustitución $i \leftrightarrow j$ se define en este documento como $T_{i,j}$, y depende del modelo evolutivo y del tiempo t . Se exploraron varios modelos y distancias PAM para la puntuación T , incluidos PAM, BLOSUM62, JTT, VT160, Gonnet, WAG, WAG* y LG (véanse las referencias anteriores). Las cifras de este informe se generaron utilizando una puntuación T basada en el modelo WAG y una distancia PAM de 250. Una distancia PAM tan grande significa que existe una posibilidad sustancial de que el aminoácido cambie (P.G. Higgs, TK Attwood, Bioinformatics and molecular evolution (Wiley-Blackwell, 2009)), y es útil para detectar relaciones distantes entre secuencias en las que los residuos pueden no ser idénticos pero las propiedades fisicoquímicas de los aminoácidos se conservan (M.O. Dayhoff, RM Schwartz, B.C. Orcutt, A model for evolutionary change M.O. Dayhoff, ed. Atlas of protein sequence and structure Vol.5, 345 (1978); P.G. Higgs, T.K. Attwood, Bioinformatics and molecular evolution (Wiley-Blackwell, 2009)).

Usando una estadística de prueba de distribución t , se compararon las puntuaciones T medias de epítomos inmunogénicos frente a no inmunogénicos de la Tabla 3 para la matriz WAG utilizando varias puntuaciones PAM (1, 10, 25, 50, 100, 150, 200 y 250). El análisis de la estadística de prueba mostró que el valor P disminuyó en forma monótona con la distancia PAM, lo que implica que una distancia PAM de 250 era la solución óptima, como era de esperar (datos no mostrados). La clasificación en H_A y H_n fue la misma para todas las matrices excepto para la matriz PAM, que es la menos precisa de todos los modelos evolutivos. De todos los modelos evolutivos, el modelo WAG250 resultó en la máxima separación entre los epítomos H_A y H_n en la Tabla 3, midiendo la separación con la estadística de prueba: [puntuación T máxima (H_A) - puntuación T mínima (H_n)] / σ (puntuación T (H_A), puntuación T (H_n)) (datos no mostrados). La misma estadística de prueba también fue máxima para una distancia PAM 250 en comparación con una distancia más pequeña.

Epítomos CD8+ publicados

Se recogieron epítomos CD8+ con aminoácidos mutados simples de la lista de antígenos tumorales resultantes de mutaciones publicadas por el Cancer Immunity Journal (P. Van der Bruggen, V. Stroobant, N. Vigneron, B. Van den Eynde. (Cancer Immun, <http://www.cancerimmunity.org/peptide/>, 2013) (<http://cancerimmunity.org/peptide/mutations/>) (<http://cancerimmunity.org/peptide/mutations/>). Los alelos de HLA se tomaron de la tabla publicada o del artículo original si este último era más preciso. Las referencias enumeradas en la Tabla 4 son las siguientes: (1) Lennerz et al., PNAS 102 (44), págs. 16013-16018 (2005); (2) Karanikas et al., Cancer Res 61 (9), págs. 3718-3724 (2001); (3) Sensi et al., Cancer Res 65 (2), págs. 632-640 (2005); (4) Linard et al., J. Immunol 168 (9), págs. 4802-4808 (2002); (5) Zorn et al., Eur. J. Immunol 29 (2), págs. 592-601 (1999); (6) Graf et al., Blood 109 (7), págs. 2985-2988 (2007)); (7) Robbins et al., J. Exp. Med 183 (3), págs. 1185-1192 (1996); (8) Vigneron et al., Cancer Immun 2, pág. 9 (2002); (9) Echchakir et al., Cancer Res 61 (10), págs 4078-4083 (2001); (10) Hogan et al., Cancer Res 58 (22), págs. 5144-5150 (1998); (11) Ito et al., En t. J. Cancer 120 (12), págs. 2618-2624 (2007); (12) Wölfel et al., Science 269 (5228), págs. 1281-1284 (1995); (13) Gjertsen et al., En t. J. Cancer 72 (5), págs. 784-790 (1997).

Ejemplo 3: Ejemplo de un esquema para ponderar las puntuaciones de mutación para mejorar la priorización de mutaciones inmunogénicas

El ARN que se inyecta en la célula, una vez traducido y escindido en péptidos cortos, puede presentarse en diferentes tipos de HLA dentro de la célula. Por lo tanto, es lógico pensar que cuantos más tipos de HLA se predice que tengan una puntuación de consenso del MHC baja (o similar), es más probable que una mutación determinada sea inmunogénica, ya que potencialmente puede mostrarse en más de un tipo de HLA en paralelo. Por lo tanto, sopesar las mutaciones por el número de tipos de HLA para los que la mutación se clasifica como H_A y/o $H_B UH_C$ o incluso sopesar cada mutación simplemente por el número de tipos de HLA que tienen una puntuación de M_{mut} baja puede mejorar la clasificación de inmunogenicidad. En la solución más general, cuando se inyecta un ARN de 27 mer o péptido en la célula, no solo existe la libertad de seleccionar el tipo de HLA, sino también la longitud del péptido y la posición de la mutación dentro de este péptido. Por lo tanto, se pueden escanear todos los tipos de HLA posibles, todas las longitudes de ventana posibles y todas las posiciones posibles para la mutación dentro de la ventana y calcular el número de soluciones (por mutación dada) que se clasifican como H_A y/o $H_B UH_C$ (Fig.12). Este puede ser un factor de ponderación importante para la priorización de mutaciones para seleccionar los epítomos más eficaces para la vacunación. En la Figura 13 se muestra un ejemplo de un diagrama de dispersión de todas estas soluciones en función de M_{mut} y $\Delta M = M_{mut} - M_{wt}$.

Listado de Secuencias

- 5 <110> BioNTech AG et al.
 <120> PREDICCIÓN DE INMUNOGENICIDAD DE LOS EPÍTOPOS DE CÉLULAS T
 <130> 674-99 PCT
- 10 <150> PCT/EP2013/001400
 <151> 2013-05-10
 <160> 94
- 15 <170> PatentIn versión 3.5
 <210> 1
 <211> 15
 <212> PRT
 20 <213> Secuencia Artificial
 <220>
 <223> Secuencia enlazadora
- 25 <220>
 <221> REPETICIÓN
 <222> (1)..(3)
 <223> Porción de secuencia repetida a veces, en la que a es independientemente un número seleccionado entre 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20
- 30 <220>
 <221> CARACTERÍSTICA MISCELÁNEA
 <222> (1)..(15)
 <223> a + b + c + d + e son diferentes de 0 y preferiblemente son 2 o más, 3 o más, 4 o más o 5 o más
- 35 <220>
 <221> REPETICIÓN
 <222> (4)..(6)
 <223> Porción de secuencia repetida b veces, en la que b es independientemente un número seleccionado entre 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20
- 40 <220>
 <221> REPETICIÓN
 <222> (7)..(9)
 <223> Porción de secuencia repetida c veces, en la que c es independientemente un número seleccionado entre 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20
- 45 <220>
 <221> REPETICIÓN
 <222> (10)..(12)
 50 <223> Porción de secuencia repetida d veces, en la que d es independientemente un número seleccionado entre 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20
- <220>
 <221> REPETICIÓN
 <222> (13)..(15)
 <223> Porción de secuencia repetida e veces, en la que e es independientemente un número seleccionado entre 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20
- 60 <400> 1
 Gly Gly Ser Gly Ser Ser Gly Gly Gly Ser Ser Gly Gly Ser Gly
 1 5 10 15
 <210> 2
 <211> 9

ES 2 871 384 T3

<212> PRT
 <213> Secuencia Artificial

 <220>
 5 <223> Secuencia enlazadora

 <400> 2

 gly gly ser gly gly gly gly ser gly
 1 5
 10
 <210> 3
 <211> 11
 <212> PRT
 <213> Secuencia Artificial
 15
 <220>
 <223> Epítopo

 <400> 3
 20
 Ala Ala Val Ile Leu Arg Asp Ala Leu His Met
 1 5 10

 <210> 4
 <211> 11
 25 <212> PRT
 <213> Secuencia Artificial

 <220>
 <223> Epítopo
 30
 <400> 4

 Ala Ala Val Ile Leu Arg Val Ala Leu His Met
 1 5 10
 35
 <210> 5
 <211> 9
 <212> PRT
 <213> Secuencia Artificial
 40
 <220>
 <223> Epítopo

 <400> 5

 gly gly Pro gly Ser Glu Lys Ser Leu
 1 5
 45
 <210> 6
 <211> 9
 <212> PRT
 50 <213> Secuencia Artificial

 <220>
 <223> Epítopo
 55
 <400> 6

 gly gly Pro gly Ser Gly Lys Ser Leu
 1 5
 60
 <210> 7
 <211> 10

ES 2 871 384 T3

<212> PRT
 <213> Secuencia Artificial

 <220>
 5 <223> Epítopo

 <400> 7

 Leu Ala Leu Pro Asn Asn Tyr Cys Asp Val
 1 5 10
 10
 <210> 8
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 15
 <220>
 <223> Epítopo

