

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 746 233**

51 Int. Cl.:

A61K 39/00 (2006.01)

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **23.05.2012 PCT/EP2012/002209**

87 Fecha y número de publicación internacional: **29.11.2012 WO12159754**

96 Fecha de presentación y número de la solicitud europea: **23.05.2012 E 12723117 (3)**

97 Fecha y número de publicación de la concesión europea: **10.07.2019 EP 2714071**

54 Título: **Vacunas individualizadas para el cáncer**

30 Prioridad:

24.05.2011 WO PCT/EP2011/002576
02.01.2012 WO PCT/EP2012/000006

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.03.2020

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 (50.0%)

72 Inventor/es:

SAHIN, UGUR;
KREITER, SEBASTIAN;
DIKEN, MUSTAFA;
DIEKMANN, JAN;
KOSLOWSKI, MICHAEL;
BRITTEN, CEDRIK;
CASTLE, JOHN;
LÖWER, MARTIN;
RENARD, BERNHARD;
OMOKOKO, TANA y
DE GRAAF, JOHANNES HENDRIKUS

74 Agente/Representante:

SÁEZ MAESO, Ana

ES 2 746 233 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

Vacunas individualizadas para el cáncer

Campo técnico de la invención

5 La presente invención se refiere a una vacuna individualizada contra el cáncer para uso en un método de tratamiento de un paciente con cáncer.

Antecedentes de la invención

10 El cáncer es una causa principal de mortalidad, que representa 1 de cada 4 de todas las muertes. El tratamiento del cáncer se ha basado tradicionalmente en la ley de promedios, lo que funciona mejor para el mayor número de pacientes. Sin embargo, debido a la heterogeneidad molecular en el cáncer, a menudo menos del 25% de las personas tratadas se benefician de las terapias aprobadas. La medicina individualizada basada en el tratamiento a medida de los pacientes se considera una solución potencial para bajas eficacias y altos costos para la innovación en el desarrollo de fármacos.

15 La inmunoterapia específica de antígeno tiene como objetivo mejorar o inducir respuestas inmunes específicas en pacientes y se ha utilizado con éxito para controlar enfermedades cancerosas. Las células T desempeñan un papel central en la inmunidad celular en humanos y animales. El reconocimiento y la unión de un antígeno particular está mediado por los receptores de células T (TCR) expresados en la superficie de las células T. El receptor de células T (TCR) de una célula T puede interactuar con péptidos inmunogénicos (epítomos) unidos a moléculas del complejo mayor de histocompatibilidad (MHC) y presentados en la superficie de las células objetivo. La unión específica del TCR desencadena una cascada de señales dentro de la célula T que conduce a la proliferación y diferenciación en una célula T efectora madura.

20

25 La identificación de un número creciente de antígenos asociados a patógenos y tumores (TAA) condujo a una amplia colección de objetivos adecuadas para inmunoterapia. Las células que presentan péptidos inmunogénicos (epítomos) derivados de estos antígenos pueden dirigirse específicamente por estrategias de inmunización activa o pasiva. La inmunización activa puede tender a inducir y expandir las células T específicas de antígeno en el paciente, que son capaces de reconocer y matar específicamente las células enfermas. Se pueden usar diferentes formatos de antígeno para la vacunación tumoral, incluidas células cancerosas completas, proteínas, péptidos o vectores inmunizantes tales como ARN, ADN o vectores virales que se pueden aplicar directamente *in vivo* o *in vitro* pulsando DC después de la transferencia al paciente.

30 Los cánceres pueden surgir de la acumulación de mutaciones genómicas y cambios epigenéticos, de los cuales una fracción puede tener un papel causal. Además de los antígenos asociados a tumores, los cánceres humanos transportan en promedio 100-120 mutaciones no sinónimas, de las cuales muchas son blanco de las vacunas. Más del 95% de las mutaciones en un tumor son únicas y específicas del paciente (Weide et al., 2008: J. Immunother. 31, 180-188). El número de mutaciones somáticas que cambian proteínas, que pueden dar lugar a epítomos de células T específicas del tumor, está en el intervalo de 30 a 400. Se ha predicho *in silico* que hay 40 a 60 epítomos restringidos de HLA clase I por paciente derivados de mutaciones somáticas específicas del tumor (Azuma et al., 1993: Nature 366, 76-79). Además, nuevamente los epítomos restringidos de HLA de clase II inmunogénicos probablemente también resulten de mutaciones asociadas a tumores, sin embargo, su número aún se desconoce.

35

40 Notablemente, algunas mutaciones no sinónimas están causalmente involucradas en la transformación neoplásica, crucial para mantener el fenotipo oncogénico (mutaciones conductoras) y pueden representar un potencial "talón de Aquiles" de células cancerosas. Como tales las mutaciones no sinónimas no están sujetas a la tolerancia inmune central, pueden ser candidatos ideales para el desarrollo de vacunas individuales contra el cáncer. Las mutaciones encontradas en el tumor primario también pueden estar presentes en metástasis. Sin embargo, varios estudios demostraron que los tumores metastásicos de un paciente adquieren mutaciones genéticas adicionales durante la evolución tumoral individual que a menudo son clínicamente relevantes (Suzuki et al, 2007: Mol. Oncol. 1 (2), 172-180; Campbell et al., 2010: Nature 467 (7319), 1109-1113). Además, también las características moleculares de muchas metástasis se desvían significativamente de las de los tumores primarios.

45

El problema técnico subyacente a la presente invención es proporcionar una vacuna individualizada altamente efectiva contra el cáncer.

50 Rammensee et al., (Genomic and Personalized Medicine; 11 de noviembre de 2008; páginas 573-589) se relaciona con consideraciones básicas con respecto a las vacunas contra el cáncer.

55 La presente invención se basa en la identificación de mutaciones de cáncer específicas del paciente y el direccionamiento de una "firma" de la mutación de cáncer individual de un paciente. Específicamente, la presente invención que implica un enfoque de inmunoterapia individualizada basada en la secuenciación del genoma, preferiblemente el exoma o el transcriptoma tiene como objetivo atacar inmunoterapéuticamente múltiples mutaciones individuales en cáncer. La secuenciación usando la secuenciación de próxima generación (NGS) permite una identificación rápida y rentable de mutaciones de cáncer específicas del paciente.

La identificación de mutaciones puntuales no sinónimas que resultan en cambios de aminoácidos que se presentarán en las moléculas del complejo mayor de histocompatibilidad (MHC) del paciente proporciona nuevos epítomos (neoepítomos) que son específicos para el cáncer del paciente pero que no se encuentran en células normales del paciente. La recopilación de un conjunto de mutaciones de las células cancerosas, tales como las células tumorales en circulación (CTC) permite la provisión de una vacuna que induce una respuesta inmune potencialmente dirigida al tumor primario, incluso si contiene subpoblaciones genéticamente distintas, así como metástasis tumorales. Para la vacunación, dichos neoepítomos identificados de acuerdo con la presente solicitud se proporcionan en un paciente en forma de un polipéptido que comprende dichos neoepítomos y después del procesamiento y presentación apropiados por las moléculas del MHC, los neoepítomos se presentan al sistema inmunitario del paciente para estimulación de células T apropiadas.

Preferentemente, dicho polipéptido se proporciona en el paciente administrando ARN que codifica el polipéptido. Una estrategia en la que el ARN transcrito *in vitro* (ARN-IVT) se inyecta directamente en un paciente mediante diferentes rutas de inmunización se ha probado con éxito en varios modelos animales. El ARN puede traducirse en células transfectadas y la proteína expresada después del procesamiento se presenta en las moléculas del MHC en la superficie de las células para provocar una respuesta inmune.

Las ventajas de usar ARN como un tipo de terapia génica reversible incluyen la expresión transitoria y un carácter no transformante. El ARN no necesita ingresar al núcleo para poder expresarse y, además, no puede integrarse en el genoma del huésped, eliminando así el riesgo de oncogénesis. Las tasas de transfección alcanzables con ARN son relativamente altas. Además, las cantidades de proteína alcanzadas corresponden a las de la expresión fisiológica.

La justificación para el direccionamiento inmunoterapéutico de múltiples mutaciones individuales es que (i) estas mutaciones se expresan exclusivamente, (ii) se puede esperar que los epítomos mutados sean ideales para la inmunoterapia con células T ya que las células T que las reconocen no se han sometido a selección tímica, (iii) el escape inmune del tumor puede reducirse, por ejemplo al direccionar a "mutaciones conductoras" que son altamente relevantes para el fenotipo tumoral, y (iv) una respuesta inmune multiepitópica tiene una mayor probabilidad de resultar en un beneficio clínico mejorado.

Descripción de la invención

Sumario de la invención

En el presente documento se describen métodos eficientes para proporcionar vacunas de cáncer recombinante individualizadas que inducen una respuesta inmune eficiente y específica en un paciente con cáncer y que apuntan potencialmente al tumor primario así como a las metástasis tumorales. Las vacunas contra el cáncer para su uso de acuerdo con la invención cuando se administran a una paciente proporcionan una colección de epítomos presentados por el MHC específicos para el tumor del paciente adecuados para estimular, cebar y/o expandir células T dirigidas contra células que expresan antígenos de los cuales se derivan los epítomos presentados por el MHC. 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 antígenos expresados por el cáncer con MHC de clase I. Dado que una vacuna para uso de acuerdo con la presente invención se dirigirá a mutaciones específicas de cáncer, será específica para el tumor del paciente.

El alcance de la presente invención se define en las reivindicaciones adjuntas.

La presente invención se refiere a una vacuna individualizada contra el cáncer para su uso en un método de tratamiento de un paciente con cáncer, comprendiendo dicho método las etapas:

(a) proporcionar dicha vacuna individualizada contra el cáncer mediante un método que comprende las etapas:

(aa) identificar mutaciones somáticas específicas de cáncer en una muestra de tumor del paciente con cáncer para proporcionar una firma de mutación de cáncer del paciente; y

(ab) proporcionar una vacuna de ARN con la firma de mutación de cáncer obtenida en la etapa (aa), en la que la vacuna de ARN con la firma de mutación del paciente comprende ARN que codifica un polipéptido poliepitópico recombinante que comprende neoepítomos basados en mutación; y

(b) administrar dicha vacuna individualizada contra el cáncer al paciente.

En una realización, el método para proporcionar dicha vacuna individualizada contra el cáncer comprende las siguientes etapas:

i) proporcionar una muestra de tumor del paciente con cáncer y una muestra no tumoral que preferiblemente se deriva del paciente con cáncer;

ii) identificar las diferencias de secuencia entre el genoma, el exoma y/o el transcriptoma de la muestra tumoral y el genoma, el exoma y/o el transcriptoma de la muestra no tumoral;

iii) diseñar un polipéptido que comprende epítomos que incorporan las diferencias de secuencia determinadas en la etapa (ii);

iv) proporcionar ARN que codifica dicho polipéptido diseñado en la etapa (iii); y

v) proporcionar una vacuna que comprenda el ARN proporcionado en la etapa (iv).

5 De acuerdo con la invención, 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 las mutaciones somáticas específicas del cáncer o las diferencias de secuencia se determinan en una o más células tumorales en circulación (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 en circulación (CTC) o una muestra que contiene una o más células tumorales o cancerosas aisladas, tales como células tumorales en circulación (CTC).

10 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 tumoral.

15 De acuerdo con la invención, 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 porción de las mutaciones cancerosas presentes en una o más células cancerosas de un paciente. Por consiguiente, la presente invención 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. En general, el método para proporcionar dicha vacuna individualizada contra el cáncer proporciona la identificación de una serie de mutaciones que proporcionan un número suficiente de neoepítomos para ser incluidos en una vacuna. Una "mutación cancerosa" se refiere a una diferencia de secuencia entre el ácido nucleico contenido en una célula cancerosa y el ácido nucleico contenido en una célula normal.

20 Preferiblemente, las mutaciones identificadas de acuerdo con la presente invención son mutaciones no sinónimas, preferiblemente mutaciones no sinónimas de proteínas expresadas en un tumor o célula cancerosa.

25 En una realización, las mutaciones somáticas específicas del cáncer o las diferencias de secuencia se determinan en el genoma, preferiblemente el genoma completo, de una muestra de tumor. Por lo tanto, el método para proporcionar dicha vacuna individualizada contra el cáncer puede comprender identificar la firma de 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 del cáncer de todo el genoma.

30 En una realización, las mutaciones somáticas específicas de cáncer o las diferencias de secuencia se determinan en el exoma, preferiblemente todo el exoma, de una muestra de tumor. El exoma es 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 funcionalmente más 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). Por lo tanto, el método para proporcionar dicha vacuna individualizada contra el cáncer 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.

35 En una realización, las mutaciones somáticas específicas del cáncer o las diferencias de secuencia se determinan en el transcriptoma, preferiblemente el transcriptoma completo, de una muestra de tumor. El transcriptoma es el conjunto de todas las moléculas de ARN, incluidos ARNm, ARNr, ARNt y otros ARN no codificantes producidos en una célula o en una población de células. En el contexto de la presente invención, 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 determinado momento. Por lo tanto, el método para proporcionar dicha vacuna individualizada contra el cáncer puede comprender identificar la firma de mutación de 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.

40 En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer o identificar diferencias de secuencia comprende la secuenciación de células individuales 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, el método para proporcionar

dicha vacuna individualizada contra el cáncer puede comprender identificar una firma de mutación de cáncer de dichas una o más células cancerosas. En una realización, las células cancerosas son células tumorales en circulación. Las células cancerosas, como las células tumorales en circulación, pueden aislarse antes de la secuenciación de células individuales.

- 5 En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer o identificar diferencias de secuencia implica usar secuenciación de próxima generación (NGS).

En una realización, la etapa de identificar mutaciones somáticas específicas de cáncer o identificar diferencias de secuencia comprende secuenciar ADN y/o ARN genómico de la muestra tumoral.

- 10 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 información de secuencia obtenida de secuenciación de ácido nucleico 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 genómico de línea germinal normal se obtiene de células mononucleares de sangre periférica (PBMC)

- 15 Una vacuna para uso de acuerdo con la presente invención se refiere a una vacuna que, cuando se administra a un paciente, proporciona preferiblemente una colección de epítomos presentados por el MHC, 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 presentados por el MHC, incorporando cambios de secuencia basados en las mutaciones identificadas o diferencias de secuencia. Dichos epítomos presentados por el MHC que incorporan cambios de secuencia basados en las mutaciones identificadas o las diferencias de secuencia también se denominan en el presente documento "neoepítomos". 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, el tumor del paciente, preferiblemente el tumor primario y las metástasis tumorales, que expresan antígenos a partir de los cuales los epítomos presentados por el MHC se derivan y presentan los mismos epítomos en la superficie de las células tumorales.

- 25 El método para proporcionar dicha vacuna individualizada contra el cáncer puede comprender la inclusión arbitraria de un número suficiente de neoepítomos (preferiblemente en forma de un ácido nucleico codificador) en una vacuna o puede comprender la etapa adicional de determinar la posibilidad de usar las mutaciones identificadas en los epítomos para la vacunación contra el cáncer. Por lo tanto, las etapas adicionales pueden incluir uno o más de los siguientes: (i) evaluar si los cambios de secuencia se encuentran en epítomos presentados por el MHC conocidos o predichos, (ii) pruebas *in vitro* y/o *in silico* si los cambios de secuencia se encuentran en los epítomos presentados por el MHC, por ejemplo, probar si los cambios de secuencia son parte de secuencias de péptidos que se procesan y/o presentan como epítomos presentados por el MHC, y (iii) probar *in vitro* si los epítomos mutados previstos, en particular cuando están presentes en su contexto de secuencia natural, por ejemplo, cuando están flanqueados por secuencias de aminoácidos que también flanquean dichos epítomos en la proteína natural, y cuando se expresan en células presentadoras de antígeno pueden estimular las células T del paciente que tienen la especificidad deseada. Dichas secuencias flanqueantes pueden comprender cada una 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 epítopo en el terminal N y/o en el terminal C.

- 40 Las mutaciones o las diferencias de secuencia determinadas de acuerdo con la invención pueden clasificarse por su posibilidad de uso como epítomos para la vacunación contra el cáncer. Por lo tanto, en un aspecto, el método para proporcionar dicha vacuna individualizada contra el cáncer comprende un proceso analítico manual o basado en ordenador en el que las mutaciones identificadas se analizan y seleccionan para su posible uso en la vacuna respectiva que se proporcionará. En una realización preferida, dicho proceso analítico es un proceso basado en un algoritmo computacional. Preferiblemente, dicho proceso analítico comprende uno o más, preferiblemente todas las siguientes etapas:

- identificación de mutaciones expresadas que modifican proteínas, por ejemplo, analizando transcritos;
- identificar mutaciones que son potencialmente inmunogénicas, es decir, comparando los datos obtenidos con los conjuntos de datos disponibles de epítomos inmunogénicos confirmados, por ejemplo, aquellos contenidos en bases de datos públicas de epítomos inmunes tales como, por ejemplo, el RECURSO DE BASE DE DATOS Y ANÁLISIS DE EPÍTOPOS INMUNITARIOS en <http://www.immunoepitope.org>

- 50 La etapa de identificar mutaciones que son potencialmente inmunogénicas puede comprender determinar y/o clasificar epítomos de acuerdo con una predicción de su capacidad de unión al MHC, preferiblemente capacidad de unión al MHC de clase I.

- 55 En otra realización de la invención, los epítomos pueden seleccionarse y/o clasificarse usando parámetros adicionales tales como impacto proteico, expresión génica asociada, unicidad de secuencia, probabilidad de presentación predicha y asociación con oncogenes.

Los análisis múltiples de CTC también permiten la selección y priorización de mutaciones. Por ejemplo, una mutación

que se encuentra en una porción más grande de CTC puede tener mayor prioridad que una mutación que se encuentra en una porción menor de CTC.

La colección de neoepítomos basados en mutaciones identificados de acuerdo con la invención y proporcionados por una vacuna para su uso de acuerdo con la invención está presente en forma de ARN que codifica un polipéptido que comprende dichos neoepítomos (polipéptido poliepitópico). Además, los neoepítomos pueden estar presentes en el polipéptido en forma de una secuencia de vacuna, es decir, presente en su contexto de secuencia natural, por ejemplo, flanqueado por secuencias de aminoácidos que también flanquean dichos epítomos en la proteína natural. Dichas secuencias flanqueantes pueden comprender cada una 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 en el terminal N y/o el 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, las secuencias de los neoepítomos y/o la vacuna están alineados en el polipéptido de un extremo al otro.

En una realización, las secuencias de los neoepítomos y/o la vacuna están separadas por enlazadores, en particular enlazadores neutros. El término "enlazador" de acuerdo con la invención se refiere a un péptido añadido entre dos dominios peptídicos tales como secuencias de epítomos o de la vacuna 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 no debe tener o solo tener pocos elementos de secuencia inmunogénica. Los enlazadores preferiblemente no deberían crear neoepítomos no endógenos como los generados a partir de la sutura de unión entre neoepítomos adyacentes, lo que podría generar reacciones inmunes no deseadas. Por lo tanto, la vacuna poliepitópica debería contener preferiblemente secuencias enlazadoras que puedan reducir el número de epítomos de empalme 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 proteasomal y, por lo tanto, el uso de secuencias enlazadoras ricas en glicina actúa para minimizar el número de péptidos contenidos en el enlazador que pueden ser procesados por el proteasoma. Además, se observó que la glicina inhibía una fuerte unión en las posiciones de ranura de unión al MHC (Abastado et al., J. Immunol. 151 (7), 3569-75, 1993). Schlessinger et al., (Proteins, 61 (1), 115-26, 2005) encontraron que los aminoácidos glicina y serina incluidos en una secuencia de aminoácidos dan como resultado una proteína más flexible que el proteasoma traduce y procesa de manera más eficiente, lo que permite un mejor acceso a los neoepítomos codificados. El enlazador puede comprender cada uno 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 de glicina y/o serina. Preferiblemente, al menos 50%, al menos 60%, al menos 70%, al menos 80%, al menos 90% o al menos 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 $(GGS)_a(GSS)_b(GGG)_c(SSG)_d(GSG)_e$ en donde a, b, c, d y e es independientemente un número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20 y en donde 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 GSGGGGGSG.

En otra realización de la presente invención, la colección de neoepítomos basados en mutación identificados de acuerdo con la invención y proporcionados por una vacuna para su uso de acuerdo con la invención está presente preferiblemente en forma de una colección de ARN que codifican polipéptidos que comprenden dichos neoepítomos en diferentes polipéptidos, en donde dichos polipéptidos comprenden cada uno, uno o más neoepítomos, que también pueden superponerse.

Un polipéptido poliepitópico de acuerdo con la presente invención se administra a un paciente en forma de ARN tal como ARN transcrito *in vitro* o sintético, que puede expresarse en células de un paciente tal como células presentadoras de antígeno para producir el polipéptido. La presente invención prevé la administración de uno o más polipéptidos multiepitópicos que para el propósito de la presente invención están comprendidos por el término "polipéptido poliepitópico", en forma de ARN tal como ARN transcrito *in vitro* o sintético, que puede expresarse en células de un paciente tal como las células presentadoras de antígeno para producir uno o más polipéptidos. En el caso de una administración de más de un polipéptido multiepitópico, los neoepítomos proporcionados por los diferentes polipéptidos multiepitópicos pueden ser diferentes o parcialmente superpuestos. Una vez presente en las células de un paciente, tal como las células presentadoras de antígeno, el polipéptido descrito en este documento se procesa para producir los neoepítomos identificados de acuerdo con la invención. La administración de una vacuna para su uso de acuerdo con la invención puede proporcionar epítomos presentados por el MHC de clase II que son capaces de provocar una respuesta de células T auxiliares 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 para uso de acuerdo con la invención puede proporcionar epítomos presentados por el MHC de clase I que son capaces de provocar una respuesta de células T CD8+ contra células que expresan antígenos de los cuales se derivan los epítomos presentados por el MHC. Además, la administración de una vacuna para su uso de acuerdo con la invención puede proporcionar uno o más neoepítomos (incluidos los neoepítomos conocidos y neoepítomos identificados de acuerdo con la invención), así como uno o más epítomos que no contienen mutaciones somáticas específicas de cáncer pero que son expresados por células cancerosas y preferiblemente que inducen una respuesta inmune contra las células cancerosas,

5 preferiblemente una respuesta inmune específica del cáncer. En una realización, la administración de una vacuna para uso de acuerdo con la invención proporciona neoepítomos que son epítomos presentados por el MHC de clase II y/o son capaces de provocar una respuesta de células T auxiliares CD4+ contra células que expresan antígenos de los cuales 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 provocar una respuesta de células T CD8+ contra células que expresan antígenos de los cuales 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 de cáncer tienen un efecto sinérgico en el tratamiento del cáncer. Preferiblemente, una vacuna para usar de acuerdo con la invención es útil para la estimulación poliepitópica de respuestas de células T citotóxicas y/o auxiliares.

15 También se describe en el presente documento una vacuna que se puede obtener mediante el método descrito en el presente documento. Por consiguiente, la presente descripción también se refiere a una vacuna que comprende un polipéptido recombinante que comprende neoepítomos basados en mutación, dichos neoepítomos resultantes de mutaciones somáticas específicas de cáncer en una muestra de tumor de un paciente con cáncer, o un ácido nucleico que codifica dicho polipéptido. Tal polipéptido recombinante también puede incluir epítomos que no incluyen mutaciones somáticas específicas de cáncer como se discutió anteriormente. Las realizaciones preferidas de dicha vacuna son como se describieron anteriormente en el contexto del método divulgado en el presente documento.

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

En el presente documento también se describe un método para inducir una respuesta inmune en un paciente, que comprende administrar al paciente una vacuna descrita en el presente documento.

25 También se describe en el presente documento un método para tratar a un paciente con cáncer que comprende administrar la vacuna descrita en el presente documento al paciente.

La divulgación también 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 prevención del cáncer.

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

30 En el presente documento también se describe un método para determinar una tasa de descubrimiento falsa basada en datos de secuenciación de próxima generación, incluyendo dicho método:

- tomar una primera muestra de material genético de un animal o humano;
- tomar una segunda muestra de material genético de un animal o humano;
- tomar una primera muestra de material genético de células tumorales;
- 35 tomar una segunda muestra de material genético de dichas células tumorales;
- determinar una comparación del tumor de cobertura común contando todas las bases del genoma de referencia que están incluidas tanto en el tumor como en al menos una de dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano;

40 determinar una cobertura común igual frente a la misma comparación contando todas las bases del genoma de referencia que están cubiertas por dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano;

dividir dicha comparación del tumor de cobertura común por dicha cobertura común similar frente a la misma comparación para formar una normalización;

45 determinar una tasa de descubrimiento falso dividiendo 1) el número de variaciones de un solo nucleótido con un puntaje de calidad mayor que Q en una comparación de dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano, por 2) el número de variaciones de un solo nucleótido con un puntaje de calidad mayor que Q en una comparación de dicha primera muestra de material genético de dichas células tumorales y dicha segunda muestra de material genético de dichas células tumorales y 3) multiplicar el resultado por dicha normalización .

50 En una realización, dicho material genético es un ADN.

En una realización, Q está determinado por:

- el establecimiento de un conjunto de propiedades de calidad $S = (s_1, \dots, s_n)$ en donde S es preferible con $T = (t_1, \dots, t_n)$, denotado por $S > T$, cuando $s_i > t_i$ para todo $i = 1, \dots, n$;
- la definición de una tasa de descubrimiento falsa intermedia dividiendo 1) el número de variaciones de un solo nucleótido con un puntaje de calidad $S > T$ en una comparación de dicha primera muestra de ADN de un animal o humano y dicha segunda muestra de ADN de un animal o humano, por 2) el número de variaciones de un solo nucleótido con un puntaje de calidad $S > T$ en una comparación de dicha primera muestra de ADN de dichas células tumorales y dicha segunda muestra de ADN de dichas células tumorales y 3) la multiplicación del resultado por dicha normalización,
- determinar el intervalo de valores para cada propiedad para m mutaciones con n propiedades de calidad cada una;
- el muestreo hasta valores p fuera de dicho intervalo de valores;
- la creación de cada combinación posible de valores de calidad muestreados que resultan en p^n puntos de datos;
- la utilización de una muestra aleatoria de dichos puntos de datos como un predictor para el entrenamiento forestal aleatorio;
- la utilización del correspondiente valor de tasa de descubrimiento falsa intermedia como respuesta para dicho entrenamiento forestal aleatorio,
- en el que la puntuación de regresión resultante de dicho entrenamiento forestal aleatorio es Q .
- En una realización, dicha segunda muestra de ADN de un animal o humano es alogénica a dicha primera muestra de ADN de un animal o humano. En una realización, dicha segunda muestra de ADN de un animal o humano es autóloga a dicha primera muestra de ADN de un animal o humano. En una realización, dicha segunda muestra de ADN de un animal o humano es xenogénica a dicha primera muestra de ADN de un animal o humano.
- En una realización, dicho material genético es un ARN.
- En una realización, Q está determinado por:
- el establecimiento de un conjunto de propiedades de calidad $S = (s_1, \dots, s_n)$ en donde S es preferible con $T = (t_1, \dots, t_n)$, denotado por $S > T$, cuando $s_i > t_i$ para todo $i = 1, \dots, n$;
- la definición de una tasa de descubrimiento falsa intermedia dividiendo 1) el número de variaciones de un solo nucleótido con un puntaje de calidad $S > T$ en una comparación de dicha primera muestra de ADN de un animal o humano y dicha segunda muestra de ADN de un animal o humano, por 2) el número de variaciones de un solo nucleótido con un puntaje de calidad $S > T$ en una comparación de dicha primera muestra de ADN de dichas células tumorales y dicha segunda muestra de ADN de dichas células tumorales y 3) la multiplicación del resultado por dicha normalización,
- determinar el intervalo de valores para cada propiedad para m mutaciones con n propiedades de calidad cada una;
- el muestreo hasta valores p fuera de dicho intervalo de valores;
- la creación de cada combinación posible de valores de calidad muestreados que resultan en p^n puntos de datos;
- la utilización de una muestra aleatoria de dichos puntos de datos como un predictor para el entrenamiento forestal aleatorio;
- la utilización del correspondiente valor de tasa de descubrimiento falsa intermedia como respuesta para dicho entrenamiento forestal aleatorio,
- en el que la puntuación de regresión resultante de dicho entrenamiento forestal aleatorio es Q .
- En una realización, dicha segunda muestra de ARN de un animal o humano es alogénica a dicha primera muestra de ARN de un animal o humano. En una realización, dicha segunda muestra de ARN de un animal o humano es autóloga a dicha primera muestra de ARN de un animal o humano. En una realización, dicha segunda muestra de ARN de un animal o humano es xenogénica a dicha primera muestra de ARN de un animal o humano.
- En una realización, dicha tasa de descubrimiento falsa se usa para hacer una formulación de vacuna. En una realización, dicha vacuna se puede administrar por vía intravenosa. En una realización, dicha vacuna se puede administrar por vía dérmica. En una realización, dicha vacuna se puede administrar muscularmente. En una realización, dicha vacuna se puede administrar por vía subcutánea. En una realización, dicha vacuna está diseñada para un paciente específico.
- En una realización, una de dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano es de dicho paciente específico.

En una realización, dicha etapa de determinar una comparación de tumor de cobertura común contando todas las bases del genoma de referencia que está incluido tanto en el tumor como en al menos una de dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano utiliza un sistema automatizado para contar todas las bases.

- 5 En una realización, dicha etapa de determinar una cobertura común igual frente a la misma comparación contando todas las bases del genoma de referencia que están cubiertas por dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano utiliza dicho sistema automatizado.

- 10 En una realización, dicha etapa de dividir dicha comparación del tumor de cobertura común por dicha cobertura común similar frente a la misma comparación para formar una normalización utiliza dicho sistema automatizado.

- 15 En una realización, dicha etapa de determinar una tasa de descubrimiento falsa dividiendo 1) el número de variaciones de nucleótidos individuales con un puntaje de calidad mayor que Q en una comparación de dicha primera muestra de material genético de un animal o humano y dicha segunda muestra de material genético de un animal o humano, por 2) el número de variaciones de un solo nucleótido con un puntaje de calidad mayor que Q en una comparación de dicha primera muestra de material genético de dichas células tumorales y dicha segunda muestra de material genético de dicho células tumorales y 3) multiplicar el resultado por dicha normalización utiliza dicho sistema automatizado.

En el presente documento también se describe un método para determinar una curva operativa estimada del receptor (ROC), incluyendo dicho método:

- 20 recibir un conjunto de datos de mutaciones, cada mutación asociada con una tasa de descubrimiento falsa (FDR); y para cada mutación:

determinar una tasa positiva verdadera (TPR) restando dicha FDR de uno; y

determinar una tasa positiva falsa (FPR) estableciendo dicho FPR igual a dicha FDR; y

formando una ROC estimada graficando, para cada mutación, un punto en los valores acumulativos de TPR y FPR hasta dicha mutación, dividido por la suma de todos los valores de TPR y FPR.

- 25 Otras características y ventajas de la presente invención serán evidentes a partir de la siguiente descripción detallada y las reivindicaciones.

Descripción detallada de la invención

- 30 Aunque la presente invención se describe en detalle a continuación, debe entenderse que esta invención no se limita a las metodologías, protocolos y reactivos particulares descritos aquí, ya que pueden variar. También debe entenderse que la terminología utilizada en el presente documento tiene el propósito de describir realizaciones particulares solamente, y no pretende limitar el alcance de la presente invención, que estará limitado solo por las reivindicaciones adjuntas. A menos que se defina lo contrario, todos los términos técnicos y científicos utilizados en este documento tienen los mismos significados que los entendidos comúnmente por un experto en la materia.

- 35 A continuación, se describirán los elementos de la presente invención. 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 y las realizaciones preferidas descritas de manera diversa no deben interpretarse como limitantes de la presente invención solo a las realizaciones descritas explícitamente. Debe entenderse que esta descripción soporta y abarca realizaciones que combinan las realizaciones descritas explícitamente con cualquier número de los elementos divulgados y/o preferidos. Además, cualquiera de las permutaciones y combinaciones de todos los elementos descritos en esta solicitud deben considerarse revelados por la descripción de la presente solicitud a menos que el contexto indique lo contrario. Por ejemplo, si en una realización preferida, el ARN comprende una cola de poli(A) que consta de 120 nucleótidos y en otra realización preferida la molécula de ARN comprende un análogo de la caperuza de 5', entonces, en una realización preferida, el ARN comprende la cola poli(A) que consta de 120 nucleótidos y la caperuza de 5'.

- 45 Preferentemente, los términos utilizados 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.

- 50 La práctica de la presente invención 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 este campo (véase, por ejemplo, Molecular Cloning: A Laboratory Manual, 2ª edición, J. Sambrook et al. Eds., Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

A lo largo de esta memoria descriptiva y las reivindicaciones que siguen, a menos que el contexto requiera lo contrario, se entenderá que la palabra "comprender", y variaciones tales como "comprende" y "que comprende", implican la

inclusión de un miembro establecido, 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 puede ser otro miembro, número entero o etapa o grupo de miembros, números enteros o etapas pueden excluirse, es decir, la materia consiste en la inclusión de un miembro establecido, número entero o etapa o grupo de miembros, números enteros o etapas. Los términos "un" y "uno, una" y "el, la" y referencias similares utilizadas en el contexto de la descripción de la invención (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 el presente documento o sea claramente contradicho por el contexto. La mención de intervalos de valores en el presente documento solo pretende servir como un método abreviado para referirse individualmente a cada valor separado que cae dentro del intervalo. A menos que se indique lo contrario en este documento, cada valor individual se incorpora a la memoria descriptiva como si se mencionara individualmente en este documento.

Todos los métodos descritos en el presente documento pueden realizarse en cualquier orden adecuado a menos que se indique lo contrario en el presente documento o que el contexto lo contradiga claramente. El uso de todos y cada uno de los ejemplos, o ejemplo de expresión (por ejemplo, "tal como"), provisto en este documento tiene la intención de ilustrar mejor la invención y no plantea una limitación en el alcance de la invención reivindicada. Ninguna expresión en la memoria descriptiva debe interpretarse como indicativo de un elemento no reivindicado esencial para la práctica de la invención.

Se citan varios documentos a lo largo del texto de esta memoria descriptiva. Nada en este documento debe interpretarse como una admisión de que la invención no tiene derecho a anteceder dicha divulgación en virtud de la invención anterior.

La vacuna para uso de acuerdo con la invención es una vacuna recombinante.

El término "recombinante" en el contexto de la presente invención significa "hecho mediante ingeniería genética". Preferiblemente, una "entidad recombinante" tal como un polipéptido recombinante en el contexto de la presente invención no es natural, y preferiblemente es el resultado de una combinación de entidades tales como secuencias de aminoácidos o ácidos nucleicos que no están combinadas en la naturaleza. Por ejemplo, un polipéptido recombinante en el contexto de la presente invención puede contener varias secuencias de aminoácidos tales como neopéptidos o secuencias de vacunas derivadas de diferentes proteínas o diferentes porciones de la misma proteína fusionadas, por ejemplo, mediante enlaces peptídicos o enlazadores apropiados.

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

De acuerdo con la invención, 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 individualizada contra el cáncer" se refiere a un paciente en particular con cáncer y significa que una vacuna contra el cáncer se adapta a las necesidades o circunstancias especiales de un paciente individual con cáncer.

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

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

Una "respuesta inmune celular", una "respuesta celular", una "respuesta celular contra un antígeno" o un término similar pretende incluir una respuesta celular dirigida a células caracterizadas por la presentación de un antígeno con MHC de clase I o clase II. La respuesta celular se relaciona con células llamadas células T o linfocitos T que actúan como "auxiliares" o "asesinos". Las células T auxiliares (también llamadas células T CD4+) juegan un papel central al regular la respuesta inmune y las células asesinas (también llamadas 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 invención implica la estimulación de una respuesta CTL antitumoral contra células tumorales que expresan uno o más antígenos expresados en tumores y que presentan preferiblemente dichos antígenos expresados en tumores con MHC de clase I.

Un "antígeno" de acuerdo con la invención cubre cualquier sustancia que provocará una respuesta inmune. En particular, un "antígeno" se refiere a cualquier sustancia, preferiblemente un péptido o proteína, que reacciona específicamente con anticuerpos o linfocitos T (células T). De acuerdo con la presente invención, el término "antígeno" comprende cualquier molécula que comprenda al menos un epítipo. Preferiblemente, un antígeno en el contexto de la presente invención es una molécula que, opcionalmente después del procesamiento, induce una reacción inmune, que es preferiblemente específica para el antígeno (incluidas las células que expresan el antígeno). De acuerdo con la presente invención, se puede usar cualquier antígeno adecuado, que sea candidato para una reacción inmune, en donde la reacción inmune es preferiblemente una reacción inmune celular. En el contexto de las realizaciones de la presente invención, el antígeno es presentado 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, que da como resultado una reacción inmune contra el antígeno. Un antígeno es preferiblemente un producto que corresponde o se deriva de un antígeno natural. Tales antígenos naturales incluyen antígenos tumorales.

En una realización preferida, el antígeno es un antígeno tumoral, 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 producen principalmente intracelularmente o como antígenos de superficie de células tumorales. Por ejemplo, los antígenos tumorales incluyen el antígeno carcinoembrionario, α 1-fetoproteína, isoformina y sulfoglicoproteína fetal, α 2-H-ferroproteína y γ -fetoproteína. De acuerdo con la presente invención, un antígeno tumoral comprende preferiblemente cualquier antígeno que se expresa y opcionalmente es característico con respecto al tipo y/o nivel de expresión para tumores o cánceres, así como para células tumorales o cancerosas. En una realización, el término "antígeno tumoral" o "antígeno asociado a tumor" se refiere a proteínas que están en condiciones normales expresadas específicamente en un número limitado de tejidos y/u órganos o en etapas de desarrollo específicas, por ejemplo, el antígeno tumoral puede estar en 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 se expresan de manera aberrante en uno o más tejidos tumorales o 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 invención incluyen, por ejemplo, antígenos de diferenciación, preferiblemente antígenos de diferenciación específicos de tipo celular, es decir, proteínas que están en condiciones normales expresadas específicamente en un determinado tipo celular en una determinada etapa de diferenciación, antígenos de cáncer/testículos, 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 invención, el antígeno tumoral que es expresado por una célula cancerosa en un sujeto, por ejemplo, un paciente que padece una enfermedad cancerosa, es preferiblemente una proteína propia en dicho sujeto. En realizaciones preferidas, el antígeno tumoral en el contexto de la presente invención se expresa en condiciones normales específicamente en un tejido u órgano que no es esencial, es decir, tejidos u órganos que cuando están dañados por el sistema inmune no conducen a la muerte del sujeto, o en órganos o estructuras del cuerpo a los que el sistema inmunitario no tiene acceso o apenas es accesible.

De acuerdo con la invención, los términos "antígeno tumoral", "antígeno expresado por el tumor", "antígeno canceroso" y "antígeno expresado por el cáncer" son equivalentes y se usan indistintamente en el presente documento. El término "inmunogenicidad" se refiere a la efectividad relativa de un antígeno para inducir una reacción inmune.

Un "péptido antigénico" de acuerdo con la invención 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 presentación del antígeno como las células enfermas, en particular las células cancerosas. Preferiblemente, un péptido antigénico 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 (CTL) sensible a antígeno. Preferiblemente, los péptidos antigénicos son péptidos presentados por el MHC de clase I y/o clase II o pueden procesarse para producir péptidos presentados por el MHC de clase I y/o clase II. Preferiblemente, los péptidos antigénicos comprenden una secuencia de aminoácidos que corresponde sustancialmente a la secuencia de aminoácidos de un fragmento de un antígeno. Preferiblemente, dicho fragmento de un antígeno es un péptido presentado al MHC de clase I y/o clase II. Preferiblemente, un péptido antigénico comprende una secuencia de aminoácidos que corresponde sustancialmente a la secuencia de aminoácidos de dicho fragmento y se procesa para producir dicho fragmento, es decir, un péptido presentado al MHC de clase I y/o clase II derivado de un antígeno.