 <400> 8
 20
 Leu Ala Leu Pro Asn Asn Tyr Cys Asp Phe
 1 5 10

 <210> 9
 <211> 10
 25 <212> PRT
 <213> Secuencia Artificial

 <220>
 <223> Epítopo
 30
 <400> 9

 Ser His Leu Asn Asn Asp Val Trp Gln Ile
 1 5 10
 35
 <210> 10
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 40
 <220>
 <223> Epítopo

 <400> 10

 Ser His Leu Asn Asn Asp Phe Trp Gln Ile
 1 5 10
 45
 <210> 11
 <211> 9
 <212> PRT
 50 <213> Secuencia Artificial

 <220>
 <223> Epítopo
 55
 <400> 11

 Tyr Tyr Met Arg Asp Val Ile Ala Ile -
 1 5
 60
 <210> 12
 <211> 9

ES 2 871 384 T3

<212> PRT
 <213> Secuencia Artificial

 <220>
 5 <223> Epítopo

 <400> 12

 Tyr Tyr Met Arg Asp Val Thr Ala Ile
 1 5
 10
 <210> 13
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 15
 <220>
 <223> Epítopo

 <400> 13
 20
 Thr Tyr Leu Gln Pro Ala Gln Ala Gln Met
 1 5 10

 <210> 14
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 25
 <220>
 <223> Epítopo
 30
 <400> 14

 Ile Tyr Leu Gln Pro Ala Gln Ala Gln Met
 1 5 10
 35
 <210> 15
 <211> 8
 <212> PRT
 <213> Secuencia Artificial
 40
 <220>
 <223> Epítopo

 <400> 15

 Gln Ser Leu Gly Phe Thr Tyr Leu
 1 5
 45
 <210> 16
 <211> 8
 <212> PRT
 <213> Secuencia artificial
 50
 <220>
 <223> Epítopo
 55
 <400> 16

 Gln Arg Leu Gly Phe Thr Tyr Leu
 1 5
 60
 <210> 17
 <211> 10

ES 2 871 384 T3

<212> PRT
 <213> Secuencia Artificial

 <220>
 5 <223> Epítopo

 <400> 22

 val Thr Phe Gln Ala Phe Ile Asp Phe Met Ser
 1 5 10
 10
 <210> 23
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 15
 <220>
 <223> Epítopo

 <400> 23
 20
 Pro Tyr Leu Thr Ala Leu Asp Asp Leu Leu
 1 5 10

 <210> 24
 <211> 10
 25 <212> PRT
 <213> Secuencia Artificial

 <220>
 <223> Epítopo
 30
 <400> 24

 Pro Tyr Leu Thr Ala Leu Gly Asp Leu Leu
 1 5 10
 35
 <210> 25
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 40
 <220>
 <223> Epítopo

 <400> 25

 Ala Gly Gly Leu Phe Val Ala Asp Ala Ile
 1 5 10
 45
 <210> 26
 <211> 10
 <212> PRT
 50 <213> Secuencia Artificial

 <220>
 <223> Epítopo
 55
 <400> 26

 Ala Gly Gly Leu Phe Val Ala Asp Glu Ile
 1 5 10
 60
 <210> 27
 <211> 10
 <212> PRT
 <213> Secuencia Artificial

ES 2 871 384 T3

<220>
 <223> Epitopo
 5 <400> 27
 val Gly Ile Asn Phe Leu Gln Ser Tyr Gln
 1 5 10
 <210> 28
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 <220>
 <223> Epitopo
 15 <400> 28
 val Gly Ile Asn Ser Leu Gln Ser Tyr Gln
 1 5 10
 20 <210> 29
 <211> 9
 <212> PRT
 <213> Secuencia Artificial
 25 <220>
 <223> Epitopo
 <400> 29
 30 Thr Arg Pro Ala Arg Asp Gly Thr Phe
 1 5
 <210> 30
 <211> 9
 <212> PRT
 <213> Secuencia Artificial
 35 <220>
 <223> Epitopo
 <400> 30
 40 Thr Arg Pro Ala Gly Asp Gly Thr Phe
 1 5
 <210> 31
 <211> 9
 <212> PRT
 <213> Secuencia Artificial
 45 <220>
 <223> Epitopo
 <400> 31
 50 Glu Pro Gln Ile Ala Met Asp Asp Met
 1 5
 <210> 32
 <211> 9
 <212> PRT
 <213> Secuencia Artificial
 60

ES 2 871 384 T3

<220>
 <223> Epítopo
 <400> 32
 5
 Glu Pro Gln Ile Asp Met Asp Asp Met
 1 5
 <210> 33
 <211> 9
 10 <212> PRT
 <213> Secuencia Artificial
 <220>
 <223> Epítopo
 15 <400> 33
 Ile Ala Met Gln Asn Thr Thr Gln Leu
 1 5
 20 <210> 34
 <211> 9
 <212> PRT
 <213> Secuencia Artificial
 25 <220>
 <223> Epítopo
 <400> 34
 30 Ile Ala Ile Gln Asn Thr Thr Gln Leu
 1 5
 <210> 35
 <211> 10
 <212> PRT
 35 <213> Secuencia Artificial
 <220>
 <223> Epítopo
 <400> 35
 40 Ala Ile Tyr His His Ala Ser Arg Ala Ile
 1 5 10
 <210> 36
 <211> 10
 45 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Epítopo
 50 <400> 36
 Ala Ile Tyr Tyr His Ala Ser Arg Ala Ile
 1 5 10
 55 <210> 37
 <211> 10
 <212> PRT
 <213> Secuencia Artificial
 60 <220>
 <223> Epítopo

ES 2 871 384 T3

<400> 37

ser Tyr Ile Ala Leu Val Asp Lys Asn Ile
1 5 10

5 <210> 38
<211> 10
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Epitopo

<400> 38

15 ser Tyr Leu Ala Leu Val Asp Lys Asn Ile
1 5 10

<210> 39
<211> 9
<212> PRT
20 <213> Secuencia Artificial

<220>
<223> Epitopo

25 <400> 39

val Ile Pro Ile Leu Glu Met Gln Phe
1 5

<210> 40
<211> 9
<212> PRT
30 <213> Secuencia Artificial

<220>
35 <223> Epitopo

<400> 40

40 val Ile Pro Ile Leu Glu Val Gln Phe
1 5

<210> 41
<211> 10
<212> PRT
45 <213> Secuencia Artificial

<220>
<223> Epitopo

<400> 41

50 gly Tyr Leu Gln phe Ala Tyr Asp Gly Arg
1 5 10

<210> 42
<211> 10
55 <212> PRT
<213> Secuencia Artificial

<220>
<223> Epitopo

60 <400> 42

ES 2 871 384 T3

		Glu Thr Val Ser Glu Gln Ser Asn Val
		1 5
5	<210> 58	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
	<220>	
10	<223> Epitopo	
	<400> 58	
		Glu Thr Val Ser Glu Glu Ser Asn Val
		1 5
15	<210> 59	
	<211> 10	
	<212> PRT	
	<213> Secuencia Artificial	
20	<220>	
	<223> Epitopo	
	<400> 59	
25		Lys Ile Leu Asp Ala Val Val Ala Gln Lys
		1 5 10
	<210> 60	
	<211> 10	
30	<212> PRT	
	<213> Secuencia Artificial	
	<220>	
	<223> Epitopo	
35	<400> 60	
		Lys Ile Leu Asp Ala Val Val Ala Gln Glu
		1 5 10
40	<210> 61	
	<211> 9	
	<212> PRT	
	<213> Secuencia Artificial	
45	<220>	
	<223> Epitopo	
	<400> 61	
50		Lys Ile Asn Lys Asn Pro Lys Tyr Lys
		1 5
	<210> 62	
	<211> 9	
	<212> PRT	
55	<213> Secuencia Artificial	
	<220>	
	<223> Epitopo	
60	<400> 62	