Si un péptido se presenta directamente, es decir, sin procesamiento, en particular sin escisión, 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 7-20 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.

Si un péptido es parte de una entidad más grande que comprende secuencias adicionales, por ejemplo, de una secuencia de vacuna o polipéptido, y debe presentarse 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 7-20 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, es decir, su secuencia corresponde sustancialmente y es preferiblemente completamente idéntica a un fragmento de un antígeno. Por lo tanto, una secuencia de un péptido antigénico o una secuencia de vacuna en una realización comprende una secuencia 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 que corresponde sustancialmente y es preferiblemente completamente idéntica a un fragmento de un antígeno y después del procesamiento de la secuencia del péptido antigénico o de la vacuna constituye el péptido presentado. Tal péptido producido por procesamiento comprende el cambio de secuencia identificado.

Un péptido o epítipo antigénico puede estar presente en una vacuna como parte de una entidad más grande tal como una secuencia de vacuna y/o un polipéptido que comprende más de un péptido o epítipo antigénico. El péptido o epítipo antigénico presentado se produce después de un procesamiento adecuado.

Los péptidos que tienen secuencias de aminoácidos que corresponden sustancialmente a una secuencia de un péptido que se presenta por el MHC de clase I pueden diferir en uno o más residuos que no son esenciales para el reconocimiento de TCR del péptido tal como se presenta por el MHC de clase I, o para la unión del péptido al MHC. Dichos péptidos sustancialmente correspondientes también son capaces de estimular un CTL sensible al antígeno y pueden considerarse inmunológicamente equivalentes. Los péptidos que tienen secuencias de aminoácidos que difieren de un péptido presentado en residuos que no afectan el reconocimiento de TCR pero que mejoran la estabilidad de la unión al MHC pueden mejorar la inmunogenicidad del péptido antigénico, y pueden denominarse en el presente documento como "péptido optimizado". Usando el conocimiento existente sobre cuáles de estos residuos pueden ser más propensos a afectar la unión al MHC o al TCR, se puede emplear un enfoque racional para el diseño de péptidos sustancialmente correspondientes. Los péptidos resultantes que son funcionales se contemplan como péptidos antigénicos.

Un péptido antigénico cuando es presentado por MHC debe ser reconocible por un receptor de células T. Preferiblemente, el péptido antigénico si es reconocido por un receptor de células T es capaz de inducir en presencia de señales coestimuladoras apropiadas, la expansión clonal de células T que porta el receptor de células T que reconoce específicamente el péptido antigénico. Preferiblemente, los péptidos antigénicos, en particular si se presentan en el contexto de moléculas del MHC, son capaces de estimular una respuesta inmune, preferiblemente una respuesta celular contra el antígeno del que derivan o células caracterizadas por la expresión del antígeno y preferiblemente caracterizadas por la presentación del antígeno. Preferiblemente, un péptido antigénico es capaz de estimular una respuesta celular contra una célula caracterizada por la presentación del antígeno con MHC de clase I y preferiblemente es capaz de estimular un CTL sensible al antígeno. Dicha célula es preferiblemente una célula objetivo.

El "procesamiento de antígeno" o "procesamiento" se refiere a la degradación de un polipéptido o antígeno en productos de procesación, que son fragmentos de dicho polipéptido o antígeno (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 antígeno, a células T específicas.

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.

Las células presentadoras de antígeno profesionales son muy eficientes en la internalización del antígeno, ya sea por fagocitosis o por endocitosis mediada por receptor, y luego muestran 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 de 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 profesionales presentadoras de antígeno.

Los tipos principales de células presentadoras de antígeno profesionales son las células dendríticas, que tienen el intervalo más amplio de presentación de antígeno, y son probablemente las células presentadoras de antígeno, macrófagos, células B y ciertas células epiteliales activadas más importantes.

Las células dendríticas (DC) son poblaciones de leucocitos que presentan antígenos capturados en los tejidos periféricos a las células T a través de las vías de presentación de antígeno 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 la inmunidad antitumoral.

Las células dendríticas se clasifican convenientemente como células "inmaduras" y "maduras", que pueden usarse como una forma simple de discriminar entre dos fenotipos bien caracterizados. Sin embargo, esta nomenclatura no debe interpretarse para excluir todas las posibles etapas intermedias de diferenciación.

5 Las células dendríticas inmaduras se caracterizan como 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 Fc γ y el 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, como el 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).

10 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 antígeno conducen al cebado de las células T, mientras que la presentación por células dendríticas inmaduras da como resultado 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.), citoquinas proinflamatorias (TNF, IL-1, IFN), ligadura de CD40 en la superficie de la célula dendrítica por CD40L, y sustancias liberadas de las células que sufren muerte celular estresante. Las células dendríticas se pueden derivar cultivando células de médula ósea *in vitro* con citoquinas, como el factor estimulante de colonias de granulocitos y macrófagos (GM-CSF) y el factor alfa de necrosis tumoral.

15 Las células presentadoras de antígeno no profesionales no expresan constitutivamente las proteínas del MHC de clase II requeridas para la interacción con las células T no alteradas; estas se expresan solo después de la estimulación de las células presentadoras de antígeno no profesionales por ciertas citoquinas tales como IFN γ .

Las "células presentadoras de antígeno" pueden cargarse con péptidos presentados por MHC de clase I traduciendo las células con ácido nucleico, preferiblemente ARN, que codifica un péptido o polipéptido que comprende el péptido a presentar, por ejemplo, un ácido nucleico que codifica el antígeno.

25 En algunas realizaciones, una composición farmacéutica que comprende un vehículo de administración génica que se dirige a una célula dendrítica u otra célula presentadora de antígeno puede administrarse a un paciente, dando como resultado una transfección que ocurre *in vivo*. La transfección *in vivo* de células dendríticas, por ejemplo, generalmente se puede realizar utilizando cualquier método conocido en la técnica, 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 invención, el término "célula presentadora de antígeno" también incluye células objetivo.

30 "Célula objetivo" significa una célula que es un objetivo para una respuesta inmune tal como una respuesta inmune celular. Las células objetivo incluyen células que presentan un antígeno o un epítipo de 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 objetivo es una célula que expresa un antígeno como se describe en el presente documento y que presenta preferiblemente dicho antígeno con MHC de clase I.

35 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 reconocido por el sistema inmune, por ejemplo, que es reconocido por una célula T, en particular cuando se presenta en el contexto de las 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 ser preferiblemente de 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 invención sea un epítipo de células T.

45 De acuerdo con la invención, 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" o "péptido antigénico". El término "péptido de unión al MHC" se refiere a un péptido que se une a una molécula del MHC de clase I y/o una molécula del MHC de clase II. En el caso de los complejos de MHC de clase I/péptido, los péptidos de unión son típicamente de 8-10 aminoácidos de longitud, aunque los péptidos más largos o más cortos pueden ser efectivos. En el caso de los complejos del MHC de clase II/péptido, los péptidos de unión son típicamente de 10-25 aminoácidos de longitud y en particular de 13-18 aminoácidos de longitud, mientras que los péptidos más largos y más cortos pueden ser efectivos.

50 Los términos "epítipo", "péptido antigénico", "epítipo antigénico", "péptido inmunogénico" y "péptido de unión al MHC" se usan indistintamente en el presente documento y preferiblemente se refieren a una representación incompleta de un antígeno que es preferiblemente capaz de provocar una respuesta inmune contra el antígeno o una célula que expresa o comprende y que presenta preferiblemente al antígeno. Preferiblemente, los términos se refieren a una porción inmunogénica de un antígeno. Preferiblemente, es una porción de un antígeno que es reconocida (es decir, unida específicamente) por un receptor de células T, en particular si se presenta en el contexto de moléculas del MHC. Tales porciones inmunogénicas preferidas se unen a una molécula del MHC de clase I o de clase II. Como se usa en el presente documento, se dice que una porción inmunogénica se "une a" una molécula del MHC de clase I o clase II si dicha unión es detectable usando cualquier ensayo conocido en la técnica.

Como se usa en el presente documento, el término "neopítipo" 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 en 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 neopítipo

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 60%, más preferiblemente al menos 70%, incluso más preferiblemente al menos 80%, y lo más preferiblemente al menos 90% de los aminoácidos de dicha secuencia de aminoácidos. Preferiblemente, si la porción es una fracción discontinua, dicha fracción discontinua está compuesta de 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 de 2, 3, 4, 5, 6, 7, 8 o más, preferiblemente no más de 4 partes de dicha secuencia de aminoácidos, en donde cada parte preferiblemente comprende 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 el presente documento y se refieren a un elemento continuo. Por ejemplo, una parte de una estructura tal como una secuencia de aminoácidos o proteína se refiere a un elemento continuo de dicha estructura. Una porción, una parte o un fragmento de una estructura comprende preferiblemente 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 de la que deriva. En el contexto de la presente invención, una "parte" de una estructura tal como una secuencia de aminoácidos comprende preferiblemente, consiste preferiblemente en 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 invención se refiere a una célula que ejerce funciones efectoras durante una reacción inmune. Una "célula inmunorreactiva" preferiblemente es capaz de unirse a un antígeno o una célula caracterizada por la presentación de un antígeno o un péptido antigénico derivado de un antígeno y que media una respuesta inmune. Por ejemplo, tales células secretan citoquinas y/o quimioquinas, secretan anticuerpos, reconocen células cancerosas y opcionalmente eliminan tales células. Por ejemplo, las células inmunorreactivas comprenden células T (células T citotóxicas, células T auxiliares, células T infiltrantes de tumores), células B, células asesinas naturales, neutrófilos, macrófagos y células dendríticas. Preferiblemente, en el contexto de la presente invención, las "células inmunorreactivas" son células T, preferiblemente células T CD4⁺ y/o CD8⁺.

Preferiblemente, una "célula inmunorreactiva" reconoce un antígeno o un péptido antigénico derivado de un antígeno 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 antígeno o células enfermas tales como células cancerosas. Preferiblemente, dicho reconocimiento permite que la célula que reconoce un antígeno o un péptido antigénico derivado de dicho antígeno sea sensible o reactiva. Si la célula es una célula T auxiliar (célula T CD4⁺) que tiene receptores que reconocen un antígeno o un péptido antigénico derivado de un antígeno en el contexto de moléculas del MHC de clase II, tal capacidad de respuesta o reactividad puede implicar la liberación de citoquinas 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 las 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, a través de apoptosis o de lisis de células mediadas por perforina. De acuerdo con la invención, la capacidad de respuesta de CTL puede incluir flujo sostenido de calcio, división celular, producción de citoquinas como IFN- γ y TNF- α , sobreexpresión de marcadores de activación tales como CD44 y CD69, y destrucción citolítica específica de células objetivo que expresan antígeno. La capacidad de respuesta de CTL también se puede determinar utilizando un indicador artificial que indica con precisión la capacidad de respuesta de CTL. Tales CTL que reconocen un antígeno o un péptido antigénico derivado de un antígeno y son sensibles o reactivos también se denominan en este documento "CTL sensible 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 usan indistintamente en el presente documento e incluyen células T auxiliares (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 juegan 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 auxiliares ayudan a otros glóbulos blancos en los procesos inmunológicos, que incluyen la maduración

de 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 auxiliares se activan cuando se presentan con antígenos peptídicos por moléculas del MHC de clase II que se expresan en la superficie de las células presentadoras de antígeno (APC). Una vez activadas, se dividen rápidamente y secretan pequeñas proteínas llamadas citoquinas que regulan o ayudan en la respuesta inmune activa.

Las células T citotóxicas destruyen células infectadas por virus y 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 objetivos 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 peptídicas separadas, que se producen a partir de los genes independientes del receptor de células T alfa y beta (TCR α y TCR β) y se denominan cadenas de TCR α y β . Las células T $\gamma\delta$ (células T gamma delta) representan un pequeño subconjunto de células T que poseen un receptor distinto de células T (TCR) 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 complejo mayor de histocompatibilidad (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 (APC) profesional, generalmente una célula dendrítica en el caso de respuestas ingenuas, aunque las células B y los macrófagos pueden ser APC importantes. Los péptidos presentados a las células T CD8+ por las moléculas del MHC de clase I tienen típicamente una longitud de 8-10 aminoácidos; los péptidos presentados a las células T CD4+ por las moléculas del MHC de clase II son típicamente más largos, ya que los extremos de la hendidura de unión de la molécula del MHC de clase II están abiertos.

De acuerdo con la presente invención, un receptor de células T es capaz de unirse a un objetivo predeterminado si tiene una afinidad significativa por dicho objetivo predeterminado y se une a dicho objetivo predeterminado en ensayos estándar. La "afinidad" o "afinidad de unión" a menudo se mide por la constante de disociación de equilibrio (K_D). Un receptor de células T no es (sustancialmente) capaz de unirse a un objetivo si no tiene una afinidad significativa por dicho objetivo y no se une significativamente a dicho objetivo en los ensayos estándar.

Un receptor de células T es preferiblemente capaz de unirse específicamente a un objetivo predeterminado. Un receptor de células T es específico para un objetivo predeterminado si es capaz de unirse a dicho objetivo predeterminado mientras que no es (sustancialmente) capaz de unirse a otros objetivos, es decir, no tiene una afinidad significativa por otros objetivos y no se une significativamente a otros objetivos en ensayos estándar.

Los linfocitos T citotóxicos pueden generarse *in vivo* mediante la incorporación de un antígeno o un péptido antigénico en células presentadoras de antígeno *in vivo*. El antígeno o péptido antigénico puede representarse como proteína, como ADN (por ejemplo, dentro de un vector) o como ARN. El antígeno puede procesarse para producir un asociado péptido para la molécula del MHC, mientras que un fragmento del mismo puede presentarse sin la necesidad de un procesamiento adicional. Este último es el caso en particular, si estos pueden unirse a las moléculas del MHC. En general, es posible la administración a un paciente mediante inyección intradérmica. Sin embargo, la inyección también puede llevarse a cabo por vía intranodal 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 los linfocitos T citotóxicos autólogos que luego se propagan.

La activación específica de las células T CD4+ o CD8+ se puede detectar de varias maneras. Los métodos para detectar la activación de células T específicas incluyen la detección de la proliferación de células T, la producción de citoquinas (por ejemplo, linfoquinas) 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.

El término "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 produce 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 los receptores de células T. Las proteínas codificadas por el MHC se expresan en la superficie de las células y muestran tanto antígenos propios (fragmentos de péptidos de la propia célula) como antígenos no propios (por ejemplo, fragmentos de microorganismos invasores) en una célula T.

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 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 inmunitario, específicamente en las células presentadoras de antígeno, las proteínas del MHC de clase II contienen cadenas α y

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

5 En humanos, los genes en la región del MHC que codifican proteínas presentadoras de antígeno en la superficie celular se denominan genes de antígeno leucocitario humano (HLA). Sin embargo, la abreviatura MHC se usa a menudo para referirse a los productos del gen del HLA. Los genes del HLA incluyen los nueve llamados genes MHC clásicos: HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA y HLA-DRB1.

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

10 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, mediante el procesamiento del antígeno, en el contexto de moléculas del MHC, en particular moléculas del MHC de 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 tal como 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 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 es naturalmente presentado por las células que expresan un antígeno.

20 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 al tipo de efecto inmunológico, 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 invención, el término "inmunológicamente equivalente" se usa preferiblemente con respecto a los efectos inmunológicos o propiedades 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 inmune de un sujeto induce una reacción inmune que tiene una especificidad de reacción con la secuencia de aminoácidos de referencia.

30 El término "funciones efectoras inmunes" en el contexto de la presente invención incluye cualquier función mediada por componentes del sistema inmunitario que resultan, por ejemplo, en la destrucción de células tumorales, o en la inhibición del crecimiento tumoral y/o inhibición del desarrollo tumoral, incluida la inhibición de la diseminación y metástasis tumoral. Preferiblemente, las funciones efectoras inmunes en el contexto de la presente invención son funciones efectoras mediadas por células T. Dichas funciones comprenden en el caso de una célula T auxiliar (célula T CD4+) el reconocimiento de un antígeno o un péptido antigénico derivado de un antígeno en el contexto de moléculas del MHC de clase II por receptores de células T, la liberación de citoquinas 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 péptido antigénico derivado de un antígeno en el contexto de las moléculas del MHC de clase I por los 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 citoquinas tales como IFN- γ y TNF- α , y destrucción citolítica específica de células objetivo que expresan antígeno.

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 las regiones codificantes de un genoma. El término "transcriptoma" se refiere al conjunto de todas las moléculas de ARN.

45 De acuerdo con la invención, un "ácido nucleico" es 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 de acuerdo con la invención ADN genómico, ADNc, ARNm, moléculas producidas recombinantemente y sintetizadas químicamente. De acuerdo con la invención, un ácido nucleico puede estar presente como una molécula monocatenaria o bicatenaria y lineal o covalentemente circularmente cerrada. Un ácido nucleico puede, de acuerdo con la invención, aislarse. El término "ácido nucleico aislado" significa, de acuerdo con la invención, que el ácido nucleico (i) se amplificó *in vitro*, por ejemplo a través de la reacción en cadena de la polimerasa (PCR), (ii) se produjo de forma recombinante por clonación, (iii) se purificó, por ejemplo, por escisión y separación por electroforesis en gel, o (iv) se sintetizó, por ejemplo, por síntesis química. Se puede emplear un ácido nucleico para la introducción en, es decir, la transfección de células, en particular, en forma de ARN que se pueden preparar por transcrito *in vitro* a partir de una plantilla de ADN. Además, el ARN puede modificarse antes de la aplicación estabilizando secuencias, protección y poliadenilación.

El término "material genético" se refiere a 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.

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 las células germinales (esperma y óvulo) 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 aminoácidos tal como una sustitución de aminoácidos en el producto de traducción.

De acuerdo con la invención, el término "mutación" incluye mutaciones puntuales, indeles, fusiones, cromotripsis y ediciones de ARN.

De acuerdo con la invención, 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 en nucleótidos. En las regiones de codificación del genoma, a menos que la longitud de un indel sea un múltiplo de 3, producen una mutación de desplazamiento de marco. Los indeles pueden contrastarse con una mutación puntual; donde un Indel inserta y elimina nucleótidos de una secuencia, una mutación puntual es una forma de sustitución que reemplaza a uno de los nucleótidos.

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 de los dos asociados de fusión. Alternativamente, un protooncogén se fusiona con un promotor fuerte y, por lo tanto, la función oncogénica se establece para funcionar mediante una sobrerregulación causada por el promotor fuerte del asociado de fusión secuencia arriba. Los transcritos de fusión oncogénicos también pueden ser causados por eventos de empalme trans o eventos de lectura.

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

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

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

Se puede usar cualquier método de secuenciación adecuado de acuerdo con la invención, prefiriéndose las tecnologías de secuenciación de próxima generación (NGS). Los métodos de secuenciación de tercera generación podrían sustituir 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 invención significan 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 al romper todo el genoma en pedazos pequeños. Dichas tecnologías NGS (también conocidas como tecnologías de secuenciación paralela masiva) pueden 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 tiempo muy cortos, 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 contexto de la presente invención, por ejemplo, aquellas 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

1) La tecnología de secuenciación por síntesis conocida como pirosecuenciación implementada por ejemplo, en el Genome Sequencer^{MR} GS-FLX 454 de la compañía asociada a Roche 454 Life Sciences (Branford, Connecticut), descrita 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 la que las perlas de unión de ADN monocatenario se encapsulan por agitación vigorosa en micelas acuosas que contienen reactivos de PCR rodeados por aceite para la 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.

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

10 3) Enfoques de secuenciación por ligadura, por ejemplo, implementado en la plataforma SOLid^{MR} de Applied Biosystems (ahora Life Technologies Corporation, Carlsbad, California). En esta tecnología, un grupo de todos los oligonucleótidos posibles de una longitud fija se marcan de acuerdo con la posición secuenciada. Los oligonucleótidos se hibridan se ligan; la ligadura preferencial por ADN ligasa para secuencias coincidentes da como resultado una señal informativa del nucleótido en esa posición. Antes de la secuenciación, el ADN se amplifica por PCR en emulsión. Las perlas resultantes, cada una conteniendo solamente copias de la misma molécula de ADN, se depositan 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 ligadura mediante el uso de una PCR de emulsión dispuesta aleatoriamente, basada en perlas, para amplificar fragmentos de ADN para secuenciación paralela.

15 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 de ADN o ARN únicas sin amplificación, definidas como secuenciación de ADN en tiempo real de molécula única (SMRT). Por ejemplo, HeliScope utiliza un sistema de detección de fluorescencia altamente sensible para detectar directamente cada nucleótido a medida que se sintetiza. Se ha desarrollado un enfoque similar basado en la transferencia de energía de resonancia de fluorescencia (FRET) de Visigen Biotechnology (Houston, Texas). Otras técnicas de molécula única basadas en fluorescencia son de U.S. Genomics (GeneEngine^{MR}) y Genovox (AnyGene^{MR}).

20 5) Nano-tecnologías para la secuenciación de molécula única en la que se utilizan varias nanoestructuras que están, por ejemplo, dispuestas en un chip para controlar el movimiento de una molécula de polimerasa en una sola cadena durante la replicación. Ejemplos no limitativos para enfoques basados en nanotecnologías son la plataforma GridON^{MR} de Oxford Nanopore Technologies (Oxford, Reino Unido), las plataformas de secuenciación de nano poros 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 nanobolas de ADN (DNB) llamada ligadura combinatoria de anclaje de sonda (cPAL^{MR}).

25 6) Tecnologías basadas en microscopía electrónica para la secuenciación de moléculas únicas, por ejemplo los desarrollados por LightSpeed Genomics (Sunnyvale, California) y Halcyon Molecular (Redwood City, California)

30 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 micro maquinados de alta densidad para realizar este proceso bioquímico de una manera masivamente paralela. Cada pozo tiene una plantilla de ADN diferente. Debajo de los pozos hay una capa sensible a iones y debajo de eso un sensor de iones patentado.

35 Preferiblemente, las preparaciones de ADN y ARN sirven como material de partida para NGS. Tales ácidos nucleicos pueden obtenerse fácilmente de muestras tales como material biológico, por ejemplo, de tejidos tumorales embebidos en parafina, congelados instantáneamente o fijados con formalina o de células recién aisladas o de CTC que están presentes en la sangre periférica de los pacientes. El ADN o ARN genómico no mutado normal se puede extraer del tejido somático normal, sin embargo, se prefieren las células de la línea germinal en el contexto de la presente invención. El ADN o ARN de línea germinal se extrae de 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 altamente fragmentados, son adecuados para aplicaciones de NGS.

40 En la literatura se describen varios métodos NGS dirigidos para la secuenciación del exoma (para revisión, véase, por ejemplo, Teer y Mullikin 2010: Human Mol Genet 19 (2), R145-51), todos los cuales se pueden usar junto con la presente invención. 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 una matriz (por ejemplo, Hodges et al. 2007: Nat. Genet. 39, 1522-1527) y basadas en líquidos (por ejemplo, Choi et al., 2009: Proc. Natl. Acad. Sci USA 106, 19096-19101). Los kits comerciales para la preparación de muestras de ADN y la posterior captura del exoma también están disponibles: 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 al comparar, por ejemplo, la secuencia de una muestra tumoral a 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 de estos tipos de muestra. Por lo tanto, se prefiere que la secuencia de una muestra de referencia tal como la secuencia de una muestra de línea germinal se determine dos veces, tres veces o más. Alternativa o adicionalmente, la secuencia de una muestra tumoral 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 tumoral 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 las réplicas de una muestra de referencia, tal como una muestra de línea germinal, se puede estimar la tasa esperada de mutaciones somáticas falsas positivas (FDR) como una cantidad estadística. Las repeticiones técnicas de una muestra deben generar resultados idénticos y cualquier mutación detectada en esta "misma comparación versus la misma comparación" 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 utilizar 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 un único puntaje 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.

De acuerdo con la invención, se puede aplicar un método de genotipado de células individuales de alto rendimiento de todo el genoma.

En una realización del genotipo de células individuales de alto rendimiento de todo el genoma, se puede usar la plataforma Fluidigm. Tal enfoque puede comprender los siguientes etapas:

1. Muestra de tejido/células tumorales y tejido sano de un paciente dado.

2. El material genético se extrae de células cancerosas y sanas y luego se secuencia su exoma (ADN) utilizando protocolos estándar de secuenciación de próxima generación (NGS). La cobertura del NGS es tal que se pueden detectar alelos heterocigóticos con al menos un 5% de frecuencia. El transcriptoma (ARN) también se extrae de las células cancerosas, se convierte en ADNc y se secuencia para determinar qué genes expresan las células cancerosas.

3. Las variaciones de nucleótidos sencillos (SNV) expresadas no sinónimas se identifican como se describe en el presente documento. Los sitios que son SNP en el tejido sano se separan por filtración.

4. Se seleccionan N = 96 mutaciones de (3) que abarcan diferentes frecuencias. Los ensayos de genotipado de SNP basados en la detección de fluorescencia están diseñados y sintetizados para estas mutaciones (los ejemplos de dichos ensayos incluyen: ensayos de SNP basados en TaqMan de Life Technologies o ensayos SNPtype de Fluidigm). Los ensayos incluirán cebadores específicos de amplificación objetivo (STA) para amplificar los amplicones que contienen el SNV dado (esto es estándar en los ensayos TaqMan y SNPtype).

5. Las células individuales se aislarán del tumor y del tejido sano, ya sea por microdissección con láser (LMD) o por desagregación en suspensiones de células individuales, seguido de una clasificación como se describió anteriormente (Dalerba P. et al., (2011) Nature Biotechnology 29: 1120-1127). Las células se pueden elegir sin preselección (es decir, imparcial) o, como alternativa, se pueden enriquecer las células cancerosas. Los métodos de enriquecimiento incluyen: tinción específica, clasificación por tamaño de célula, inspección histológica durante LMD, etc.

6. Las células individuales se aislarán en tubos de PCR que contengan una mezcla maestra con los cebadores STA y se amplificarán los amplicones que contienen las SNV. Alternativamente, el genoma de la célula individual se amplificará a través de la amplificación del genoma completo (WGA) como se describió previamente (Frumkin D. et al., (2008) Cancer Research 68: 5924). La lisis celular se logrará mediante la etapa de calentamiento a 95 °C o a través de un regulador de lisis dedicado.

7. Las muestras amplificadas con STA se diluyen y se cargan en la matriz de genotipado Fluidigm.

8. Las muestras de tejido sano se usarán como controles positivos para determinar grupos de alelos homocigóticos (sin mutación). Dado que los datos de NGS indican que las mutaciones homocigóticas son extremadamente raras, generalmente se esperan solo dos grupos: XX y XY, con X = saludable.

9. El número de matrices que se pueden ejecutar no está limitado, lo que permite, en la práctica, analizar hasta ~1000 células individuales (~10 matrices). Si se realiza en 384 placas, la preparación de la muestra se puede reducir a unos pocos días.

10. Luego se determinan las SNV para cada célula.

En otra realización el genotipado de células individuales de alto rendimiento de todo el genoma, se puede usar la plataforma NGS. Tal enfoque puede comprender las siguientes etapas:

1. Las etapas 1 a 6 anteriores son idénticas, excepto que N (número de SNV analizados) puede ser mucho mayor que 96. En el caso de WGA, se realizarán después varios ciclos de STA. Los cebadores de STA contendrán dos secuencias de etiquetas universales en cada cebador.

5 2. Después de la STA, los cebadores de códigos de barras se amplificarán por PCR en los amplicones. Los cebadores de códigos de barras contienen secuencias de códigos de barras únicas y las secuencias de etiquetas universales anteriores. Cada célula contendrá así un código de barras único.

3. Los amplicones de todas las células se mezclarán y secuenciarán mediante NGS. La limitación práctica sobre el número de células que se pueden multiplexar es el número de placas que se pueden preparar. Como las muestras se pueden preparar en 384 placas, un límite práctico sería de ~5000 células.

10 4. En función de los datos de secuencia, se detectan SNV (u otras anomalías estructurales) de las células individuales.

15 Para priorizar antígenos, se puede usar la reconstrucción filogenética tumoral basada en el genotipado de células individuales ("priorización de antígeno filogenético") de acuerdo con la invención. Además de la priorización del antígeno basada en criterios como la expresión, el tipo de mutación (no sinónimo versus otro), las características de unión al MHC, etc., se puede utilizar una dimensión adicional para la priorización diseñada para hacer frente a la heterogeneidad intra e intertumoral y al sesgo de biopsia como se describe por ejemplo a continuación.

1. Identificación de los antígenos más abundantes.

20 La frecuencia de cada SNV se puede estimar con precisión basándose en el ensayo de células individuales descrito anteriormente en relación con el método de genotipado de células individuales de alto rendimiento de todo el genoma y las SNV más abundantes presentes se pueden seleccionar para proporcionar vacunas individualizadas para el cáncer (IVAC).

2. Identificación de antígenos basales primarios basados en análisis de árbol enraizado

25 Los datos de NGS de tumores sugieren que las mutaciones homocigóticas (aciertos en ambos alelos) son eventos raros. Por lo tanto, no se requiere para obtener el de haplotipo y se puede crear un árbol filogenético de las mutaciones somáticas tumorales a partir del conjunto de datos SNV de la célula individual. La secuencia de la línea germinal se usará para enraizar el árbol. Usando algoritmos para reproducir secuencias ancestrales, se reproducirán las secuencias de nodos cerca de la raíz del árbol. Estas secuencias contienen las primeras mutaciones que se predijo que existían en el tumor primario (definido en este documento como las mutaciones/antígenos basales primarios). Debido a la baja probabilidad de que ocurran dos mutaciones en los mismos alelos en la misma posición en el genoma, se predice que las mutaciones en las secuencias ancestrales se fijan en el tumor.

30 Priorizar los antígenos basales primarios no es equivalente a priorizar las mutaciones más frecuentes en la biopsia (aunque se espera que las mutaciones basales primarias se encuentren entre las más frecuentes en la biopsia). La razón es la siguiente: digamos que dos SNV parecen estar presentes en todas las células derivadas de una biopsia (y por lo tanto tienen la misma frecuencia: 100%), pero una mutación es basal y la otra no, entonces la mutación basal debe seleccionarse para IVAC. Esto se debe a que es probable que la mutación basal se presente en todas las regiones del tumor, mientras que la última mutación puede ser una mutación más reciente que por casualidad se reparó en la región donde se tomó la biopsia. Además, es probable que existan antígenos basales en los tumores metastásicos derivados del tumor primario. Por lo tanto, al priorizar los antígenos basales para IVAC, se puede aumentar en gran medida la posibilidad de que IVAC pueda erradicar todo el tumor y no solo una parte del tumor.

40 Si existen tumores secundarios y de estos también se tomaron muestras, se puede estimar un árbol evolutivo de todos los tumores. Esto puede mejorar la robustez del árbol y permitir la detección de mutaciones basales a todos los tumores.

3. Identificación de antígenos que abarcan al máximo el tumor o tumores

45 Otro enfoque para obtener antígenos que cubran al máximo todos los sitios tumorales es tomar varias biopsias del tumor. Una estrategia sería seleccionar los antígenos identificados por el análisis NGS para que estén presentes en todas las biopsias. Para mejorar las probabilidades de identificar mutaciones basales, se puede realizar un análisis filogenético basado en mutaciones de células individuales de todas las biopsias.

En caso de metástasis, se pueden obtener biopsias de todos los tumores y se pueden seleccionar mutaciones identificadas a través de NGS que son comunes a todos los tumores.

4. Uso de CTC para priorizar antígenos que inhiben la metástasis

50 Se cree que los tumores metastásicos se derivan de células individuales. Por lo tanto, al genotipar células individuales extraídas de diferentes tumores de un paciente determinado junto con el genotipado de células tumorales en circulación (CTC) del paciente, se puede reconstruir la historia evolutiva del cáncer. La expectativa es observar el tumor metastásico que evoluciona a partir del tumor original a través de un clado de CTC derivado del tumor primario.

5 A continuación se describe (un método imparcial para identificar, contar y sondear genéticamente CTC) una extensión del método de genotipado de células individuales de alto rendimiento del genoma total para un aislamiento imparcial y de análisis genómico de CTC. Usando el análisis descrito anteriormente, se puede reconstruir un árbol filogenético del tumor cebador, CTC y tumores secundarios que surgen de metástasis (si existen). Con base en este árbol, se pueden identificar mutaciones (pasajeras o conductoras) que ocurrieron en el momento o muy de cerca después de que los CTC se separaron por primera vez del tumor primario. La expectativa es que los genomas de CTC que surgen del tumor primario son evolutivos más similares a los genomas tumorales primarios que a los genomas tumorales secundarios. Además, se espera que los genomas de CTC que surgen del tumor primario contengan mutaciones únicas que se fijan en los tumores secundarios, o que probablemente se fijaran si se forman tumores secundarios en el futuro. Estas mutaciones únicas se pueden priorizar para que IVAC ataque (o prevenga) la metástasis.

10 La ventaja de priorizar las mutaciones CTC frente a las mutaciones basales primarias es que los antígenos derivados de CTC pueden movilizar células T específicamente para atacar la metástasis y, por lo tanto, serán un brazo independiente de las células T que se dirigen al tumor primario (usando diferentes antígenos). Además, si hay pocos tumores secundarios (o ninguno), se espera que la posibilidad de escape inmune de los antígenos derivados de CTC sea menor, ya que la probabilidad de escape tumoral debería aumentar con el número de células cancerosas que transportan el antígeno dado.

5. Identificación de antígenos que se producen conjuntamente en la misma célula (el "cóctel" de IVAC)

20 Se cree que el tumor evoluciona para suprimir mutaciones debido a la presión de selección del sistema inmune y la terapia. Las vacunas contra el cáncer dirigidas a múltiples antígenos que se producen conjuntamente en la misma célula y que también son frecuentes en el tumor tienen una mayor probabilidad de anular los mecanismos de escape del tumor y, por lo tanto, reducen la posibilidad de una recaída. Tales "vacunas de cóctel" serían análogas a la terapia de combinación antirretroviral para pacientes VIH+. Las mutaciones concurrentes se pueden identificar mediante análisis filogenético o inspeccionando la alineación del SNV de todas las células.

25 Además, de acuerdo con la invención, se puede usar un método imparcial para identificar, contar y sondear genéticamente CTC. Tal enfoque puede comprender las siguientes etapas:

1. Obtener una biopsia de tumor o tumores y determinar el atlas de mutaciones somáticas.

2)

Opción 1: Seleccionar $N \geq 96$ mutaciones para una mayor investigación basada en esquemas de priorización previamente establecidos.

30 Opción 2: Realizar el ensayo de células individuales (véase el método de genotipado de células individuales de alto rendimiento del genoma total descrito anteriormente) seguido del análisis filogenético para seleccionar $N \geq 96$ mutaciones basales primarias y posiblemente mutaciones más recientes para maximizar la diversidad. Las primeras mutaciones son útiles para identificar las CTC (véase más abajo), y las últimas para generar un análisis filogenético (véase la sección "Identificación de antígenos que se producen conjuntamente en la misma célula (el "cóctel" de IVAC)").

35 3. Obtener sangre completa del paciente con cáncer.

4. Lisar los glóbulos rojos

5. Remover los glóbulos blancos agotando las células CD45+ (por ejemplo, mediante la clasificación, perlas magnéticas conjugadas con el anticuerpo anti CD45, etc.) para enriquecer las CTC.

40 6. Remover el ADN libre por digestión de ADNasa. El origen del ADN libre puede ser el ADN presente en la sangre o el ADN de células muertas.

7. Clasificar las células restantes en tubos de PCR, realizar el STA (en función de las mutaciones seleccionadas) y detectar en Fluidigm (método de genotipado de células individuales de alto rendimiento de todo el genoma descrito anteriormente). Las CTC generalmente deberían ser positivas para múltiples SNV.

45 8. Las células identificadas como cancerosas (= CTC) se pueden analizar luego filogenéticamente en función del panel de SNV seleccionados (véase sección "Identificación de antígenos que se producen conjuntamente en la misma célula (el "cóctel" de IVAC)").

50 También es posible combinar este método con métodos establecidos previamente para CTC aisladas. Por ejemplo, se puede clasificar por células EpCAM+, o células positivas para citoqueratinas (Rao CG. et al., (2005) International journal of oncology 27: 49; Allard WJ. et al., (2004) Clinical Cancer Research 10: 6897-6904) Estas CTC putativas se pueden verificar/perfilar en el Fluidigm/NGS para derivar sus mutaciones.

Este método puede usarse para contar CTC. Dado que el método no se basa en un marcador particular, que puede o no expresarse por las células cancerosas, sino más bien en el perfil de mutación de las mutaciones somáticas del

cáncer exclusivas del paciente, este es un método imparcial para detectar y enumerar las CTC.

De acuerdo con la invención, se puede usar un enfoque que implica la reconstrucción filogenética tumoral basada en el genotipado de células individuales para enriquecer las mutaciones conductoras ("filtrado filogenético").

5 En una realización de este enfoque, se realiza un análisis filogenético pan-tumoral para recuperar mutaciones conductoras. Por ejemplo, se pueden detectar mutaciones conductoras de $n = 1$ tumor.

10 En la sección anterior "Identificación de antígenos basales primarios basados en el análisis de árboles enraizados", se describe un método para recuperar secuencias ancestrales y/o identificar células que tienen secuencias cercanas a la raíz del árbol. Se espera que el número de mutaciones en estas secuencias sea significativamente menor que el número de mutaciones en la muestra global del cáncer, ya que, por definición, se trata de secuencias cercanas a la raíz del árbol. Por lo tanto, al seleccionar secuencias cercanas a la raíz del árbol, se espera que muchas mutaciones pasajeras sean "filtradas filogenéticamente". Este procedimiento tiene el potencial de enriquecer enormemente las mutaciones conductoras. Las mutaciones conductoras pueden usarse para identificar/seleccionar el tratamiento para un paciente o pueden usarse como guías para nuevas terapias.

En otro ejemplo, pueden detectarse mutaciones conductoras de $n > 1$ tumores de un tipo dado.

15 Al reconstruir las mutaciones basales primarias de muchos tumores de un tipo particular, se puede aumentar en gran medida la posibilidad de detectar mutaciones conductoras. Dado que las secuencias basales cerca de la raíz del árbol filtran muchas mutaciones pasajeras, se espera que la relación señal/ruido en la detección de mutaciones conductoras aumente considerablemente. Por lo tanto, este método tiene el potencial de detectar (1) mutaciones conductoras menos frecuentes (2) mutaciones conductoras frecuentes a partir de menos muestras.

20 En otra realización del enfoque que implica la reconstrucción filogenética tumoral basada en el genotipo de células individuales para enriquecer las mutaciones conductoras ("filtrado filogenético"), se realiza un análisis filogenético para recuperar la metástasis que causa mutaciones conductoras.

25 En la sección anterior "Uso de CTC para priorizar antígenos que inhiben la metástasis", se describe un método para detectar mutaciones asociadas a CTC. Este método también se puede utilizar para enriquecer las mutaciones conductoras que conducen a la metástasis. Por ejemplo, al mapear la filogenia combinada del tumor cebador, los tumores secundarios y las CTC, las CTC derivadas del tumor primario deben conectarse entre los clados de los tumores secundarios primarios. Tal análisis filogenético puede ayudar a identificar con precisión las mutaciones únicas en esta transición entre el cebador y los tumores secundarios. Una fracción de estas mutaciones pueden ser mutaciones conductoras. Además, al comparar mutaciones únicas de CTC de diferentes instancias del mismo cáncer (es decir, $n > 1$ tumores), se pueden enriquecer aún más las mutaciones conductoras únicas que causan metástasis.

30 De acuerdo con la invención, se puede usar el análisis filogenético para identificar tumores primarios versus secundarios.