ES 2 871 384 T3

Lys Glu Leu Glu Gly Ile Leu Leu Leu
 1 5

5 <210> 68
 <211> 9
 <212> PRT
 <213> Secuencia Artificial

10 <220>
 <223> Epitopo
 <400> 68

Lys Glu Leu Glu Gly Ile Leu Leu Pro
 1 5

15 <210> 69
 <211> 10
 <212> PRT
 <213> Secuencia Artificial

20 <220>
 <223> Epitopo
 <400> 69

25 Thr Leu Asp Trp Leu Leu Gln Thr Pro Lys
 1 5 10

30 <210> 70
 <211> 10
 <212> PRT
 <213> Secuencia Artificial

35 <220>
 <223> Epitopo
 <400> 70

Thr Leu Gly Trp Leu Leu Gln Thr Pro Lys
 1 5 10

40 <210> 71
 <211> 10
 <212> PRT
 <213> Secuencia Artificial

45 <220>
 <223> Epitopo
 <400> 71

Phe Ile Ala Ser Asn Gly Val Lys Leu Val
 1 5 10

50 <210> 72
 <211> 10
 <212> PRT
 <213> Secuencia Artificial

55 <220>
 <223> Epitopo
 <400> 72

ES 2 871 384 T3

Phe Ile Ala Ser Lys Gly Val Lys Leu Val
 1 5 10

5 <210> 73
 <211> 9
 <212> PRT
 <213> Secuencia artificial

10 <220>
 <223> Epitopo

<400> 73

Val Val Pro Cys Glu Pro Pro Glu Val
 1 5

15 <210> 74
 <211> 9
 <212> PRT
 <213> Secuencia Artificial

20 <220>
 <223> Epitopo

<400> 74

25 Val Val Pro Tyr Glu Pro Pro Glu Val
 1 5

30 <210> 75
 <211> 10
 <212> PRT
 <213> Secuencia Artificial

35 <220>
 <223> Epitopo

<400> 75

Ala Cys Asp Pro His Ser Gly His Phe Val
 1 5 10

40 <210> 76
 <211> 10
 <212> PRT
 <213> Secuencia artificial

45 <220>
 <223> Epitopo

<400> 76

Ala Arg Asp Pro His Ser Gly His Phe Val
 1 5 10

50 <210> 77
 <211> 9
 <212> PRT
 <213> Secuencia artificial

55 <220>
 <223> Epitopo

<400> 77

60

ES 2 871 384 T3

val val val Gly Ala val Gly val Gly
1 5

5 <210> 78
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Epítopo

<400> 78

val val val Gly Ala Gly Gly val Gly
1 5

15 <210> 79
<211> 9
<212> PRT
<213> Secuencia Artificial

20 <220>
<223> Epítopo

<400> 79

25 Gly Leu Ala Gly Gly Leu Leu Ala Leu
1 5

30 <210> 80
<211> 9
<212> PRT
<213> Secuencia Artificial

35 <220>
<223> Epítopo

<400> 80

val Leu Phe Trp Gly Leu Leu Leu Leu
1 5

40 <210> 81
<211> 9
<212> PRT
<213> Secuencia artificial

45 <220>
<223> Epítopo

<400> 81

Gly Ile Leu Gly Trp val Leu Tyr Leu
1 5

50 <210> 82
<211> 9
<212> PRT
<213> Secuencia artificial

55 <220>
<223> Epítopo

60 <400> 82

ES 2 871 384 T3

Val Leu Phe Trp Gly Leu Leu Phe Tyr
1 5

5 <210> 83
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Epitopo
<400> 83

Gly Leu Ala Gly Gly Leu Leu Ala Tyr
1 5

15 <210> 84
<211> 9
<212> PRT
<213> Secuencia artificial

20 <220>
<223> Epitopo
<400> 84

Val Ile Phe Trp Gly Leu Leu Val Leu
1 5

25 <210> 85
<211> 10
<212> PRT
30 <213> Secuencia artificial

<220>
<223> Epitopo

35 <400> 85

Lys Ile Leu Asp Ala Val Val Ala Gln Glu
1 5 10

40 <210> 86
<211> 10
<212> PRT
<213> Secuencia Artificial

45 <220>
<223> Epitopo

<400> 86

Lys Ile Leu Asp Ala Val Val Ala Gln Lys
1 5 10

50 <210> 87
<211> 9
<212> PRT
<213> Secuencia artificial

55 <220>
<223> Epitopo

<400> 87

ES 2 871 384 T3

Gln His Ser Ala Ala Pro Gly Pro Pro
1 5

5 <210> 88
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Epitopo

<400> 88

Gln His Ser Ala Ala Pro Gly Pro Leu
1 5

15 <210> 89
<211> 9
<212> PRT
<213> Secuencia artificial

20 <220>
<223> Epitopo

<400> 89

25 Glu Ile Asn Lys Asn Pro Lys Tyr Lys
1 5

30 <210> 90
<211> 9
<212> PRT
<213> Secuencia Artificial

35 <220>
<223> Epitopo

<400> 90

Lys Ile Asn Lys Asn Pro Lys Tyr Lys
1 5

40 <210> 91
<211> 10
<212> PRT
<213> Secuencia artificial

45 <220>
<223> Epitopo

<400> 91

50 Tyr Val Asp Phe Arg Glu Tyr Glu Tyr Asp
1 5 10

55 <210> 92
<211> 10
<212> PRT
<213> Secuencia Artificial

60 <220>
<223> Epitopo

<400> 92

ES 2 871 384 T3

Tyr Val Asp Phe Arg Glu Tyr Glu Tyr Tyr
1 5 10

5 <210> 93
<211> 9
<212> PRT
<213> Secuencia Artificial

10 <220>
<223> Epitopo

<400> 93

Ser Tyr Leu Asp Ser Gly Ile His Ser
1 5

15 <210> 94
<211> 9
<212> PRT
<213> Secuencia Artificial

20 <220>
<223> Epitopo

<400> 94

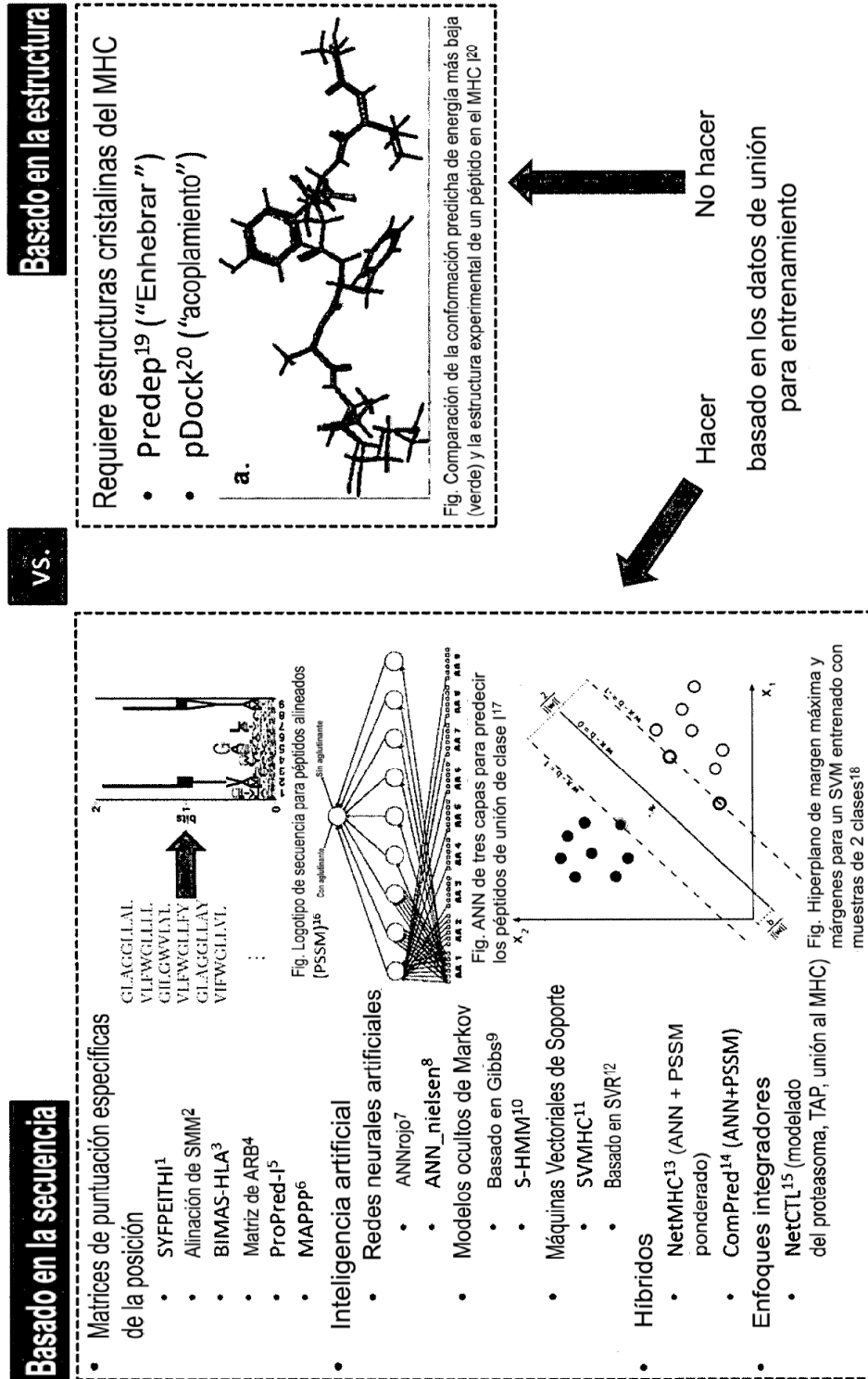
25 Ser Tyr Leu Asp Ser Gly Ile His Phe
1 5