35 En caso de metástasis, si se muestrean todos los tumores, se puede usar un árbol enraizado para predecir el orden temporal en que aparecieron los tumores: qué tumor es el tumor primario (ganglios más cercanos a la raíz del árbol) y qué tumores son los más recientes. Esto puede ser útil en casos en los que es difícil determinar qué tumor es el primario.

40 En el contexto de la presente invención, el término "ARN" se refiere a una molécula que comprende al menos un residuo de ribonucleótido y que preferiblemente está compuesta total o sustancialmente de 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 difiere del ARN natural por adición, eliminación, sustitución y/o alteración de uno o más nucleótidos. Dichas alteraciones pueden incluir la adición de material no nucleotídico, tal como al 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 natural.

45 De acuerdo con la presente invención, el término "ARN" incluye y preferiblemente se refiere a "ARNm". El término "ARNm" significa "ARN mensajero" y se refiere a un "transcrito" que se genera usando una plantilla de ADN y codifica un péptido o polipéptido. Típicamente, un ARNm comprende una 5'-UTR, una región codificante de proteínas y una 3'-UTR. El ARNm solo posee una vida media limitada en las células e *in vitro*. En el contexto de la presente invención, el ARNm puede generarse por transcrito *in vitro* a partir de una plantilla de ADN. La metodología de transcrito *in vitro* es conocida por la persona experta. Por ejemplo, hay una variedad de kits de transcrito *in vitro* disponibles comercialmente.

55 De acuerdo con la invención, la estabilidad y la eficiencia de traducción del ARN pueden modificarse según se requiera. Por ejemplo, el ARN puede estabilizarse y su traducción aumentarse en una o más modificaciones que tengan efectos

estabilizadores y/o aumenten la eficiencia de traducción del ARN. Dichas modificaciones se describen, por ejemplo, en el documento PCT/EP2006/009448. Para aumentar la expresión del ARN usado de acuerdo con la presente invención, puede modificarse dentro de la región de codificación, 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.

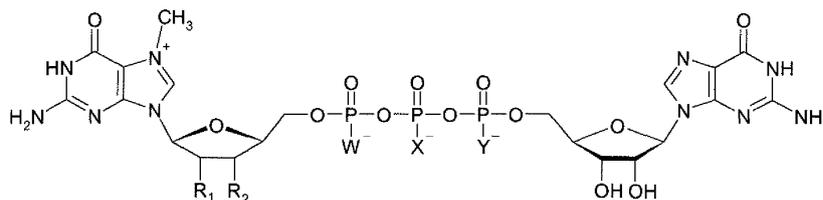
El término "modificación" en el contexto del ARN utilizado en la presente invención incluye cualquier modificación de un ARN que no está presente de forma natural en dicho ARN.

En una realización de la invención, el ARN utilizado de acuerdo con la invención no tiene 5'-trifosfatos sin proteger. La eliminación de tales 5'-trifosfatos sin proteger se puede lograr tratando el ARN con una fosfatasa.

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

En una realización, el término "modificación" se refiere a proporcionar un ARN con una caperuza 5' o un análogo de caperuza 5'. El término "caperuza 5'" se refiere a una estructura de caperuza 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 5' a 5'-trifosfato. En una realización, esta guanosina se metila en la posición 7. El término "caperuza 5' convencional" se refiere a una caperuza 5' de ARN natural, preferiblemente a la caperuza de 7-metilguanosa (m⁷G). En el contexto de la presente invención, el término "caperuza 5'-" incluye un análogo de caperuza 5' que se asemeja a la estructura de la caperuza 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.

Preferiblemente, el extremo 5' del ARN incluye una estructura de Cap que tiene la siguiente fórmula general:



en la que R₁ y R₂ son independientemente hidroxilo o metoxi y W, X e Y son independientemente oxígeno, azufre, selenio o BH₃. En una realización preferida, R₁ y R₂ son hidroxilo y W, X e Y son oxígeno. En una realización preferida adicional, uno de R₁ y R₂, preferiblemente R₁ es hidroxilo y el otro es metoxi y W, X e Y son oxígeno. En una realización preferida adicional, R₁ y R₂ son hidroxilo y uno de W, X e Y, preferiblemente X es azufre, selenio o BH₃, preferiblemente azufre, mientras que los otros son oxígeno. En una realización preferida adicional, uno de R₁ y R₂, preferiblemente R₂ es hidroxilo y el otro es metoxi y uno de W, X e Y, preferiblemente X es azufre, selenio o BH₃, preferiblemente azufre mientras que los otros son oxígeno.

En la fórmula anterior, el nucleótido en el lado derecho está conectado a la cadena de ARN a través de su grupo 3'.

Aquellas estructuras de Cap en las que al menos uno de W, X e Y es azufre, es decir, que tienen una fracción fosforotioato, existen en diferentes formas diastereoisoméricas, todas las cuales están abarcadas en el presente documento. Además, la presente invención abarca todos los tautómeros y estereoisómeros de la fórmula anterior.

Por ejemplo, la estructura de Cap que tiene la estructura anterior en la que R₁ es metoxi, R₂ es hidroxilo, X es azufre y W e Y son oxígeno existe en dos formas diastereoisoméricas (Rp y Sp). Estos pueden resolverse mediante HPLC de fase inversa y se denominan D1 y D2 de acuerdo con su orden de elución de la columna de HPLC de fase inversa. De acuerdo con la invención, el isómero D1 de m₂^{7,2'-O}GppspG es particularmente preferido.

El suministro de un ARN con una caperuza 5' o un análogo de caperuza 5' se puede lograr mediante la transcripción *in vitro* de una plantilla de ADN en presencia de dicha caperuza 5' o un análogo de caperuza 5', en el que dicha caperuza 5' se incorpora cotranscripcionalmente en la cadena de ARN generada, o el ARN se puede generar, por ejemplo, mediante transcripción *in vitro*, y la caperuza 5' se puede unir al ARN después de la transcripción usando enzimas de protección, por ejemplo, enzimas de protección del virus vacuna.

El ARN puede comprender modificaciones adicionales. Por ejemplo, una modificación adicional del ARN usado en la presente invención puede ser una extensión o truncamiento de la cola de poli(A) natural o una alteración de las regiones no traducidas (UTR) 5' o 3', tal como la introducción de un 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 uno o más, preferiblemente dos copias de una 3'-UTR derivada de un gen de globina, tales como alfa2-globina, alfa-globina, beta-

globina, preferiblemente beta-globina, más preferiblemente beta-globina humana.

El ARN que tiene una secuencia de poli-A no enmascarada se traduce más eficientemente que el ARN que tiene una secuencia de poli-A enmascarada. El término "cola de poli(A)" o "secuencia de 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 una "secuencia de poli-A no enmascarada" significa que la secuencia de poli-A en el extremo 3' de una molécula de ARN termina con una A de la secuencia de poli-A y no está seguida de nucleótidos distintos de A ubicados en el extremo 3', es decir, secuencia abajo, de la secuencia poli-A. Además, una secuencia larga de poli-A de aproximadamente 120 pares de bases da como resultado una estabilidad de transcrito óptima y eficiencia de traducción de ARN.

Por lo tanto, para aumentar la estabilidad y/o la expresión del ARN usado de acuerdo con la presente invención, puede modificarse para que esté presente junto con una secuencia de poli-A, que preferiblemente tiene 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 utilizado de acuerdo con la invención, la secuencia de poli-A puede desenmascararse.

Además, la incorporación de una región 3' no traducida (UTR) en la región 3' no traducida de una molécula de ARN puede dar como resultado una mejora en la eficiencia de la traducción. Se puede lograr un efecto sinérgico incorporando dos o más de tales 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 en la estabilidad del ARN y aumento en la eficiencia de la traducción.

El término "estabilidad" de ARN se refiere a la "vida media" de 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 invención, 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 invención 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 como se describió anteriormente, aumentando la estabilidad y/o eficiencia de traducción del ARN.

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

De acuerdo con la invención, el término expresión también incluye una "expresión aberrante" o "expresión anormal". "Expresión aberrante" o "expresión anormal" significa de acuerdo con la invención que la expresión está alterada, preferiblemente aumentada, 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 en la expresión se refiere a un aumento de al menos 10%, en particular al menos 20%, al menos 50% o al menos 100%, o más. En una realización, la expresión solo se encuentra en un tejido enfermo, mientras que la expresión en un tejido sano se reprime.

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. Por lo tanto, una proteína que se expresa exclusivamente en las células de la mucosa gástrica y, en menor medida, en cualquier otro tejido, como los testículos, se expresa específicamente en las 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 tejido u órganos, pero preferiblemente en no más de 3 tipos diferentes de tejido u órgano. 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 un grado aproximadamente igual en pulmón y estómago, dicho antígeno tumoral se expresa específicamente en pulmón y estómago.

En el contexto de la presente invención, el término "transcrito" 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 invención, el término "transcrito" comprende "transcrito *in vitro*", en la que el término "transcrito *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 transcritos. Estos vectores de clonación se designan generalmente como vectores de transcrito y están

de acuerdo con la presente invención abarcados por el término "vector". De acuerdo con la presente invención, el ARN utilizado en la presente invención preferiblemente es ARN transcrito *in vitro* (ARN-IVT) y puede obtenerse mediante la transcripción *in vitro* de una plantilla de ADN apropiada. El promotor para controlar la transcripción puede ser cualquier promotor para 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 invención está controlada por un promotor T7 o SP6. Se puede obtener 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 se puede obtener por transcripción inversa de ARN.

El término "traducción" de acuerdo con la invención 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 formar un péptido o polipéptido.

Las secuencias de control de expresión o secuencias reguladoras, que de acuerdo con la invención pueden estar unidas funcionalmente con un ácido nucleico, pueden ser homólogas o heterólogas con respecto al ácido nucleico. Una secuencia de codificación y una secuencia reguladora están unidas "funcionalmente" si están unidas de forma covalente, 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 de codificación se va a traducir en una proteína funcional, con el enlace funcional de una secuencia reguladora con la secuencia de codificación, la inducción de la secuencia reguladora conduce a una transcripción de la secuencia de codificación, sin causar un cambio en el marco de lectura en la secuencia de codificación o incapacidad de la secuencia de codificación para traducirse en la proteína o péptido deseado.

El término "secuencia de control de expresión" o "secuencia reguladora" comprende, de acuerdo con la invención, 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 ciertas realizaciones de la invención, las secuencias reguladoras pueden controlarse. La estructura precisa de las secuencias reguladoras puede variar de acuerdo con la especie o el 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, tal como la caja TATA, una secuencia de protección, secuencia de CAAT y similares. En particular, las secuencias reguladoras 5' no transcritas comprenden una región promotora que incluye una secuencia promotora para el control transcripcional del gen funcionalmente unido. Las secuencias reguladoras también pueden comprender secuencias potenciadoras o secuencias activadoras secuencia arriba.

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

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

Términos 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 invención, la célula puede formar parte de un órgano, un tejido y/o un organismo. De acuerdo con la presente invención, 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 está en forma de ácidos nucleicos desnudos. Preferiblemente, el ARN se administra en combinación con sustancias estabilizantes tales como inhibidores de RNasa. La presente invención también prevé la introducción repetida de ácidos nucleicos en las células para permitir la expresión sostenida durante períodos de tiempo prolongados.

Las células se pueden transfectar con cualquier vehículo 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 resulta en una mayor estabilidad del ARN en comparación con el ARN desnudo. Los portadores útiles de acuerdo con la invención incluyen, por ejemplo, portadores 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 invención.

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 vehículo 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 superficial en la célula objetivo o un ligando para un receptor en la célula objetivo puede incorporarse al vehículo de ácido nucleico o puede unirse al mismo. En caso de que el ácido nucleico sea administrado por liposomas, las proteínas que se unen a una proteína de membrana superficial que está asociada con la endocitosis pueden incorporarse en la formulación de liposomas para permitir la selección y/o captación. Dichas proteínas abarcan

proteínas de la cápside de fragmentos de las mismas que son específicas para un tipo de célula particular, anticuerpos contra proteínas que se internalizan, proteínas que se dirigen a una ubicación intracelular, etc.

De acuerdo con la presente invención, 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 o 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 el presente documento.

10 De acuerdo con la invención, el término "cambio de secuencia" con respecto a péptidos o proteínas se refiere a 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 acuerdo con la invención pueden potencialmente crear nuevos epítos.

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

Las variantes de adición de aminoácidos comprenden fusiones en los terminales 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 por que al menos un residuo en la secuencia se elimina y otro residuo se inserta en su lugar.

El término "derivado" significa de acuerdo con la invención que una entidad particular, en particular una secuencia particular, está presente en el objeto del que se deriva, en particular un organismo o molécula. En el caso de secuencias de aminoácidos, especialmente regiones de secuencias particulares, "derivado" en particular significa que la secuencia de aminoácidos relevante se deriva de una secuencia de aminoácidos en la que está presente.

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 invención a cualquier célula que pueda transformarse o transfectarse con un ácido nucleico exógeno. El término "célula" incluye de acuerdo con la invención 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 el ADN mitocondrial. Las células de mamífero son particularmente preferidas, tal como las células de humanos, ratones, hámsteres, cerdos, cabras y primates. Las células pueden derivarse de una gran cantidad de tipos de tejidos e incluyen células primarias y líneas celulares. Los ejemplos específicos incluyen queratinocitos, leucocitos de sangre periférica, células madre de 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 un macrófago.

40 Una célula que comprende una molécula de ácido nucleico preferiblemente expresa 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 invención, el término se usa preferiblemente en el contexto de una respuesta inmunológica en la que los linfocitos son estimulados por un antígeno, proliferan, y el linfocito específico que reconoce dicho antígeno se amplifica. Preferiblemente, la expansión clonal conduce a la diferenciación de los linfocitos.

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

50 Los términos tales como "aumentar", "mejorar", "promover" o "prolongar" se refieren preferiblemente a un aumento, mejora, promoción o prolongación de aproximadamente al menos 10%, preferiblemente al menos 20%, preferiblemente al menos 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 relacionarse con un aumento, mejora, promoción o prolongación desde cero o un nivel no medible o no detectable a un nivel superior a cero o un nivel medible o detectable.

55

- Los agentes, composiciones y métodos descritos en el presente documento pueden usarse 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 péptido antigénico. 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.
- De acuerdo con la invención, el término "enfermedad" se refiere a cualquier estado patológico, incluyendo enfermedades cancerosas, en particular aquellas formas de enfermedades cancerosas descritas en el presente documento.
- El término "normal" se refiere al estado saludable o las condiciones en un sujeto o tejido sano, es decir, condiciones no patológicas, en las que "sano" significa preferiblemente no canceroso.
- "Enfermedad que implica células que expresan un antígeno" significa de acuerdo con la invención 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 10%, en particular al menos 20%, al menos 50%, al menos 100%, al menos 200%, al menos 500%, al menos 1000%, al menos 10000% o incluso más. En una realización, la expresión solo se encuentra en un tejido enfermo, mientras que la expresión en un tejido sano se reprime. De acuerdo con la invención, las enfermedades que implican o están asociadas con células que expresan un antígeno incluyen enfermedades cancerosas.
- El cáncer (término médico: neoplasia maligna) es una clase de enfermedades en las que un grupo de células muestra un crecimiento incontrolado (división más allá de los límites normales), invasión (intrusión y destrucción de tejidos adyacentes) y, a veces, metástasis (propagación a otras ubicaciones en el cuerpo a través de la linfa o la sangre). Estas 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.
- El tumor maligno es esencialmente sinónimo de cáncer. Malignidad, neoplasia maligna y tumor maligno son esencialmente sinónimos de cáncer.
- De acuerdo con la invención, 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 lesión. Por "célula tumoral" se entiende una célula anormal que crece mediante una proliferación celular rápida e incontrolada y continúa creciendo después de que cesan los estímulos que iniciaron el nuevo crecimiento. Los tumores muestran una falta parcial o completa de organización estructural y coordinación funcional con el tejido normal, y generalmente forman una masa de tejido distinta, que puede ser benigna, premaligna o maligna.
- Un tumor benigno es un tumor que carece de las tres propiedades malignas de un cáncer. Por lo tanto, por definición, un tumor benigno no crece de manera ilimitada y agresiva, no invade los tejidos circundantes y no se propaga a los tejidos no adyacentes (metástasis).
- Un neoplasma es una masa anormal de tejido como resultado de la 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 tejidos normales a su alrededor. El crecimiento persiste de la misma manera excesiva incluso después del cese de los estímulos. Generalmente causa un bulto o tumor. Las neoplasias pueden ser benignas, premalignas o malignas.
- El "crecimiento de un tumor" o "crecimiento tumoral" de acuerdo con la invención 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 fines de la presente invención, los términos "cáncer" y "enfermedad cancerosa" se usan indistintamente con los términos "tumor" y "enfermedad tumoral".
- Los cánceres se clasifican por el tipo de célula que se asemeja al tumor y, por lo tanto, al tejido que se supone que es el origen del tumor. Estas son la histología y la ubicación, respectivamente.
- El término "cáncer" de acuerdo con la invención comprende leucemias, seminomas, melanomas, teratomas, linfomas, neuroblastomas, gliomas, cáncer rectal, cáncer de endometrio, cáncer de riñón, cáncer suprarrenal, cáncer de tiroides, cáncer de sangre, cáncer de piel, cáncer de cerebro, cáncer cervical, cáncer intestinal, cáncer de hígado, cáncer de colon, cáncer de estómago, cáncer de intestino, cáncer de cabeza y cuello, cáncer gastrointestinal, cáncer de ganglios linfáticos, cáncer de esófago, cáncer colorrectal, cáncer de páncreas, cáncer de oído, nariz y garganta (ENT), cáncer de mama, cáncer de próstata, cáncer de útero, 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 cervicales o metástasis de los tipos de cáncer o tumores descritos anteriormente. El término cáncer de acuerdo con la invención también comprende metástasis de cáncer y recaída de cáncer.
- Los principales tipos de cáncer de pulmón son el carcinoma de pulmón de células pequeñas (SCLC) y el carcinoma de pulmón de células no pequeñas (NSCLC). Existen tres subtipos principales de carcinomas de pulmón de células

no pequeñas: carcinoma de pulmón de células escamosas, adenocarcinoma y carcinoma de pulmón de células grandes. Los adenocarcinomas representan aproximadamente el 10% de los cánceres de pulmón. Este cáncer generalmente se ve periféricamente en los pulmones, a diferencia del cáncer de pulmón de células pequeñas y el cáncer de pulmón de células escamosas, que tienden a estar ubicados más centralmente.

- 5 El cáncer de piel es un crecimiento maligno en la piel. Los cánceres de piel más comunes son el cáncer de células basales, el cáncer de células escamosas y el melanoma. El melanoma maligno es un tipo grave de cáncer de piel. Se debe al crecimiento incontrolado de las células pigmentarias, llamadas melanocitos.

De acuerdo con la invención, un "carcinoma" es un tumor maligno derivado de células epiteliales. Este grupo representa los cánceres más comunes, incluidas las formas comunes de cáncer de mama, próstata, pulmón y colon.

- 10 El "carcinoma bronquiolar" es un carcinoma de pulmón, que se cree que deriva del epitelio de los bronquiolos terminales, en el que el tejido neoplásico se extiende a lo largo de las paredes alveolares y crece en pequeñas masas dentro de los alvéolos. La mucina puede demostrarse en algunas de las células y en el material de los alvéolos, que también incluye células desnudas.

- 15 El "adenocarcinoma" es un cáncer que se origina en el tejido glandular. Este tejido también forma parte de una categoría de tejido más grande conocida como tejido epitelial. El tejido epitelial incluye piel, glándulas y una variedad de otros tejidos que recubren las cavidades y los órganos del cuerpo. El epitelio se deriva embriológicamente del ectodermo, el endodermo y el mesodermo. Para clasificarse como adenocarcinoma, las células no necesariamente deben ser parte de una glándula, siempre que tengan propiedades secretoras. Esta forma de carcinoma puede ocurrir en algunos mamíferos superiores, incluidos los humanos. Los adenocarcinomas bien diferenciados tienden a parecerse al tejido glandular del que derivan, mientras que los mal diferenciados pueden no serlo. Al teñir las células de una biopsia, un patólogo determinará si el tumor es un adenocarcinoma o algún otro tipo de cáncer. Los adenocarcinomas pueden surgir en muchos tejidos del cuerpo debido a la naturaleza ubicua de las glándulas dentro del cuerpo. Si bien es posible que cada glándula no esté secretando la misma sustancia, siempre que haya una función exocrina en la célula, se considera glandular y, por lo tanto, su forma maligna se denomina adenocarcinoma. Los adenocarcinomas malignos invaden otros tejidos y a menudo hacen metástasis con el tiempo suficiente para hacerlo. El adenocarcinoma de ovario es el tipo más común de carcinoma de ovario. Incluye los adenocarcinomas serosos y mucinosos, el adenocarcinoma de células claras y el adenocarcinoma endometriode.