REIVINDICACIONES

1. Un método implementado por ordenador para predecir péptidos modificados inmunogénicos que comprenden modificaciones de aminoácidos, siendo dichas modificaciones sustituciones, comprendiendo el método las etapas:
- 5 a) determinar una puntuación para la unión de un péptido modificado a una o más moléculas del MHC,
 b) determinar una puntuación para la unión del péptido no modificado a la una o más moléculas del MHC, y
 c) determinar una puntuación para la unión del péptido modificado cuando está presente en un complejo MHC-péptido
- 10 a uno o más receptores de células T, que comprende determinar una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados, en la que la puntuación de las similitudes químicas y físicas se determina sobre la base de la probabilidad de que los aminoácidos se intercambien en la naturaleza, en la que cuanto más frecuentemente se intercambian los aminoácidos en la naturaleza, más similares se consideran los aminoácidos, en la que se determinan las similitudes químicas y físicas utilizando matrices de sustitución,
- 15 en el que el péptido no modificado y el péptido modificado son idénticos pero para la modificación o modificaciones, en el que el péptido no modificado y el péptido modificado tienen una longitud de 8 a 15 aminoácidos, y en el que se predice que el péptido modificado es inmunogénico si (i) el péptido no modificado tiene una puntuación para unirse a una o más moléculas del MHC que satisfacen un umbral que indica la unión a una o más moléculas del MHC y (ii) el péptido modificado tiene una puntuación para unirse a una o más moléculas del MHC que satisfacen un umbral que
- 20 indica la unión a una o más moléculas del MHC y (iii) los aminoácidos no modificados y modificados tienen una puntuación para las similitudes químicas y físicas que satisfacen un umbral que indica la diferencia química y física.
2. El método de la reivindicación 1, en el que el péptido modificado comprende un fragmento de una proteína modificada, comprendiendo dicho fragmento la modificación o modificaciones presentes en la proteína.
- 25 3. El método de la reivindicación 1 o 2, en el que el péptido no modificado tiene el aminoácido de línea germinal en la posición o posiciones correspondientes a la posición o posiciones de la modificación o modificaciones en el péptido modificado.
- 30 4. El método de una cualquiera de las reivindicaciones 1 a 3, en el que el péptido no modificado y el péptido modificado tienen una longitud de 8 a 12 aminoácidos.
5. El método de una cualquiera de las reivindicaciones 1 a 4, en el que una o más moléculas del MHC comprenden diferentes tipos de moléculas del MHC, en particular diferentes alelos del MHC.
- 35 6. El método de una cualquiera de las reivindicaciones 1 a 5, en el que una o más moléculas del MHC son moléculas del MHC de clase I y/o moléculas del MHC de clase II.
- 40 7. El método de una cualquiera de las reivindicaciones 1 a 6, en el que la puntuación para la unión a una o más moléculas del MHC se determina mediante un proceso que comprende una comparación de secuencia con una base de datos de motivos de unión al MHC.
- 45 8. El método de una cualquiera de las reivindicaciones 1 a 7, en el que tanto la etapa a) como la etapa b) comprenden determinar si dicha puntuación satisface un umbral predeterminado para unirse a una o más moléculas del MHC, y en el que el umbral aplicado en la etapa a) es diferente al umbral aplicado en la etapa b).
9. El método de la reivindicación 8, en el que el umbral predeterminado de unión a una o más moléculas del MHC refleja una probabilidad de unión a una o más moléculas del MHC.
- 50 10. El método de una cualquiera de las reivindicaciones 1 a 9, en el que las similitudes químicas y físicas se determinan usando matrices logarítmicas de probabilidades con base en la evolución.
11. El método de una cualquiera de las reivindicaciones 1 a 10, en el que la modificación no está en una posición de anclaje para unirse a una o más moléculas del MHC.
- 55 12. El método de una cualquiera de las reivindicaciones 1 a 11, que comprende realizar la etapa a) en dos o más péptidos modificados diferentes, comprendiendo dichos dos o más péptidos modificados diferentes la misma modificación o modificaciones.
- 60 13. El método de la reivindicación 12, en el que los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones comprenden fragmentos diferentes de una proteína modificada, comprendiendo dichos fragmentos diferentes la misma modificación o modificaciones presentes en la proteína.
- 65 14. El método de la reivindicación 12 o 13, en el que los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones comprende todos los fragmentos de unión al MHC potenciales de una proteína modificada, comprendiendo dichos fragmentos la misma modificación o modificaciones presentes en la proteína.

- 5 15. El método de cualquiera de las reivindicaciones 12 a 14, que comprende además seleccionar (el, los) péptido o péptidos de los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones que tienen una probabilidad o tienen la probabilidad más alta de unirse a una o más moléculas del MHC.
16. El método de una cualquiera de las reivindicaciones 12 a 15, en el que los dos o más péptidos modificados diferentes que comprenden la misma modificación o modificaciones difieren en la longitud y/o posición de la modificación o modificaciones.
- 10 17. El método de una cualquiera de las reivindicaciones 1 a 16, que comprende realizar la etapa a) y opcionalmente uno o ambas etapas b) y c) en dos o más péptidos modificados diferentes.
18. El método de la reivindicación 17, en el que dichos dos o más péptidos modificados diferentes comprenden la misma modificación o modificaciones y/o comprenden modificaciones diferentes.
- 15 19. El método de la reivindicación 18, en el que las diferentes modificaciones están presentes en la misma y/o en diferentes proteínas.
- 20 20. El método de una cualquiera de las reivindicaciones 12 a 19, que comprende comparar las puntuaciones de dos o más de dichos péptidos modificados diferentes.
- 25 21. El método de la reivindicación 20, en el que una puntuación para la unión del péptido modificado a una o más moléculas del MHC se pondera más alta que una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados, y una puntuación para las similitudes químicas y físicas entre los aminoácidos no modificados y modificados se ponderan más alto que una puntuación para la unión del péptido no modificado a una o más moléculas del MHC.
- 30 22. El método de una cualquiera de las reivindicaciones 1 a 21, que comprende además identificar mutaciones no sinónimas en una o más regiones codificantes de proteínas.
- 35 23. El método de una cualquiera de las reivindicaciones 1 a 22, en el que las modificaciones se identifican secuenciando parcial o completamente el genoma o transcriptoma de una o más células tales como una o más células cancerosas y opcionalmente una o más células no cancerosas e identificando mutaciones en una o más regiones codificantes de proteínas.
- 40 24. El método de la reivindicación 22 o 23, en el que dichas mutaciones son mutaciones somáticas.
25. El método de una cualquiera de las reivindicaciones 22 a 24, en el que dichas mutaciones son mutaciones cancerosas.
- 45 26. Un método para proporcionar una vacuna que comprende las etapas:
- a) identificar (a) modificación o modificaciones o (a) péptido o péptidos modificados que se predice que son inmunogénicos mediante el método de una cualquiera de las reivindicaciones 1 a 25, y
- b) proporcionar una vacuna que comprende un péptido o polipéptido que comprende la modificación o modificaciones o un péptido o péptidos modificados que se predice que son inmunogénicos, o un ácido nucleico que codifica el péptido o polipéptido.

Fig. 1



¹Rammensee HJ (1999) Immunogenetics, ²Sette A (2005) Bioinformatics, ³Parker KC (1994) J Immunol, ⁴Sette A (2005) Immunogenetics, ⁵Sing H (2003) Bioinformatics, ⁶Hakenberg J (2003) Appl. Bioinformatics, ⁷Bhatia V (2004) Infect. Immun., ⁸Nielsen M (2003) Protein Sci, ⁹Nielsen M (2004) Bioinformatics, ¹⁰Brusic V (2002) J Biosci Bioeng., ¹¹Dönnes P (2006) NAR, ¹²Flower et al. (2006) BMC Bioinform., ¹³Lundegaard C (2008) Bioinformatics, ¹⁴Brusic V (2004) Methods, ¹⁵Larsen MV (2005) Eur J Immunol., ¹⁶Lund, Nielsen: Immunological Bioinformatics, 2005. MIT Press. Página 117, ¹⁷Brusic V (1994) Complex Systems, ¹⁸http://en.wikipedia.org/wiki/Support_vector_machine, ¹⁹Altuvia Y (2004) Methods, ²⁰Khan JM (2010). Imm Research

Fig. 2

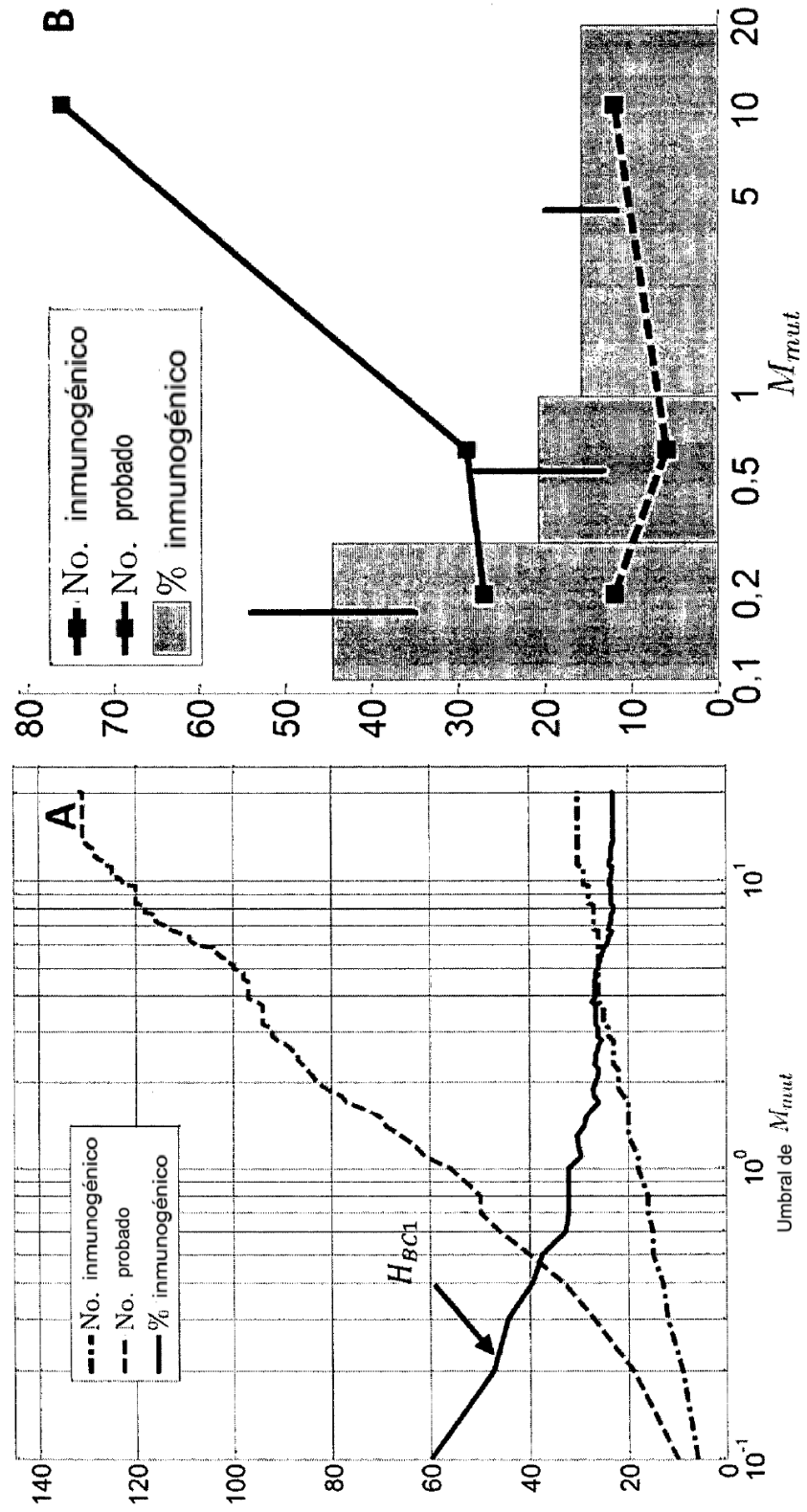


Fig. 3

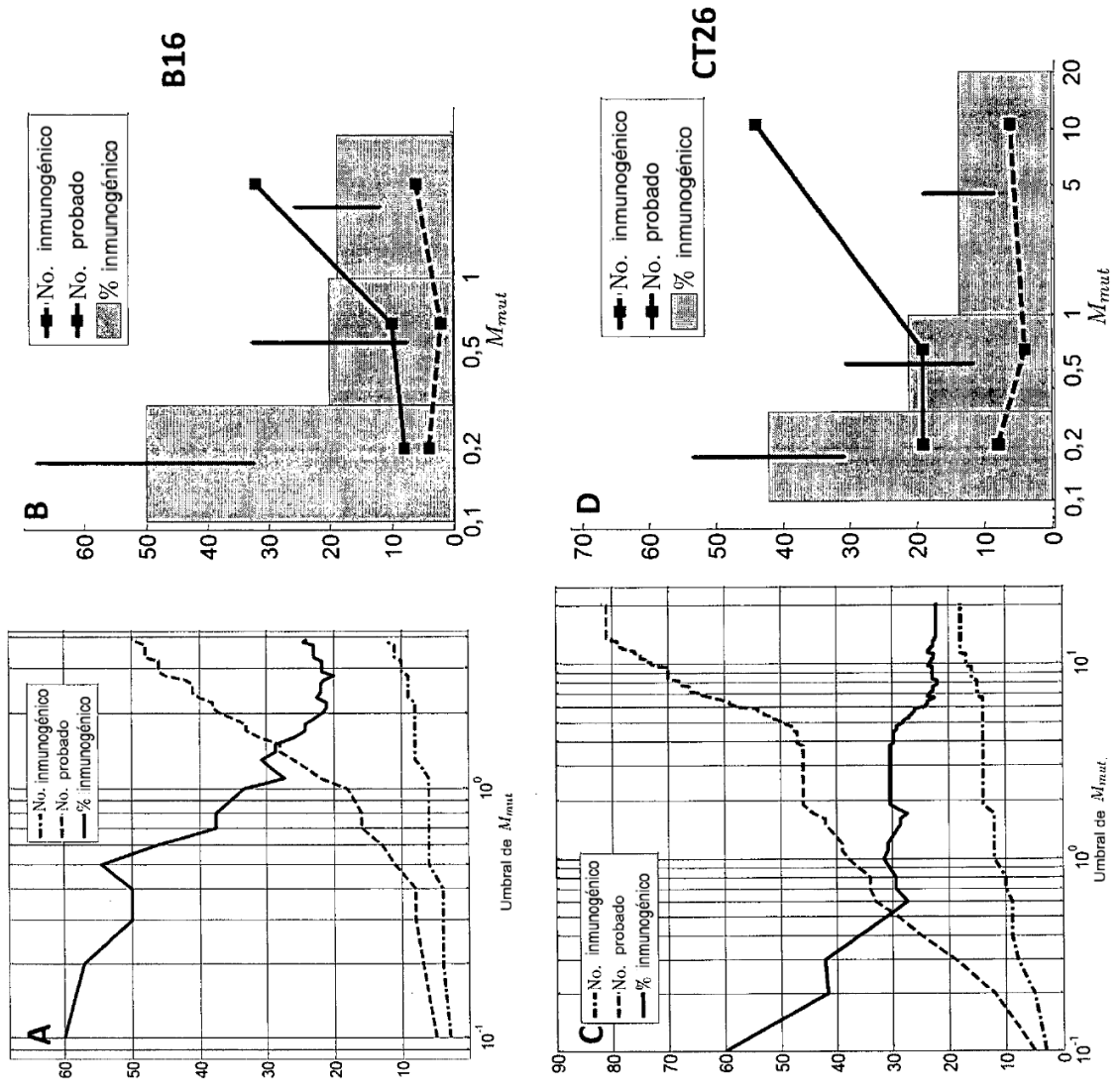


Fig. 4

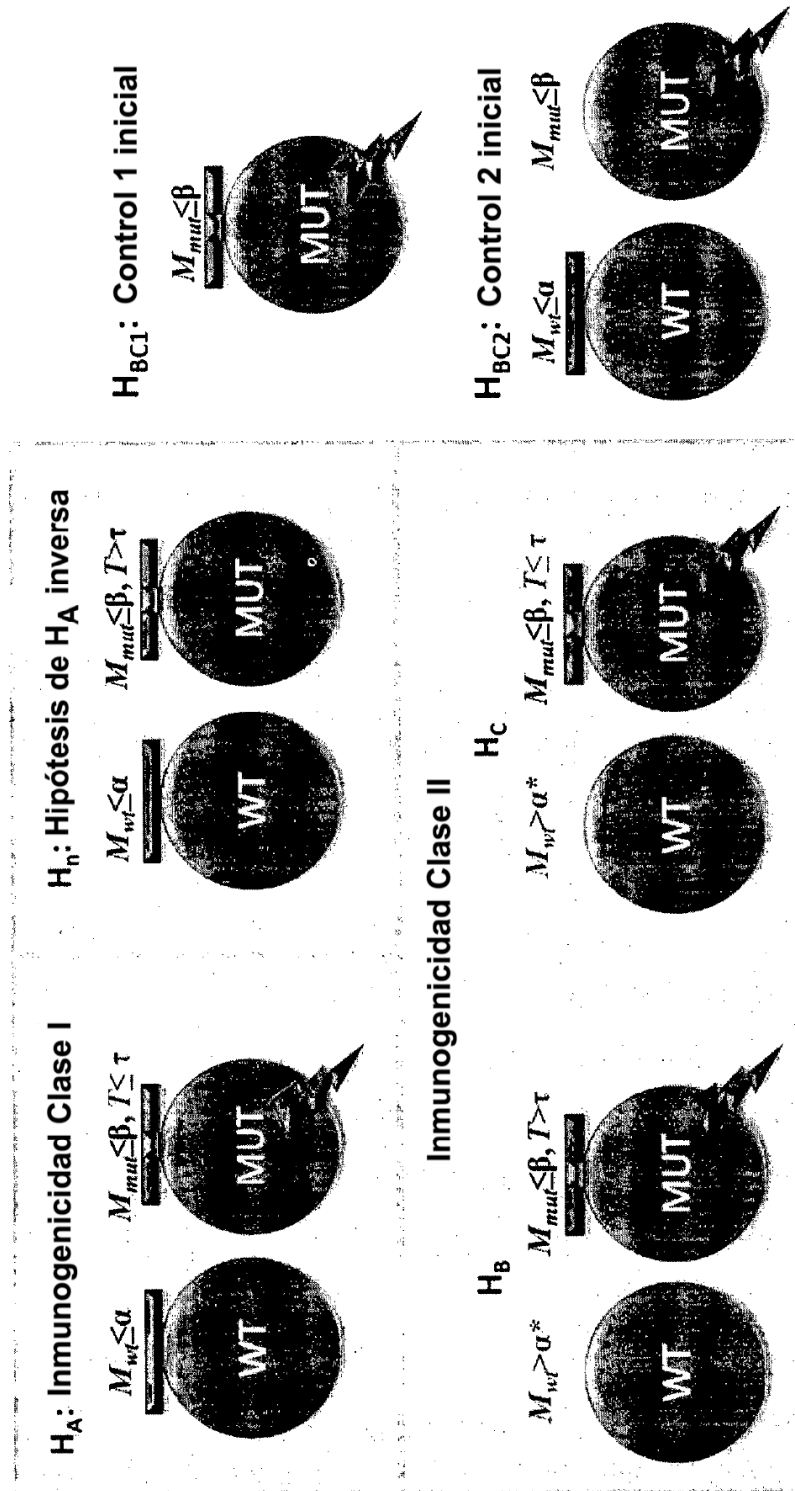


Fig. 5

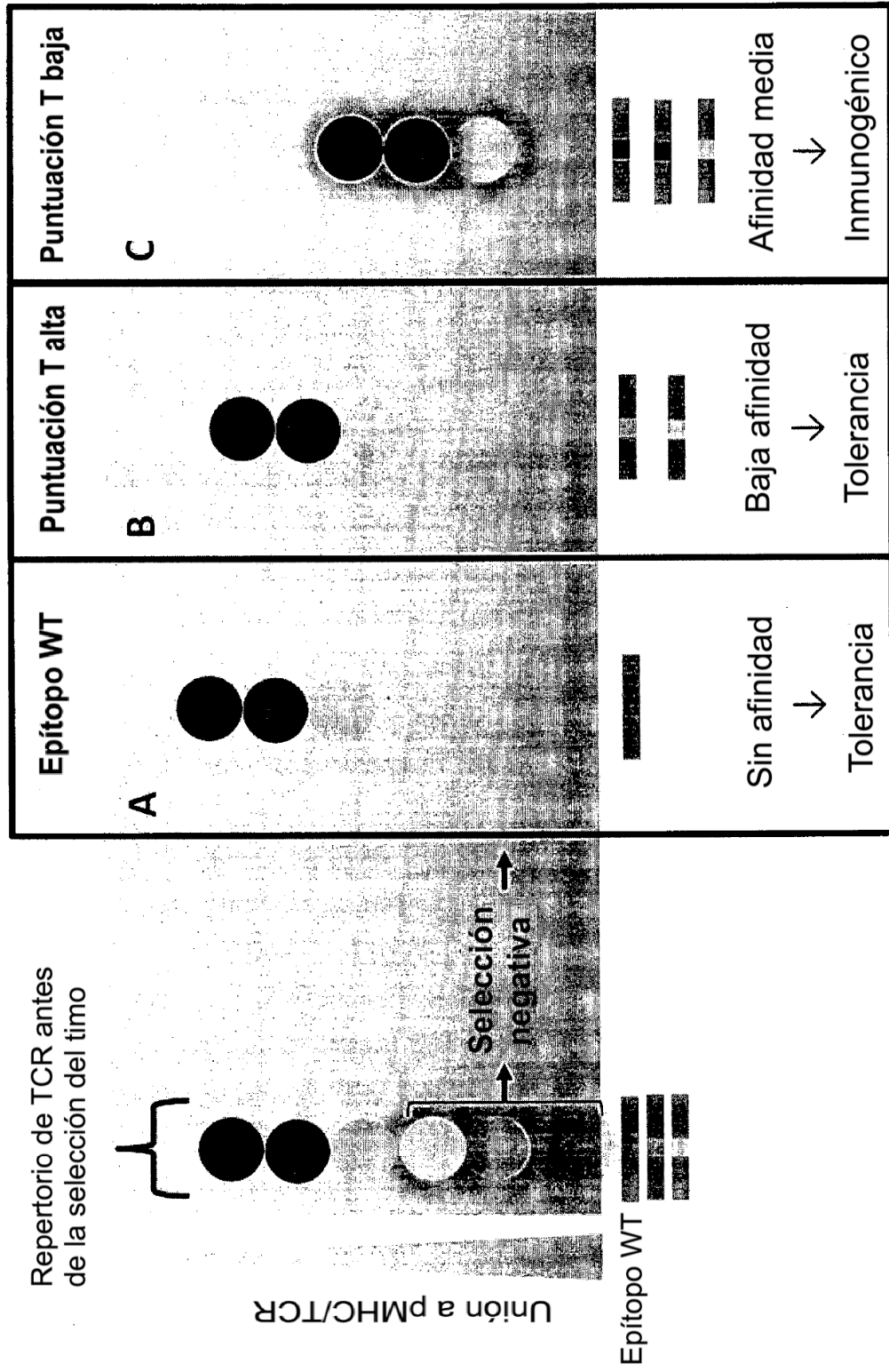


Fig. 6

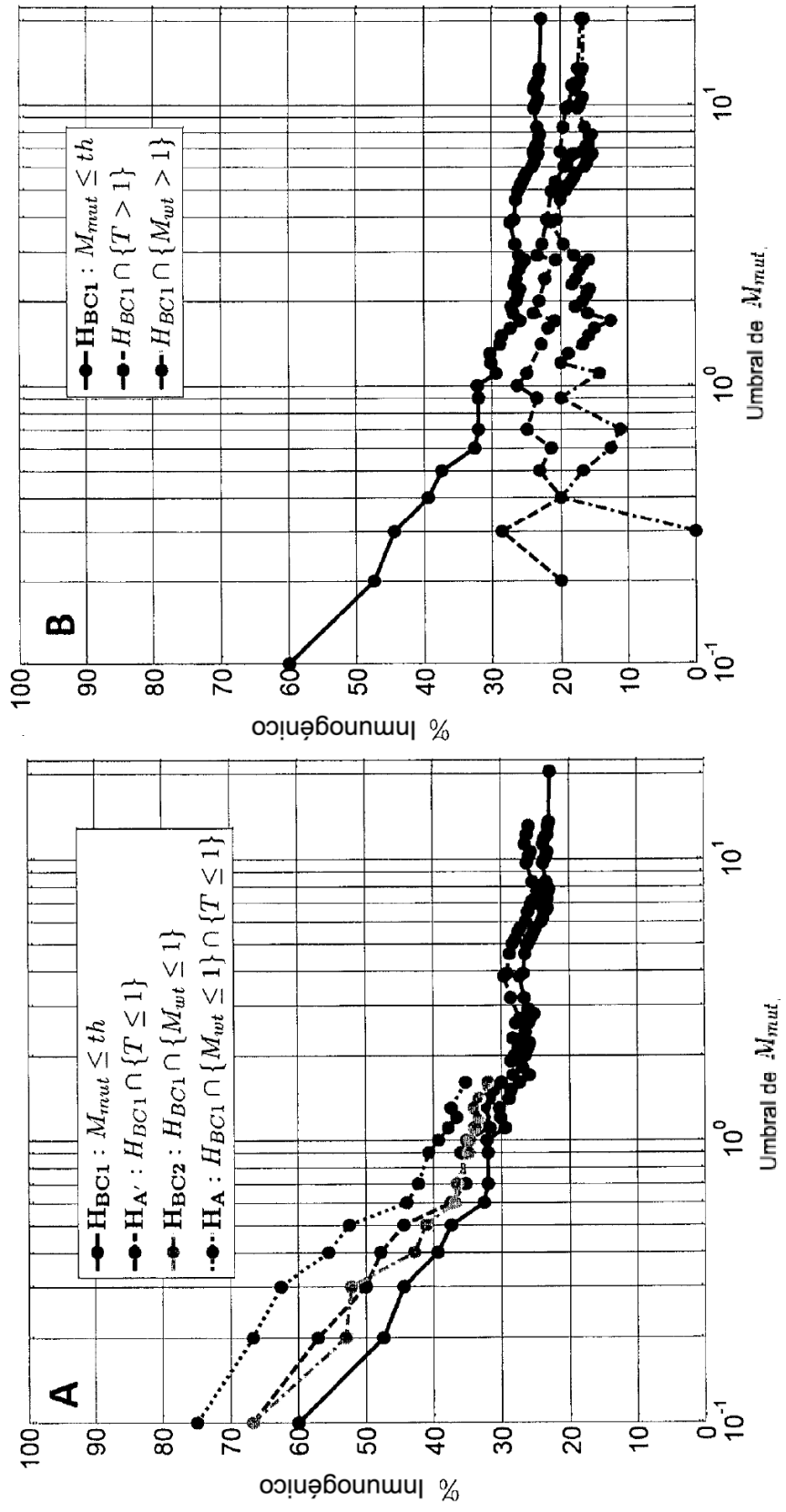


Fig. 7

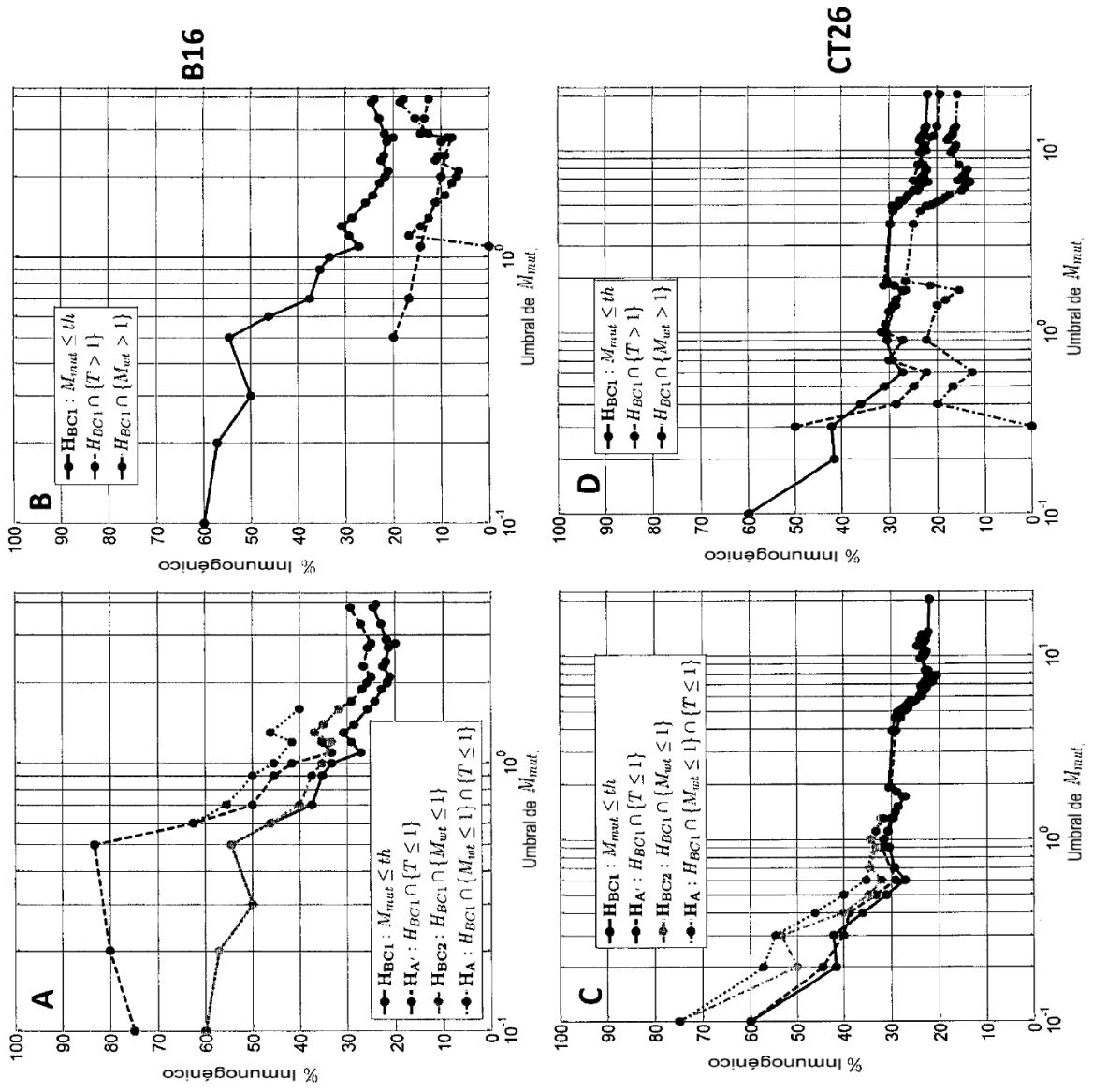


Fig. 8

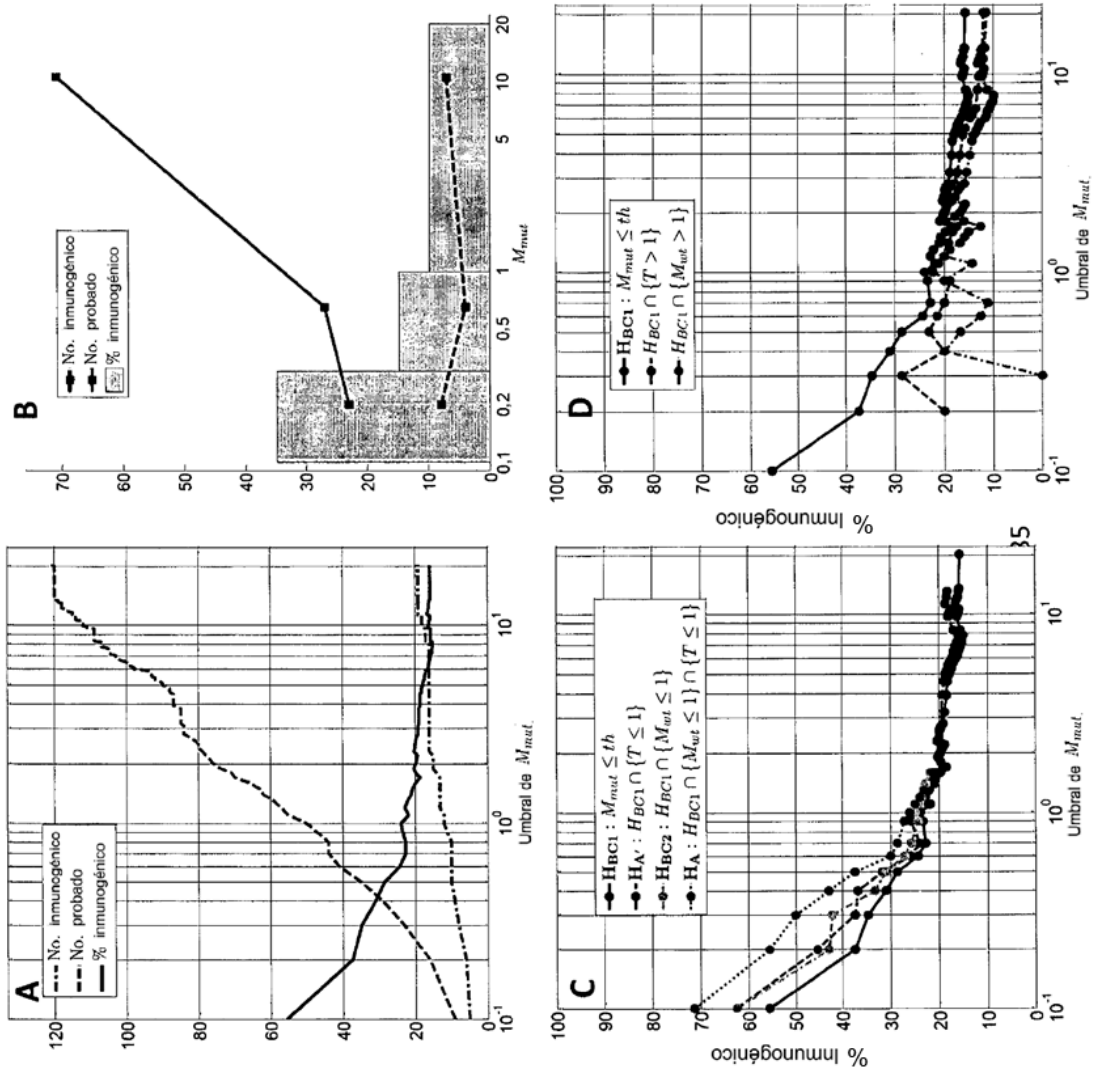


Fig. 9

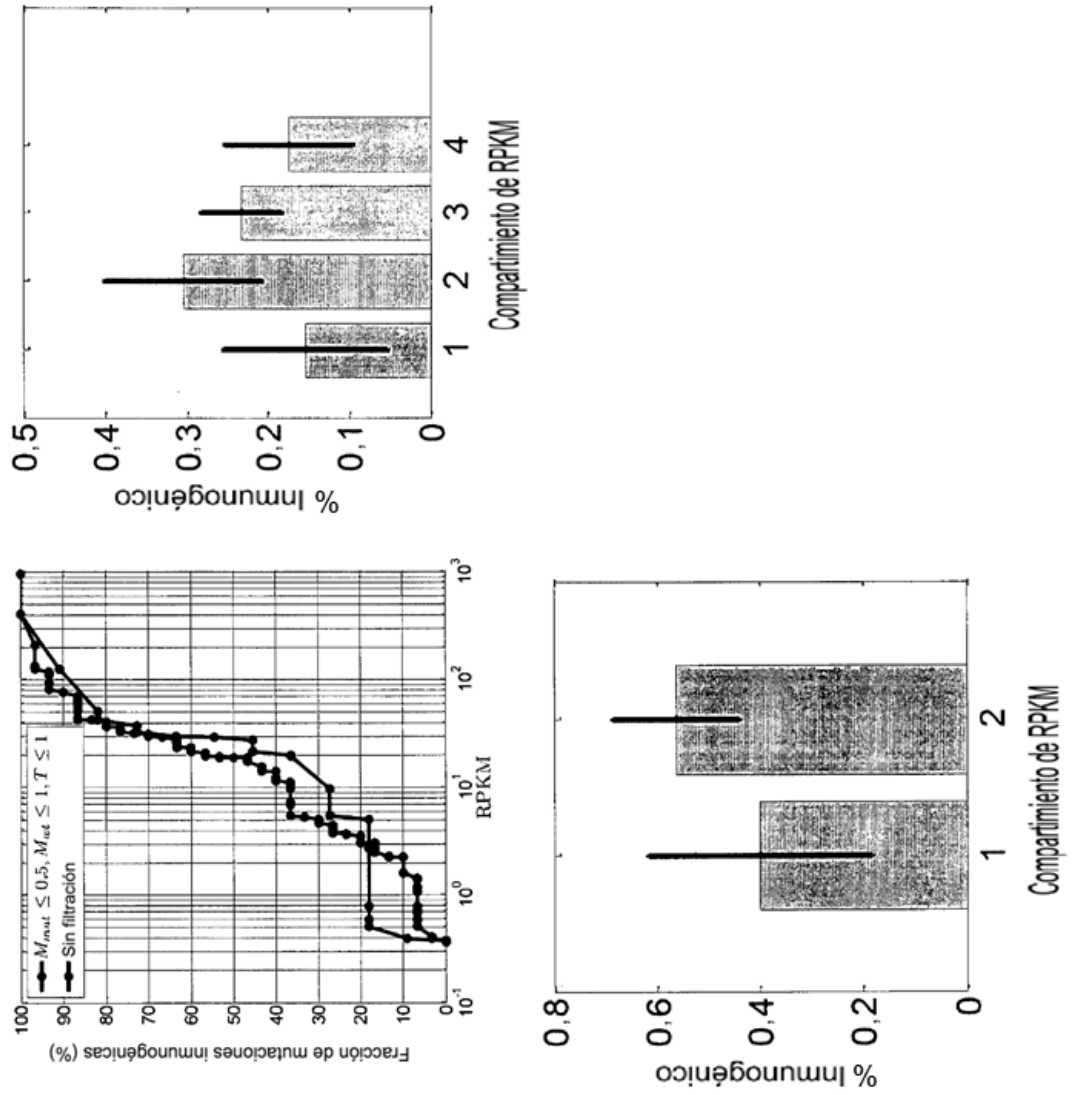


Fig. 10