- 20 El carcinoma de células renales también conocido como cáncer de células renales o adenocarcinoma de células renales es un cáncer de riñón que se origina en el revestimiento del túbulo contorneado proximal, los tubos muy pequeños en el riñón que filtran la sangre y eliminan los productos de desecho. El carcinoma de células renales es, con mucho, el tipo más común de cáncer de riñón en adultos y el más letal de todos los tumores genitourinarios. Los subtipos distintos de carcinoma de células renales son el carcinoma de células renales de células claras y el carcinoma de células renales papilares. El carcinoma de células renales de células claras es la forma más común de carcinoma de células renales. Cuando se observa al microscopio, las células que forman el carcinoma de células renales de células claras aparecen muy pálidas o claras. El carcinoma papilar de células renales es el segundo subtipo más común. Estos cánceres forman pequeñas proyecciones similares a dedos (llamadas papilas) en algunos, si no en la mayoría de los tumores.

El linfoma y la leucemia son tumores malignos derivados de células hematopoyéticas (formadoras de sangre).

- 40 El tumor blástico o blastoma es un tumor (generalmente maligno) que se asemeja a un tejido inmaduro o embrionario. Muchos de estos tumores son más comunes en niños.

- 45 Por "metástasis" se entiende la propagación 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 corporal y los vasos, y luego, después de ser transportado por la sangre, infiltración de órganos objetivo. Finalmente, el crecimiento de un nuevo tumor, es decir, un tumor secundario o tumor metastásico, en el sitio objetivo 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 invención se refiere a "metástasis a distancia" que se refiere a una metástasis que está alejada del tumor primario y del sistema regional de ganglios linfáticos.

- 50 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.

- 55 En el cáncer de ovario, la metástasis puede ocurrir de las siguientes maneras: por contacto directo o extensión, puede invadir tejidos u órganos cercanos ubicados cerca o alrededor del ovario, tales como las trompas de Falopio, el útero, la vejiga, el recto, etc.; al sembrar o arrojar a la cavidad abdominal, que es la forma más común de propagación del cáncer de ovario. Las células cancerosas se desprenden de la superficie de la masa ovárica y "caen" a otras estructuras del abdomen, como el hígado, el estómago, el colon o el diafragma; desprendiéndose de la masa ovárica,

invadiendo los vasos linfáticos y luego viajando a otras áreas del cuerpo u órganos distantes como el pulmón o el hígado; desprendiéndose de la masa ovárica, invadiendo el sistema sanguíneo y viajando a otras áreas del cuerpo u órganos distantes.

5 De acuerdo con la invención, el cáncer de ovario metastásico incluye cáncer en las trompas de Falopio, cáncer en órganos del abdomen, tales como cáncer en el intestino, cáncer en el útero, cáncer en la vejiga, cáncer en el recto, cáncer en el hígado, cáncer en el estómago, cáncer en el colon, cáncer en el diafragma, cáncer en los pulmones, cáncer en el revestimiento del abdomen o la pelvis (peritoneo) y cáncer en el cerebro. Del mismo modo, el cáncer de pulmón metastásico se refiere al cáncer que se ha diseminado desde los pulmones a sitios distantes y/o varios en el cuerpo e incluye cáncer en el hígado, cáncer en las glándulas suprarrenales, cáncer en los huesos y cáncer en el cerebro.

10 El término "células tumorales en circulación" o "CTC" se refiere a células que se han separado de un tumor primario o metástasis tumorales y circulan en el torrente sanguíneo. Las CTC pueden constituir semillas para el posterior crecimiento de tumores adicionales (metástasis) en diferentes tejidos. Las células tumorales en circulación se encuentran en frecuencias del orden de 1-10 CTC por mL de sangre completa en pacientes con enfermedad metastásica. Se han desarrollado métodos de investigación para aislar CTC. Se han descrito varios métodos de investigación en la técnica para aislar CTC, por ejemplo, técnicas que utilizan el hecho de que las células epiteliales comúnmente expresan la proteína de adhesión celular EpCAM, que está ausente en las células sanguíneas normales. La captura inmunomagnética basada en perlas implica el tratamiento de muestras de sangre con anticuerpos contra EpCAM que se ha conjugado con partículas magnéticas, seguido de la separación de las células marcadas en un campo magnético. Luego, las células aisladas se tiñen con anticuerpos contra otro marcador epitelial, la citoqueratina, así como con 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 de 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 el flujo de sangre completa a través de una cámara incrustada con 80.000 micropostes que se han vuelto funcionales al recubrirlos con anticuerpos contra EpCAM. Luego, las CTC se tiñen con anticuerpos secundarios contra citoqueratina o marcadores específicos de tejido, tales como PSA en cáncer de próstata o HER2 en 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 en circulación positivas para citoqueratinación en pacientes con un rendimiento medio de 50 células/mL y una pureza que oscila entre 1-80% (Nagrath et al., 2007: Nature 450, 1235-1239). Otra posibilidad para aislar las CTC es usar la prueba de células tumorales en circulación (CTC) CellSearch^{MR} de Veridex, LLC (Raritan, NJ) que captura, identifica y cuenta los 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 la enumeración de CTC en sangre completa que se basa en una combinación de etiquetado inmunomagnético y microscopía digital automatizada. Existen otros métodos para aislar CTC descritos en la literatura, todos los cuales se pueden usar junto con la presente invención.

Una recaída o recurrencia ocurre cuando una persona se ve afectada nuevamente por una afección que la afectó en el pasado. Por ejemplo, si un paciente ha sufrido una enfermedad tumoral, ha recibido un tratamiento exitoso de dicha enfermedad y nuevamente desarrolla dicha enfermedad, dicha enfermedad recientemente desarrollada puede considerarse una recaída o recurrencia. Sin embargo, de acuerdo con la invención, 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 ovárico y ha recibido un tratamiento exitoso, una recaída o recurrencia puede ser la aparición de un tumor ovárico 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 se produce 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 tal como se describe en el presente documento a un sujeto para prevenir o eliminar una enfermedad, que incluye reducir el 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 severidad de los síntomas y/o recurrencias en un sujeto que actualmente tiene o que previamente ha tenido una enfermedad; 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 empeoramiento, o prevenir o retrasar la aparición de una enfermedad o sus síntomas.

55 Por "estar en riesgo" se entiende un sujeto, es decir, un paciente, que se identifica que tiene 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 de cáncer.

60 El término "inmunoterapia" se refiere a un tratamiento que implica la activación de una reacción inmune específica. En el contexto de la presente invención, términos tales como "proteger", "prevenir", "profiláctico", "preventivo" o "protector"

se relacionan con la prevención o el tratamiento o la ocurrencia 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 un tumor, como se describió anteriormente, sería un candidato para la terapia para prevenir un tumor.

- 5 Una administración profiláctica de una inmunoterapia, por ejemplo, una administración profiláctica de la composición descrita en el presente documento, preferiblemente protege al receptor del desarrollo de una enfermedad. Una administración terapéutica de una inmunoterapia, por ejemplo, una administración terapéutica de la composición descrita en el presente documento, 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 de la
10 progresión 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 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.

- 15 En ciertas 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 se proporcionan en este documento).

- 20 Los agentes y composiciones proporcionados en el presente 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).

El término "inmunización" o "vacunación" describe el proceso de tratamiento de 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.

- 25 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 invención son humanos, primates no humanos, animales domesticados tal como perros, gatos, ovejas, vacas, cabras, cerdos, caballos, etc., animales de laboratorio tal como ratones, ratas, conejos, conejillos de indias, etc., así como animales en cautiverio, tal como
30 animales de zoológicos. El término "animal" como se usa en el presente 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.

El término "autólogo" se usa para describir cualquier cosa que se deriva 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.

- 35 El término "heterólogo" se usa para describir algo que consiste en múltiples elementos diferentes. Como ejemplo, la transferencia de la médula ósea de un individuo a un individuo diferente constituye un trasplante heterólogo. Un gen heterólogo es un gen derivado de una fuente distinta del sujeto.

- 40 Como parte de la composición para una inmunización o una vacuna, preferiblemente uno o más agentes como se describe en el presente 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 mejoran o aceleran una respuesta inmune. La composición descrita en el presente documento ejerce preferiblemente su efecto sin adición de adyuvantes. Aún así, la composición de la presente solicitud puede contener cualquier adyuvante conocido. Los adyuvantes comprenden un grupo heterogéneo de compuestos tales como emulsiones en aceite (por
45 ejemplo, adyuvantes de Freund), compuestos minerales (como el alumbre), productos bacterianos (como la toxina de *Bordetella pertussis*), liposomas y complejos inmunoestimulantes. Ejemplos de adyuvantes son lípido A monofosforilado (MPL Smithkline Beecham). Saponinas tales como QS21 (Smithkline Beecham), DQS21 (Smithkline Beecham; 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, montanida, alumbre, oligonucleótidos CpG (Krieg et al., 1995, Nature 374: 546-549) y varias emulsiones de agua en aceite que se preparan a partir de aceites biológicamente
50 degradables tales como el escualeno y/o el tocoferol.

También se pueden administrar otras sustancias que estimulan una respuesta inmune del paciente. Es posible, por ejemplo, usar citoquinas en una vacuna, debido a sus propiedades reguladoras en los linfocitos. Dichas citoquinas comprenden, por ejemplo, interleuquina 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 vacuna. Dichos compuestos comprenden moléculas coestimulantes proporcionadas en forma de proteínas o ácidos nucleicos

tales como B7-1 y B7-2 (CD80 y CD86, respectivamente).

De acuerdo con la invención, una "muestra tumoral" es una muestra tal como una muestra corporal que contiene células tumorales o cancerosas tales como células tumorales en circulación (CTC), en particular una muestra de tejido, que incluye fluidos corporales, y/o una muestra celular. De acuerdo con la invención, un "especimen no tumoral" es una muestra tal como una muestra corporal que no contiene células tumorales o cancerosas tales como células tumorales en circulación (CTC), en particular una muestra de tejido, que incluye fluidos corporales, y/o una muestra celular. Tales muestras corporales se pueden obtener de la manera convencional, tal como una biopsia de tejido, incluida una biopsia por punción, y tomando sangre, aspirado bronquial, esputo, orina, heces u otros fluidos corporales. De acuerdo con la invención, el término "muestra" también incluye muestras procesadas tales como fracciones o aislados de muestras biológicas, por ejemplo, ácido nucleico o aislados celulares.

Los agentes terapéuticamente activos, las vacunas y las composiciones descritas en el presente documento pueden administrarse a través de cualquier ruta convencional, incluyendo por inyección o infusión. La administración puede llevarse a cabo, por ejemplo, por vía oral, intravenosa, intraperitoneal, intramuscular, subcutánea o transdérmica. En una realización, la administración se lleva a cabo en forma intranodal, tal como 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 ácidos nucleicos descritos en este documento, seguido de la administración de las células presentadoras de antígeno.

Los agentes descritos en este documento se administran en cantidades efectivas. Una "cantidad efectiva" 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 la demora del inicio o la prevención de la aparición de dicha enfermedad o afección.

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

De acuerdo con la presente invención, las composiciones farmacéuticas son preferiblemente estériles y contienen una cantidad eficaz de la sustancia terapéuticamente activa para generar la reacción deseada o el efecto deseado.

De acuerdo con la presente invención, las composiciones farmacéuticas 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 generalmente pueden contener sales, sustancias reguladoras, conservantes, vehículos, sustancias que complementan la inmunidad, tal como adyuvantes, por ejemplo oligonucleótidos CpG, citoquinas, quimiocinas, saponina, GM-CSF y/o ARN y, cuando corresponda, otros compuestos terapéuticamente activos. Cuando se usa 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 se incluyen de acuerdo con la invención. Las sales farmacológica y farmacéuticamente compatibles de este tipo comprenden de forma no limitativa las preparadas a partir de los siguientes ácidos: ácidos clorhídrico, bromhídrico, sulfúrico, nítrico, fosfórico, maleico, acético, salicílico, cítrico, fórmico, malónico, succínico y similares. Las sales farmacéuticamente compatibles también se pueden preparar como sales de metales alcalinos o sales de metales alcalinotérreos, tales como sales de sodio, sales de potasio o sales de calcio.

De acuerdo con la presente invención, una composición farmacéutica 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 invención, el término "vehículo farmacéuticamente compatible" incluye uno o más rellenos, diluyentes o sustancias encapsulantes sólidos o líquidos compatibles, que son adecuados para la administración a un paciente. De acuerdo con la presente invención, los componentes de la composición farmacéutica son usualmente tales que no se produce interacción que perjudique sustancialmente la eficacia farmacéutica deseada.

De acuerdo con la presente invención, las composiciones farmacéuticas pueden contener sustancias reguladoras adecuadas tales como ácido acético en una sal, ácido cítrico en una sal, ácido bórico en una sal y ácido fosfórico en una sal.

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

Las composiciones farmacéuticas se proporcionan habitualmente en una forma de dosificación uniforme y se pueden

preparar de una manera ya conocida. De acuerdo con la presente invención, las composiciones farmacéuticas pueden estar en forma de cápsulas, tabletas, pastillas, soluciones, suspensiones, jarabes, elixires o en forma de una emulsión, por ejemplo.

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

10 La presente invención se describe en detalle por las figuras y ejemplos a continuación, que se usan solo con fines ilustrativos y no pretenden ser limitantes. Debido a la descripción y los ejemplos, otras formas de realización que también se incluyen en la invención son accesibles para el experto en la materia.

Figuras

Figura 1:

Arriba: Proceso para descubrir y priorizar probables mutaciones somáticas inmunogénicas en muestras tumorales masivas. Abajo: proceso aplicado al sistema B16 y Black6.

15 Figura 2: Ejemplo de mutación validada en Kif18b

Una mutación identificada en el gen Kif18b por secuenciación del exoma de NGS que fue confirmada por secuenciación de Sanger. En las células de tipo silvestre, la secuencia es T/T. En las células tumorales, la secuencia es una mezcla de T/G.

Figura 3: reactividad inmunológica contra secuencias mutadas

20 Los ratones (n = 5) se inmunizaron dos veces (d0, d7) con secuencias de péptidos mutados (100 µg + 50 µg de Poli I:C; s.c.). En el día 12 se sacrificaron los ratones y se recogieron las células del bazo. El ELISpot de IFNγ se realizó utilizando 5x10⁵ células de bazo/pozo como efectores y 5x10⁴ células dendríticas de médula ósea cargadas con péptidos (2 µg/mL durante 2 horas a 37 °C y 5% de CO₂) como células objetivo. Las células efectoras del bazo se probaron contra el péptido mutado, el péptido de tipo silvestre y un péptido de control (nucleoproteína del virus de la estomatitis vesicular, VSV-NP, aa 52-59). Se muestra el número medio de puntos medidos del que se restaron los puntos de fondo contra VSV-NP para cada ratón (círculos vacíos: ratones inmunizados con péptido de tipo silvestre; cajas llenas: ratones inmunizados con péptidos mutados). Los datos se muestran para cada ratón y se representa la media ± SEM.

25

30 Figura 4: Beneficio de supervivencia para ratones vacunados con la secuencia de péptidos mutados recientemente identificada

Las células B16F10 (7,5 x 10⁴) se inocularon subcutáneamente en d0. Los ratones se vacunaron con el péptido 30 (Jerini Peptide Technologies (Berlín); 100 µg de péptido + 50 µg de Poli I:C s.c. (Invitrogen)) el día -4, el día +2, el día +9. El grupo de control recibió solo Poli I:C (50 µg s.c.). El crecimiento tumoral se controló hasta el día +16 *, p <0,05 en la prueba Log-Rank (Mantel-Cox).

35 Figura 5:

(A) Ejemplos de expresión de proteínas mejorada (izquierda eGFP, derecha Luciferasa) con ARN optimizado para estabilidad y eficiencia traduccional (B) Ejemplo de expansión poliepitópica de células T CD8+ y CD4+ específicas de antígeno con ARN optimizado para enrutamiento efectivo del antígeno (s. Referencia Kreiter, Konrad, Sester et al., Cancer Immunol. Immunother. 56: 1577-1587, 2007). T (C) Ejemplo de una prueba preclínica de eficacia antitumoral en el modelo de melanoma B16 utilizando una vacuna de ARN que codifica un solo epítipo (OVA-SIINFEKL). Se obtuvieron datos de supervivencia para ratones tratados con vacuna sola o vacuna en combinación con adyuvante. (D) Diseño de vacuna individualizada, poli-neoepitópica. El vehículo de la vacuna integra elementos funcionales para una mayor expresión y una inmunogenicidad optimizada. Se pueden integrar hasta 30 epítipos mutados por molécula que están separados por enlazadores en su contexto de secuencia natural.

40

45 Figura 6: Diseño del constructo

(A) Diagrama esquemático de un constructo del poliepitopo de ARN. Cap: análogo de caperuza; 5'UTR: región no traducida 5'; L: enlazador; Seq. 1: secuencia de ARN que codifica el péptido que contiene aa mutados; 3'UTR: secuencia no traducida 3'; poli-A: cola de poli-A. (B) Secuencia de los constructos de ARN que codifican 2 secuencias de aa que incluyen una aa mutada de B16F10. El codón de inicio y parada, así como el péptido señal y la secuencia MITD no forman parte de la figura esquemática que se simboliza mediante "...".

50

Figura 7: Funcionalidad del poli epítipo de ARN

(A-C) Datos para ELISpot de IFNγ utilizando 5 x 10⁵ células de bazo por pozo como efectores y 5 x 10⁴ BMDC como

células objetivo. Las BMDC se cargaron con péptido (2 µg/mL durante 2 horas a 37 °C y 5% de CO₂) o se transfectaron con ARN (20 µg) por electroporación. El ARN de control era eGFP (panel izquierdo) o un constructo de ARN que codifica para 2 péptidos no relacionados que contienen aa mutados, separados por un enlazador. Los datos se muestran como media ± SEM. (A) Se muestran los datos para el péptido de mutación 30, el péptido de tipo silvestre 30 y la codificación de ARN para la mutación 30 y 31. (B) Se muestran los datos para el péptido de mutación 12, el péptido de tipo silvestre 12 y la codificación de ARN para la mutación 12 y 39. (C) Se representa el barrido de ELISpot representativo de un solo ratón de la lectura que se muestra en (B).

Figura 8: Dos realizaciones de vacunas de ARN poli-neoepitópico que muestran epítomos de unión

La vacuna de ARN puede construirse con (arriba) o sin enlazadores (abajo) entre péptidos que codifican mutaciones. Los buenos epítomos incluyen aquellos que incluyen la mutación somática ("**") y se unen a las moléculas del MHC. Los epítomos malos incluyen epítomos que se unen a moléculas del MHC pero contienen partes de dos péptidos (abajo) o partes de secuencias de péptidos y enlazadores (arriba).

Figura 9: Descubrimiento y caracterización del "mutanoma farmacológico de células T"

(A) El diagrama de flujo ofrece una visión general del procedimiento experimental a partir de las muestras B16F10 y C57BL/6 para la lectura ELISPOT. (B) Se muestra el número de resultados para cada etapa de evaluación y el proceso para la selección de mutaciones para la validación del ADN y las pruebas de inmunogenicidad. Las mutaciones seleccionadas para la validación y las pruebas de inmunogenicidad fueron aquellas predichas como inmunogénicas y en genes expresados en RPKM > 10. (C) El mutanoma farmacológico de células T se mapeó en el genoma de B16F10. Los anillos de afuera hacia adentro representan los siguientes subconjuntos: (1) presentes en todos los triplicados, (2) tienen una FDR < 0,05, (3) están ubicados en regiones de codificación de proteínas, (4) causan cambios no sinónimos, (5) están localizados en genes expresados, y (6) están en el conjunto validado. Cromosomas de ratón (círculo externo), densidad génica (verde), expresión génica (verde (bajo)/amarillo/rojo (alto)) y mutaciones somáticas (naranja).

Figura 10: respuestas inmunitarias provocadas *in vivo* por vacunación de ratones con mutación que representa péptidos sintéticos largos

(A, B) Análisis ELISPOT de IFN-γ de efectores de células T de ratones vacunados con péptidos codificadores de mutación. Las columnas representan medias (± SEM) de 5 ratones por grupo. Los asteriscos indican diferencias estadísticamente significativas de reactividad contra la mutación y el péptido de tipo silvestre (prueba t de Student; valor p < 0,05). (A) Los esplenocitos de ratones vacunados se estimularon nuevamente con BMDC transfectadas con el péptido codificador de mutación utilizado para la vacunación, el péptido de tipo silvestre correspondiente y un péptido de control irrelevante (VSV-NP). (B) Para el análisis de la reactividad de las células T frente a mutaciones procesadas endógenamente, se estimularon nuevamente esplenocitos de ratones vacunados con BMDC transfectadas con ARN de control (eGFP) o un ARN que codifica la mutación acusada. (C) Mutación 30 (gen Kif18B, proteína Q6PFD6, mutación p.K739N). Rastro de secuenciación de Sanger y secuencia de mutación (arriba). Dominios de proteínas y ubicación de la mutación (abajo).

Figura 11: Efectos antitumorales de vacunas peptídicas mutadas en ratones con tumores B16F10 de crecimiento agresivo

(A) Se inocularon ratones C57BL/6 (n = 7) con 7,5 x 10⁴ células B16F10 s.c., en el costado de los ratones. En los días 3 y 10 después de la inoculación del tumor, los ratones se vacunaron con 100 µg de péptido MUT30 o MUT44 + 50 µg de poli (I:C) o con adyuvante solo. (B) Los ratones C57BL/6 (n = 5) recibieron una inmunización de 100 µg de péptido MUT30 + 50 µg de poli (I:C) el día -4. El día 0 se inocularon 7,5 x 10⁴ células B16F10 s.c., en el costado de los ratones. Las inmunizaciones de refuerzo con péptido MUT30 (+ poli (I:C)) se realizaron los días 2 y 9. Transferencia de supervivencia de Kaplan-Meier (izquierda). Cinética de crecimiento tumoral (derecha).

Figura 12: La vacunación con ARN que codifican la mutación conduce a respuestas de células T CD4⁺ y CD8⁺

Datos de análisis de tinción intracelular de citoquinas para IFN-γ en efectores de células T CD4⁺ y CD8⁺ de ratones vacunados con ARN que codifica la mutación. Los ARN codificaban 1 (monoepítomo, fila superior), 2 (biepítomo, fila central) o 16 (poliepítomo, fila inferior) mutaciones diferentes. Los puntos representan medias de 3 ratones por grupo. Los asteriscos indican diferencias estadísticamente significativas de reactividad contra la mutación y el péptido de control (VSV-NP) (prueba t de Student; valor p < 0,05). Las gráficas de FACS muestran efectores del animal que secreta IFN-γ más alto para cada mutación e indican el fenotipo de la respuesta de las células T.

Figura 13: La vacunación con mutación que codifica el ARN de poliepítomo conduce las respuestas de las células T contra varias mutaciones

Análisis ELISPOT de IFN-γ de efectores de células T de ratones vacunados con mutación que codifica poliepítomo que incluye 16 mutaciones diferentes. Las columnas representan medias (± SEM) de 3 ratones por grupo. La fotografía muestra los pozos por triplicado de células de un ejemplo de animal reestimulado con los péptidos indicados.

Figura 14: La vacunación con 5 epítomos modelo diferentes codificados por un ARN conduce a respuestas inmunes

contra todos los epítomos codificados

- 5 A) Análisis ELISPOT de IFN- γ de efectores de células T de ratones vacunados con el modelo de codificación de mutación de poliepítomo que incluye 5 diferentes epítomos modelo (SIINFEKL, Trp2, VSV-NP, Inf-NP, OVA clase II). Los esplenocitos se reestimularon con los péptidos indicados. Las manchas representan las medias de pozos por triplicado de 5 ratones por grupo. B) Tinción con pentámero de linfocitos de sangre de un ratón de control y un ratón inmunizado con el poliepítomo modelo. Las células CD8⁺ teñidas con pentámero Inf-NP son específicas para el péptido Inf-NP.

Figura 15: Una mutación inductora de células T CD4⁺ puede inducir un potente efecto antitumoral del melanoma B16F10 en sinergia con un epítomo débil de células T CD8⁺

- 10 Se inocularon ratones C57BL/6 (n = 8) con 1 x 10⁵ células B16F10 s.c., en el costado de los ratones. El día 3, 10 y 17 después de la inoculación del tumor, los ratones fueron vacunados con 100 μ g de MUT30, Trp2 o ambos péptidos + 50 μ g de poli (I:C). A) Se muestran las cinéticas medias de crecimiento tumoral de cada grupo. El día 28, los valores medios entre los grupos de tratamiento único y los animales no tratados y el grupo de combinación son estadísticamente diferentes (prueba de Mann-Whitney, valor p <0,05). B) Gráfico de supervivencia de Kaplan-Meyer de los diferentes grupos. Las curvas de supervivencia de los ratones vacunados con MUT30 y MUT30 + Trp2 son estadísticamente diferentes (prueba Log-Rank, valor p = 0,0029).

Figura 16: Descripción general del proceso para encontrar mutaciones somáticas en B16

- 20 Los números para las etapas individuales se dan como ejemplo para una muestra de B16, en comparación con una muestra black6. "Exones" se refiere a las coordenadas del exón definidas por todos los transcritos de RefSeq que codifican proteínas.

Figura 17: Diagrama de Venn que muestra el número de variaciones somáticas en los exones de codificación de proteínas, encontradas por las herramientas de software individual, dos o las tres, respectivamente

Los números se calcularon después del filtrado y representan el consenso de las tres muestras.

- 25 Figura 18: **A** Ejemplos de variaciones de un solo nucleótido encontradas: una mutación somática encontrada en las tres muestras de B16 (izquierda), una mutación no somática encontrada en todas las muestras de B16 y black6 (centro) y una mutación encontrada en una sola muestra de black6 (derecha) **B** La distribución FDR calculada para el conjunto de datos del cual se seleccionaron las mutaciones validadas; la distribución se visualiza como una curva ROC estimada promedio con las barras grises que dan el intervalo de confianza del 95% para la media en ambas dimensiones en posiciones muestreadas uniformemente. La media se obtuvo de la distribución de las curvas ROC estimadas de las FDR para todas las 18 combinaciones posibles (véase el texto).

- 30 Figura 19: **A** Curvas ROC estimadas para la comparación de las tres herramientas de software diferentes (duplicados, cobertura 38x). **B** Curvas ROC estimadas para la comparación de diferentes profundidades de secuenciación promedio (SAMtools, sin replicación). 38x denota la cobertura obtenida por el experimento, mientras que otras coberturas se redujeron a partir de estos datos. **C** Curvas ROC estimadas que visualizan el efecto de la replicación del experimento (cobertura 38x, SAMtools). **D** Curvas ROC estimadas para diferentes protocolos de secuenciación (SAMtools, sin replicación). Las curvas se calcularon utilizando los resultados de la biblioteca de 2x100 nt.

- 35 Figura 20: **A** Diez mutaciones validadas con las FDR más bajas, seleccionadas utilizando el conjunto óptimo de parámetros de un conjunto final de 2396 variaciones. Ninguna de estas mutaciones está presente en dbSNP (versión 128; ensamblaje del genoma mm9). **B** Cantidad relativa de variaciones encontradas en el mismo conjunto de datos que A para un corte de FDR dado, graficada por separado para todas las variantes en el conjunto de datos y las mutaciones validadas. Para mayor claridad visual, solo se muestran valores de 0 a 10% de FDR.

Figura 21: Actividad antitumoral de una vacuna de ARN de poliepítomo que codifica la mutación

- 45 Se inocularon ratones C57BL/6 (n = 10) con 1 x 10⁵ células B16F10 s.c., en el costado de los ratones. En los días 3, 6, 10, 17 y 21 después de la inoculación del tumor, los ratones fueron vacunados con un ARN poliepítomo que formuló un reactivo de transfección de ARN liposomal. El grupo control recibió liposomas sin ARN. La figura muestra el gráfico de supervivencia de Kaplan-Meyer de los diferentes grupos. Las curvas de supervivencia son estadísticamente diferentes (prueba Log-Rank, valor p = 0,0008).

Ejemplos

- 50 Las técnicas y métodos utilizados en este documento se describen aquí 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, 2nd Edition (1989) Cold Spring Harbor Laboratory Prensa, 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: detección de mutaciones y priorización

Primero se demostró el perfil de secuencia de muestras tumorales y normales para identificar mutaciones somáticas de manera imparcial. Se demostró que esto no es solo para muestras tumorales masivas sino que, por primera vez, se demostró la capacidad para identificar mutaciones de células tumorales en circulación individuales. A continuación, se priorizaron las mutaciones para su inclusión en una vacuna polineoepitópica basada en la inmunogenicidad predicha de la mutación y se demostró que las mutaciones identificadas son de hecho inmunogénicas.

Detección de mutaciones

La razón para usar CTC: la detección de células tumorales en circulación (CTC) de la sangre periférica de pacientes con cáncer es un marcador de pronóstico independiente reconocido para el curso clínico de los tumores (Pantel et al., Trends Mol Med 2010; 16 (9): 398-406). Durante muchos años, la importancia clínica de las CTC ha sido objeto de una intensa investigación científica y clínica en oncología. Se ha demostrado que la detección de CTC en la sangre de pacientes con cáncer metastásico de mama, próstata y colorrectal tiene relevancia pronóstica, proporcionando información adicional a las técnicas de imagenología convencionales y otros biomarcadores tumorales pronósticos. Las muestras de sangre secuenciales extraídas de un paciente antes, durante una etapa temprana y después del tratamiento con un agente terapéutico (sistémico o dirigido) proporcionan información sobre la respuesta/fracaso del tratamiento. El análisis molecular de las CTC resistentes a los medicamentos puede proporcionar una visión adicional de los mecanismos de resistencia (por ejemplo, mutaciones en vías de señalización específicas o pérdida de la expresión del objetivo) en pacientes individuales. Una posibilidad adicional del perfil y la caracterización genética de las CTC es la identificación de nuevos objetivos de cáncer para el desarrollo de nuevas terapias dirigidas. Esta nueva estrategia de diagnóstico se conoce como "Biopsia de tumor líquido". Como este perfil podría realizarse de forma rápida y repetitiva, requiriendo solo sangre del paciente y sin cirugía, esto proporcionaría una visión en "tiempo real" del estado del tumor.

Mutaciones de células tumorales: Se demostró la capacidad para identificar mutaciones usando células de melanoma B16, captura de exoma para extraer regiones codificantes de proteínas, secuenciación de próxima generación usando HiSeq 2000, seguido de análisis bioinformático usando un canal de información de software "iCAM" (Figura 1) se identificaron 2448 mutaciones no sinónimas y se seleccionaron 50 para confirmación. Se pudieron confirmar todas las 50 mutaciones somáticas.

El siguiente es un ejemplo del impacto de la proteína de una mutación somática descubierta en células de melanoma B16:

Kif18b, NM_197959, exón 3

Mutación (+15 aa)

SPSKPSFQEFVDWENVSPELNSTDQPFLPS

Tipo silvestre (+15 aa)

SPSKPSFQEFVDWEKVSPELNSTDQPFLPS

Mutaciones de células tumorales en circulación individuales (CTC): A continuación, se pudieron identificar mutaciones somáticas específicas de tumor a partir de perfiles NGS de ARN de CTC individuales. Las células de melanoma B16 marcadas se inyectaron por vía intravenosa en las colas de los ratones, se sacrificaron los ratones, se recogió sangre de los corazones, las células se clasificaron para recuperar las células B16 en circulación marcadas (CTC), se extrajo el ARN, se realizó una síntesis de ADNc basada en SMART y se realizó una amplificación inespecífica, seguida por el ensayo de Seq-ARN de NGS y análisis de datos de subsecuencia (a continuación).

Se perfilaron ocho CTC individuales y se identificaron mutaciones somáticas. Además, en ocho de ocho células, se identificaron mutaciones somáticas previamente identificadas. En múltiples casos, los datos mostraron heterogeneidad a nivel celular individual. Por ejemplo, en la posición 144078227 en el cromosoma 2 (ensamblaje mm9), en el gen Snx15, dos células mostraron el nucleótido de referencia (C) mientras que dos células mostraron el nucleótido mutado (T).

Esto demuestra que se pueden perfilar CTC individuales para identificar mutaciones somáticas, un camino fundamental hacia una iVAC (vacuna individualizada) en "tiempo real", en el que los pacientes se perfilan de forma repetitiva y los resultados reflejan el estado actual del paciente en lugar del estado en un punto de tiempo anterior. Además, esto demuestra que se pueden identificar mutaciones somáticas heterogéneas que están presentes en un subconjunto de células tumorales, lo que permite la evaluación de la frecuencia de mutación, tal como la identificación de mutaciones importantes y mutaciones raras.

Métodos

Muestras: para el experimento de perfilado, las muestras incluyeron muestras de cola de 5-10 mm de ratones C57BL/6 ("Black6") y células de melanoma murino B16F10 altamente agresivas ("B16"), que se derivan originalmente de ratones Black6.

- Se crearon células tumorales en circulación (CTC) usando células de melanoma B16 marcadas con fluorescencia. Las células B16 se resuspendieron en PBS y se añadió a las células un volumen igual de solución CFSE recién preparada (5 μ M en PBS). La muestra se mezcló suavemente por vórtice seguido de incubación durante 10 minutos a temperatura ambiente. Para detener la reacción de marcado, se añadió a la muestra la misma cantidad de PBS que contenía un 20% de FSC y se mezcló suavemente mediante vórtice. Después de 20 minutos de incubación a temperatura ambiente, las células se lavaron dos veces usando PBS. Finalmente, las células se resuspendieron en PBS y se inyectaron por vía intravenosa (i.v.) en ratones. Después de 3 minutos, los ratones fueron sacrificados y se recogió sangre.
- Se lisaron los eritrocitos de las muestras de sangre mediante la adición de 1,5 mL de solución PharmLyse recién preparada (Beckton Dickinson) por 100 μ L de sangre. Después de una etapa de lavado, se añadió 7-AAD a la muestra y se incubó durante 5 minutos a temperatura ambiente. La incubación fue seguida por dos etapas de lavado y la muestra se resuspendió en 500 μ L de PBS.
- Las células B16 en circulación marcadas con CFSE se clasificaron con un clasificador de células Aria I (BD). Las células individuales se clasificaron en placas de 96 pozos con fondo en V preparadas con 50 μ L/pozo de regulador RLT (Qiagen). Después de terminar la clasificación, las placas se almacenaron a -80 °C hasta que comenzó la extracción de ácido nucleico y la preparación de la muestra.
- Extracción de ácido nucleico y preparación de muestras: se extrajeron ácidos nucleicos de células B16 (ADN y ARN) y tejido de cola de Black6 (ADN) usando el kit de sangre y tejido DNeasy de Qiagen (ADN) y el kit de RNeasy Micro de Qiagen (ARN).
- Para las CTC individuales clasificadas, se extrajo ARN y se realizó una síntesis de ADNc basada en SMART y una amplificación inespecífica. El ARN de las células CTC clasificadas se extrajo con el Kit de RNeasy Micro (Qiagen, Hilden, Alemania) de acuerdo con las instrucciones del proveedor. Se usó un protocolo BD SMART modificado para la síntesis de ADNc: la transcriptasa inversa Mint (Evrogen, Moscú, Rusia) se combinó con un cebador oligo(dT)-T largo para cebar la reacción de síntesis de la primera cadena y TS corto (Eurogentec SA, Seraing, Bélgica) introduciendo una secuencia de oligo (riboG) para permitir la creación de una plantilla extendida por la actividad de la transferasa terminal de la transcriptasa inversa y para el cambio de plantilla [Chenchik, A., Y. et al., 1998. Generation and use of high quality cDNA from small amounts of total RNA by SMART PCR. En Gene Cloning and Analysis by RT-PCR. P. L. J. Siebert, ed. BioTechniques Books, MA, Natick. 305-319]. El ADNc de la primera cadena sintetizado de acuerdo con las instrucciones del fabricante se sometió a 35 ciclos de amplificación con 5 U de ADN polimerasa de alta fidelidad PfuUltra Hotstart (Stratagene, La Jolla, CA) y cebador TS-PCR 0,48 μ M en presencia de dNTP 200 μ M (condiciones de ciclado: 2 minutos a 95 °C durante 30 segundos a 94 °C, durante 30 segundos a 65 °C, 1 minuto a 72 °C, para extensión final de 6 minutos a 72 °C). La amplificación exitosa de los genes CTC se controló con cebadores específicos para controlar la actina y la GAPDH.
- Secuenciación de próxima generación, secuenciación de ADN: la captura del exoma para la resecuenciación de ADN se realizó usando el ensayo de captura basado en la solución Sure-Select de Agilent [Gnirke A et al.: Solution hybrid selection with ultra-long Oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009, 27: 182-189], en este caso diseñado para capturar todas las regiones de codificación de proteínas de ratón.
- En breve, se fragmentaron 3 μ g de ADN genómico purificado hasta 150-200 pb utilizando un dispositivo de ultrasonido Covaris S2. Los fragmentos de ADNg se repararon finalmente usando ADN polimerasa T4, ADN polimerasa Klenow y se fosforilaron en 5' usando polinucleótido quinasa T4. Los fragmentos de ADNg de extremos romos se adenilaron en 3' usando el fragmento de Klenow (3' a 5' exo menos). Los adaptadores del extremo 3' emparejados con salientes T individuales Illumina se ligaron a los fragmentos de ADNg usando una relación molar 10:1 de adaptador a inserto de ADN genómico usando ADN ligasa T4. Los fragmentos de ADNg ligados al adaptador se enriquecieron antes de la captura y se agregaron secuencias específicas de células de flujo usando los cebadores de PCR Illumina PE 1.0 y 2.0 y la Herculasa II polimerasa (Agilent) usando 4 ciclos de PCR.
- Se hibridaron 500 ng de fragmentos de ADNg enriquecidos por PCR ligados al adaptador a cebos de biblioteca de ARN de exoma completo de ratón biotinilado SureSelect de Agilent durante 24 horas a 65 °C. Los complejos de cebo hibridado de ADNg/ARN se eliminaron usando perlas magnéticas recubiertas con estreptavidina. Los complejos de cebo de ADNg/ARN se lavaron y los cebos de ARN se escindieron durante la elución en regulador de elución SureSelect, dejando los fragmentos de ADNg enriquecidos por PCR ligados al adaptador capturado. Los fragmentos de ADNg se amplificaron por PCR después de la captura usando ADN polimerasa Herculasa II (Agilent) y cebadores de PCR SureSelect GA durante 10 ciclos.
- Todas las limpiezas se realizaron con un volumen de 1,8x de perlas magnéticas AMPure XP (Agencourt). Todos los controles de calidad se realizaron con el ensayo Qubit HS de Invitrogen y el tamaño del fragmento se determinó con el ensayo de ADN 2100 Bioanalyzer HS de Agilent.
- Las bibliotecas de ADNg enriquecidas con exomas se agruparon en el cBot usando el kit de agrupación Truseq SR v2.5 usando 7 μ M y se secuenciaron 50 pb en el Illumina HiSeq2000 usando el kit Truseq SBS-HS de 50 pb.
- Secuenciación de próxima generación, secuenciación de ARN (Seq-ARN): se prepararon bibliotecas de ADNc Seq-