Muestra ¹	Clase	Gen	Alelo de HLA	Posiciones de anclaje ²	Epítipo (WT) ³	Epítipo (MUT) ³	M _{mut}	M _{wr}	Mutación de la posición de anclaje	Inmunogénico
MZ7-MEL	(H ₂ UHL) clase II	SNRP116	HLA-A*03	1 2 3 4 5 6 7 8 9	K L DAV V AQE	K L DAV V AQ K	0,15	10,20	sí	sí
MZ7-MEL	(H ₂ UHL) clase II	ZNF335	HLA-B*38:01	1 2 3 4 5 6 7 8 9	Q H SAAPGPP	Q H SAAPG L	0,20	18,35	sí	desconocido
Zorn <i>et al.</i>	(H ₂ UHL) clase II	MYO1B	HLA-A*03	1 2 3 4 5 6 7 8 9	E I KNPKY K	K I KNPKY K	0,25	2,25	no	sí
Graf <i>et al.</i>	(H ₂ UHL) clase II	FLT3	HLA-A*01	1 2 3 4 5 6 7 8 9	YVDFREYED	YVDFREY Y	0,20	10,95	sí	sí
Robbins <i>et al.</i>	(H ₂ UHL) clase II	CTNNB1	HLA-A24	1 2 3 4 5 6 7 8 9	S Y LDSGIHS	S Y LDSGI H F	0,38	10,58	sí	sí

¹ Ver la Tabla 3

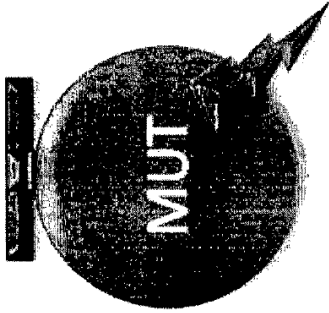
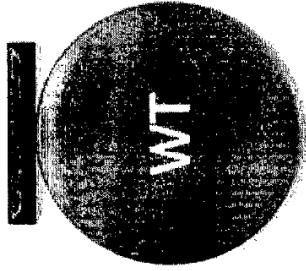
² Las posiciones de los anclajes (negrita) y los anclajes auxiliares (subrayadas) están indicadas

³ Residuos de anclaje conservado (negrita) y de anclaje auxiliar (subrayada) para el tipo dado de HLA y el epítipo están impresos en letras grandes. Residuos mutados resaltados en rojo.

Fig. 11

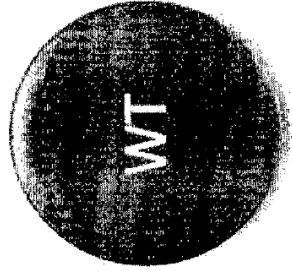
A Inmunogenicidad Clase 1 (H_A)

$$M_{WT} \leq \alpha \quad M_{MUT} \leq \beta, T \leq \tau$$



B Inmunogenicidad Clase 2 ($H_B UH_C$)

$$M_{WT} > \alpha^*$$



$$M_{MUT} \leq \beta$$



Fig. 12

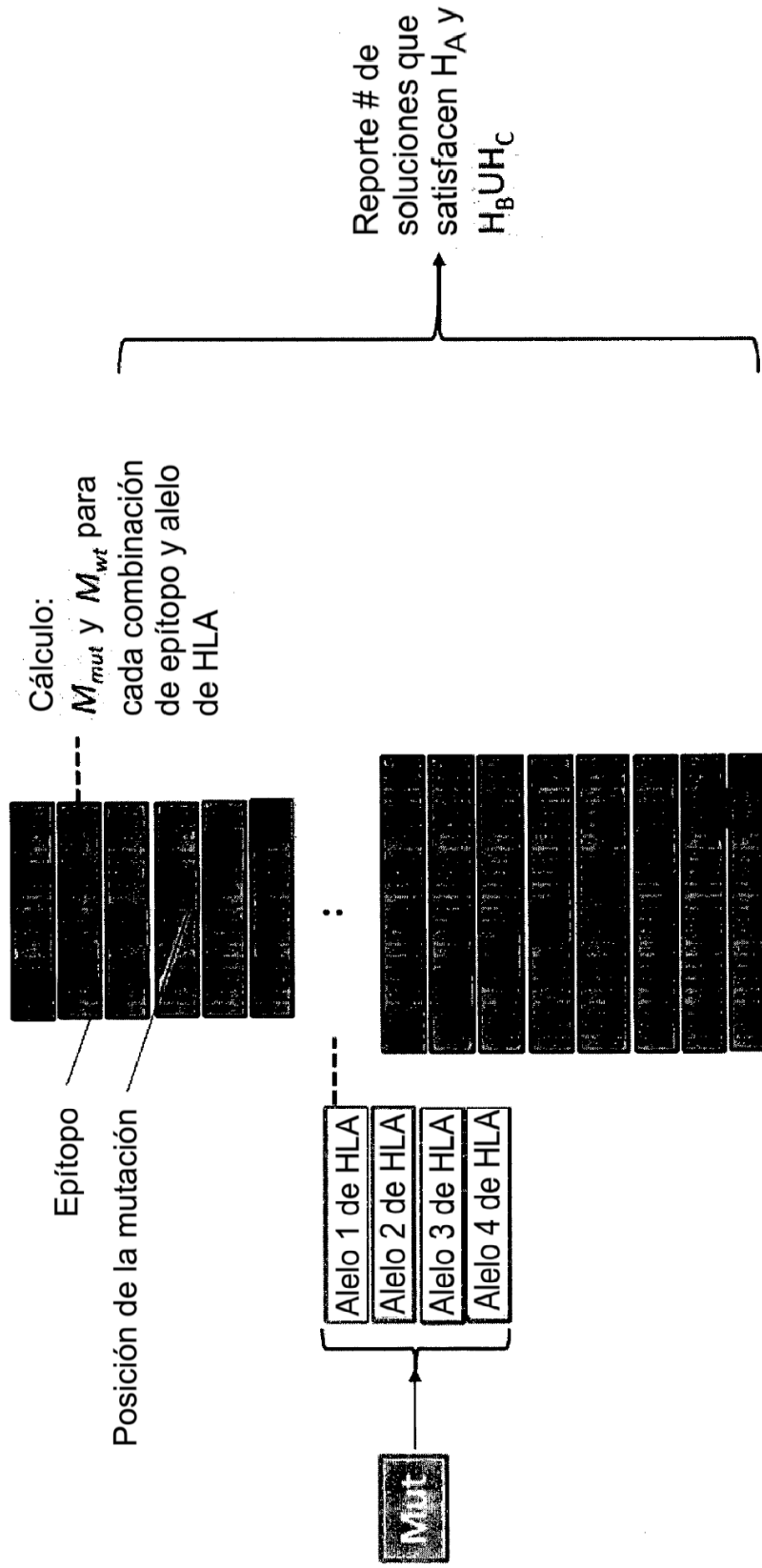


Fig. 13