- ARNm con código de barras a partir de 5 µg de ARN total usando una versión modificada del protocolo Seq-ARNm de Illumina. El ARNm se aisló usando perlas magnéticas Oligo(dT) Seramag (Thermo Scientific). El ARNm aislado se fragmentó usando cationes divalentes y calor, dando como resultado fragmentos que oscilan entre 160-220 pb. El ARNm fragmentado se convirtió en ADNc usando cebadores aleatorios y SuperScriptIII (Invitrogen) seguido de la síntesis de la segunda cadena usando ADN polimerasa I y RNasaH. El ADNc se reparó al final con ADN polimerasa T4, ADN polimerasa Klenow y se fosforiló en 5' con polinucleótido quinasa T4. Los fragmentos de ADNc con extremos romos se adenilaron en 3' usando el fragmento de Klenow (3' a 5' exo menos). Los adaptadores específicos múltiplex 3' con salientes T individuales Illumina se ligaron usando una relación molar 10:1 de adaptador a inserto de ADNc usando ADN ligasa T4.
- Se purificaron las bibliotecas de ADNc y se seleccionó el tamaño a 200-220 pb usando el gel E-Gel SizeSelect al 2% (Invitrogen). El enriquecimiento, la adición de Illumina con índice de seis bases y secuencias específicas de células de flujo se realizó por PCR usando ADN polimerasa Phusion (Finnzymes). Todas las limpiezas se realizaron con un volumen 1,8x de bolas magnéticas Agencourt AMPure XP. Todos los controles de calidad se realizaron usando el ensayo Qubit HS de Invitrogen y el tamaño del fragmento se determinó usando el ensayo de ADN 2100 Bioanalyzer HS de Agilent.
- Las bibliotecas de Seq-ARN con código de barras se agruparon en el cBot usando el kit de agrupación Truseq SR v2.5 usando 7 pM y se secuenciaron 50 pb en el Illumina HiSeq2000 usando el kit Truseq SBS-HS de 50 pb.
- CTC: para el perfil de Seq-ARN de las CTC, se usó una versión modificada de este protocolo en la que se usaron 500-700 ng de ADNc amplificado SMART, se ligaron adaptadores finales emparejados y se realizó enriquecimiento por PCR usando cebadores de PCR Illumina PE 1.0 y 2.0.
- Análisis de datos de NGS, expresión génica: para determinar los valores de expresión, los resultados de las lecturas de secuencia de las muestras de ARN del Illumina HiSeq 2000 se procesaron previamente de acuerdo con el protocolo estándar de Illumina. Esto incluye el filtrado para lecturas de baja calidad y de desmultiplexación. Para el análisis del transcriptoma de Seq-ARN, las lecturas de secuencia se alinearon con la secuencia genómica de referencia [Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420, 520-562 (2002)] utilizando Bowtie (versión 0.12.5) [Langmead B. et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome*. *Genome Biol* 10: R25] usando los parámetros "-v2-best" para las alineaciones del genoma y los parámetros predeterminados para las alineaciones de transcritos. Las coordenadas de alineación se compararon con las coordenadas de exón de los transcritos RefSeq [Pruitt KD. et al., *NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins*. *Nucleic Acids Res.* de enero 1 de 2005; 33(publicación de la base de datos): D501-4] y para cada transcrito se registraron los recuentos de alineaciones superpuestas. Las lecturas de secuencia no alineables con la secuencia genómica se alinearon con una base de datos de todas las posibles secuencias de unión exón-exón de los transcritos de RefSeq. Los recuentos de lecturas que se alinean con las uniones de empalme se agregaron con los respectivos recuentos de transcritos obtenidos en la etapa anterior y se normalizaron a RPKM (número de lecturas que se mapean por kilobase de modelo de exón por millón de lecturas mapeadas [Mortazavi, A. et al., (2008 Mapping and quantifying mammalian transcriptomes by rna-seq. *Nat Methods*, 5 (7): 621-628]) para cada transcrito. Tanto la expresión génica como los valores de expresión de exón se calcularon en función del número normalizado de lecturas que se superponen a cada gen o exón, respectivamente.
- Descubrimiento de mutaciones, tumor en masa: 50 nt, extremo único, las lecturas de Illumina HiSeq 2000 se alinearon usando bwa (versión 0.5.8c) [Li H. y Durbin R. (2009) *Fast and accurate short read alignment with Burrows-Wheeler Transform*. *Bioinformatics*, 25: 1754-60] utilizando las opciones predeterminadas para el ensamblaje del genoma del ratón de referencia mm9. Las lecturas ambiguas, esas lecturas que se asignan a múltiples ubicaciones del genoma, se eliminaron, las alineaciones restantes se clasificaron, indexaron y convirtieron a un formato binario y comprimido (BAM) y los puntajes de calidad de lectura se convirtieron del puntaje phred+ 64 estándar de Illumina al puntaje de calidad estándar de Sanger utilizando programas de instrucciones de algoritmos.
- Para cada carril de secuenciación, se identificaron mutaciones usando tres programas de software: incluyendo SAMtools (versión 0.1.8) [Li H. *Improving SNP discovery by base alignment quality*. *Bioinformatics*. Abril 15 de 2011; 27 (8): 1157-8. Epub, febrero 13 de 2011], GATK (versión 1.0.4418) [McKenna A. et al. *The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data*. *Genome Res*. Septiembre de 2010; 20 (9): 1297-303. Epub Jul 19 de 2010], y SomaticSniper (<http://genome.wustl.edu/software/somaticsniper>). Para SAMtools, se utilizaron las opciones recomendadas por el autor y los criterios de filtro, incluido el filtrado de primera ronda, cobertura máxima 200. Para el filtrado de segunda ronda de SAMtools, el puntaje mínimo de calificación de indel fue 50, la calidad mínima de mutación puntual fue 30. Para llamadas de mutación GATK, se siguieron las pautas de mejores prácticas diseñadas por el autor presentadas en el manual del usuario de GATK (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit). La etapa de recalibración de puntuación variante se omitió y se reemplazó por la opción de filtrado duro. Para las llamadas de mutación SomaticSniper, se utilizaron las opciones predeterminadas y solo se consideraron más las mutaciones predichas con una "puntuación somática" de 30 o más.
- Descubrimiento de mutaciones, CTC: de acuerdo con el proceso iCAM, de tumores masivos 50 nt, extremo único, las

lecturas del Illumina HiSeq 2000 se alinearon usando bwa (versión 0.5.8c) [5]) usando opciones predeterminadas para el ensamblaje del genoma del ratón de referencia mm9. Como las lecturas de NGS de CTC se derivaron del ensayo de Seq-ARN, las lecturas también se alinearon con las secuencias de transcriptoma, incluidas las uniones exón-exón, utilizando Bowtie (más arriba). Usando todas las alineaciones, las secuencias de nucleótidos de las lecturas se compararon tanto con el genoma de referencia como con las mutaciones B16 derivadas del tumor masivo. Las mutaciones identificadas se evaluaron utilizando programas de instrucciones de Perl y manualmente utilizando el programa de software SAMtools y el IGV (Integrated Genome Viewer) para obtener imágenes de los resultados. El resultado del "descubrimiento de mutaciones" es la identificación de mutaciones somáticas en las células tumorales, desde la muestra hasta los datos de NGS y una lista de mutaciones. En las muestras B16, se identificaron 2448 mutaciones somáticas mediante resecuenciación del exoma.

Priorización de mutaciones

A continuación, se demostró la posibilidad de una fuente de información de priorización de la mutación para inclusión de la vacuna. Este método, llamado "canal de detección de mutación de un cáncer individual" (iCAM) identifica y prioriza las mutaciones somáticas a través de una serie de etapas que incorporan múltiples algoritmos de vanguardia y métodos bioinformáticos. El resultado de este proceso es una lista de mutaciones somáticas, priorizadas en función de la probable inmunogenicidad.

Identificación de mutaciones somáticas: las mutaciones se identifican usando tres algoritmos diferentes, tanto para las muestras B16 como para Black6 (descubrimiento de mutaciones, ver más arriba). La primera etapa de iCAM es combinar las listas de salida de cada algoritmo para generar una lista de mutaciones somáticas de alta confianza. GATK y SAMtools informan variantes en una muestra en relación con un genoma de referencia. Para seleccionar mutaciones de alta confianza con pocos falsos positivos para una muestra dada (es decir, tumor o normal), se seleccionan mutaciones que se identifican en todas las réplicas. Luego, se seleccionan variantes que están presentes en la muestra tumoral pero que no están presentes en la muestra normal. SomaticSniper informa automáticamente las posibles variaciones somáticas de los pares de datos tumorales y normales. Se filtró aún más los resultados a través de la intersección de los resultados obtenidos de las réplicas. Para eliminar tantas llamadas de falsos positivos como sea posible, se intersectó la lista de mutaciones derivadas del uso de los tres algoritmos y todas las réplicas. La etapa final para cada mutación somática es asignar un valor de confianza (valor p) para cada mutación en función de la profundidad de la cobertura, la calidad del SNP, la calidad del consenso y la calidad del mapeo.

Impacto de la mutación: el impacto de las mutaciones somáticas filtradas y de consenso se determina mediante un programa de instrucciones dentro del canal de mutaciones iCAM. Primero, las mutaciones que ocurren en regiones genómicas que no son únicas dentro del genoma, tal como las que ocurren para algunos parálogos de proteínas y pseudogenes, se excluyen del análisis a medida que se eliminan las lecturas de secuencias que se alinean en múltiples ubicaciones. En segundo lugar, se determina si la mutación ocurre en un transcrito. Tercero, se determina si la mutación ocurre en una región codificante de proteínas. Cuarto, la secuencia del transcrito se traduce con y sin la mutación para determinar si hay un cambio en la secuencia de aminoácidos.

Expresión de la mutación: el canal de iCAM selecciona mutaciones somáticas que se encuentran en genes y exones que se expresan en células tumorales. Los niveles de expresión se determinan a través de Seq-ARN de NGS de células tumorales (ver más arriba). El número de lecturas que se superponen a un gen y un exón indica niveles de expresión. Estos recuentos se normalizan a RPKM (lecturas por kilobase del modelo de exón por millón de lecturas mapeadas, [Mortazavi A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. Julio de 2008; 5 (7): 621-8. Epub mayo 30 de 2008]) y se seleccionan las 10 RPKM expresadas anteriormente.

Unión al MHC: para determinar la probabilidad de que un epítipo que contiene el péptido mutado se una a una molécula del MHC, el canal de iCAM ejecuta una versión modificada del software de predicción de MHC de la base de datos de inmuno epítipos (<http://www.iedb.org/>). La instalación local incluye modificaciones para optimizar el flujo de datos a través del algoritmo. Para los datos de B16 y Black6, la predicción se ejecutó utilizando todos los alelos del MHC de clase I de black6 disponibles y todos los epítipos para las longitudes de péptido respectivas. Se seleccionan las mutaciones que caen en un epítipo clasificado en el percentil 95 de la distribución de puntaje de predicción de los datos de entrenamiento de IEDB (<http://mhcbindingpredictions.immuneepitope.org/dataset.html>), considerando todos los alelos del MHC y todos los epítipos potenciales que se superponen a la mutación .

Criterios de selección de mutaciones: las mutaciones somáticas se seleccionan mediante los siguientes criterios: a) tienen un contenido de secuencia único, b) identificado por los tres programas, c) alta confianza de mutación, d) cambio en la proteína no sinónimos, e) alta expresión de transcrito, f) y predicción favorable de unión al MHC de clase I.

El resultado de este proceso es una lista de mutaciones somáticas, priorizadas en función de la probable inmunogenicidad. En las células de melanoma B16, hay 2448 mutaciones somáticas. 1247 de estas mutaciones se encuentran en transcritos de genes. De estos, 734 causan cambios en la proteína no sinónimos. De estos, 149 están en genes expresados en las células tumorales. De éstos, se predice que 102 de estas mutaciones no sinónimas expresadas se presentarán en las moléculas del MHC. Estas 102 mutaciones inmunogénicas probables se pasan luego a la confirmación de mutación (a continuación).

Confirmación de mutación

Las mutaciones somáticas de la resecuenciación del exoma de ADN se confirmaron por cualquiera de los dos métodos, la resecuenciación de la región mutada y el análisis de Seq-ARN.

- 5 Para la confirmación de las mutaciones por resecuenciación, una región genómica que contiene la mutación se amplificó por PCR estándar a partir de 50 ng de ADN tumoral y de ADN de control normal. El tamaño de los productos amplificados estaba en el intervalo de 150 a 400 nt. La especificidad de la reacción se controló cargando el producto de PCR en el dispositivo Qiaxel (Qiagen). Los productos de PCR se purificaron utilizando el kit de purificación de PCR minElute (Qiagen). Los productos de PCR específicos se secuenciaron utilizando el método de secuenciación estándar de Sanger (Eurofins), seguido de un análisis por electroferograma.
- 10 La confirmación de la mutación también se logró mediante el examen del ARN tumoral. El gen tumoral y los valores de expresión de exón se generaron a partir de Seq-ARN (NGS de ARN), que genera secuencias de nucleótidos que se mapearon en transcritos y se contaron. Se examinaron los datos de secuencia mismos para identificar mutaciones en la muestra tumoral [Berger MF. et al. Integrative analysis of the melanoma transcriptome. Abril de 2010; 20 (4): 413-27. Epub Feb 23 de 2010], que proporciona una confirmación independiente de las mutaciones somáticas identificadas derivadas del ADN.
- 15

Tabla 1: Lista de genes que contienen las 50 mutaciones validadas

Genes que contienen las 50 mutaciones somáticas identificadas y confirmadas, con anotaciones sobre el símbolo del gen, el nombre del gen y la localización y función predichas.			
ID	Símbolo	Entrada del nombre del gen	Localización
NM_021895	ACTN4	actinina, alfa 4	Citoplasma
NM_028840	ARMC1	repetición de armadillo que contiene 1	desconocida
NM_029291	ASCC2	subunidad 2 del complejo cointegrador 1 de señal de activación	desconocida
NM_024184	ASF1B	homólogo B de la función antisilenciadora 1 ASF1 (S. cerevisiae)	Núcleo
NM_138679	ASH1L	tipo ash1 (ausente, pequeño, u homeótico) (Drosophila)	Núcleo
NM_015804	ATP11A	ATPasa, clase VI, tipo 11A	Membrana plasmática
NM_009730	ATRN	atractina	Espacio extracelular
NM_028020	CPSF3L	tipo factor 3 específico de excisión y poliadenilación	Núcleo
NM_010017	DAG1	distroglicano 1 (glicoproteína 1 asociada a distrofina)	Membrana plasmática
NM_015735	DDB1	proteína 1 de unión al AND específica del daño, 127 kDa	Núcleo
NM_001080981	DDX23	polipéptido 23 de la cada DEAD (Asp-Glu-Ala-Asp)	Núcleo
NM_054046	DEF8	diferencialmente expresado en el homólogo de FDCP 8 (ratón)	desconocida
NM_019965	DNAJB12	Homólogo de DnaJ (Hsp40), subfamilia B, miembro 12	Citoplasma
NM_011262	DPF2	D4, familia 2 de dedos PHD de zinc y dobles	Núcleo
NM_007907	EEF2	factor 2 de alargamiento de la traducción eucariota	Citoplasma

ES 2 746 233 T3

Genes que contienen las 50 mutaciones somáticas identificadas y confirmadas, con anotaciones sobre el símbolo del gen, el nombre del gen y la localización y función predichas.			
ID	Símbolo	Entrada del nombre del gen	Localización
NM_001081286	FAT1	homólogo 1 supresor del tumor FAT (Drosophila)	Membrana plasmática
NM_173182	FNDC3B	dominio tipo III de fibronectina que contiene 3B	desconocida
NM_008057	FZD7	homólogo 7 rizado (Drosophila)	Membrana plasmática
NM_201617	GNAS	Locus del complejo GNAS	Membrana plasmática
NM_030035	GOLGB1	golgina B1	Citoplasma
NM_011365	ITSN2	intersectina 2	Citoplasma
NM_029841	KIAA2013	KIAA2013	desconocida
NM_197959	KIF18B	miembro 18B de la familia de las quinesinas	desconocida
NM_145479	KLHL22	tipo kelch 22 (Drosophila)	desconocida
NM_018810	MKRN1	proteína 1 del dedo anular makorina	desconocida
NM_001170785	MTHFD1L	tipo metilentetrahidrofolato deshidrogenasa 1 (dependiente de NADP+)	Citoplasma
NM_133947	NUMA1	proteína 1 del aparato mitótico nuclear	Núcleo
NM_178884	OBSL1	tipo obscurina 1	desconocida
NM_008765	ORC2	complejo de reconocimiento del origen, subunidad 2	Núcleo
NM_023209	PBK	Quinasa de unión a PDZ	Citoplasma
NM_033594	PCDHGA11	subfamilia A de la protocaderina gamma, 11	Membrana plasmática
NM_025951	PI4K2B	fosfatidilinositol 4-quinasa tipo 2 beta	Citoplasma
NM_011961	PLOD2	procolágeno-lisina, 2-oxoglutarato 5-dioxigenasa 2	Citoplasma
NM_023200	PPP1R7	proteín fosfatasa 1, subunidad reguladora 7 (inhibidora)	Núcleo
NM_008986	PTRF	polimerasa I y factor de liberación del transcrito	Núcleo
NM_011240	RANBP2	proteína 2 de unión RAN	Núcleo
NM_009438	RPL13A	proteína ribosomal L13a	Citoplasma
NM_009113	S100A13	proteína A13 de unión a calcio S100	Citoplasma
NM_001081203	SBNO1	dominio sema del homólogo 1 de muesca de fresa (Drosophila), dominio de inmunoglobulina (Ig), dominio	desconocido

Genes que contienen las 50 mutaciones somáticas identificadas y confirmadas, con anotaciones sobre el símbolo del gen, el nombre del gen y la localización y función predichas.			
ID	Símbolo	Entrada del nombre del gen	Localización
		básico corto, secretado	
NM_009153	SEMA3B	(semaforina) 3B	Espacio extracelular
NM_026912	SNX15	clasificación de nexina 15	Citoplasma
NM_024225	SNX5	clasificación de nexina 5	Citoplasma
NM_008188	THUMPD3	dominio THUMP que contiene 3	desconocido
NM_133352	TM9SF3	miembro 3 de la superfamilia transmembrana 9	Citoplasma
NM_177296	TNPO3	transportina 3	Citoplasma
NM_011640	TP53	proteína p53 tumoral	Núcleo
NM_023279	TUBB3	tubulina, beta 3	Citoplasma
NM_029896	WDR82	dominio 32 de repetición WD	desconocido
NM_025830	WWP2	dominio WW que contiene proteína ligasa 2 ubiquitina E3	Citoplasma
NM_001081056	XPOT	exportina, ARNt (receptor de exportación nuclear para ARNt)	Núcleo

Ejemplo 2: el algoritmo de selección IVAC permite la detección de mutaciones inmunogénicas

5 Para investigar si se podían inducir respuestas específicas de células T contra las mutaciones confirmadas de células de melanoma B16F10, se inmunizaron ratones C57BL/6 (n = 5/péptido) inalterados dos veces (d0, d7) por vía subcutánea con 100 µg de péptido (+ 50 µg de Poli I:C como adyuvante) que comprende la secuencia aa mutada o de tipo silvestre (véase la Tabla 2). Todos los péptidos tenían una longitud de 27 aa con el tipo aa mutado/silvestre en la posición central. En el día 12 se sacrificaron los ratones y se recogieron las células del bazo. Como método de lectura, se realizó ELISpot de IFNγ utilizando 5x10⁵ células de bazo/pozo como efectores y 5x10⁴ células dendríticas de médula ósea cargadas con péptidos (2 µg/mL) como células objetivo. Las células efectoras del bazo se probaron contra el péptido mutado, el péptido de tipo silvestre y un péptido de control (nucleoproteína del virus de la vesiculostomatitis, VSV-NP). Con 44 secuencias probadas, se observó que 6 de ellas inducían una inmunidad de células T dirigida solo contra la secuencia mutada pero no al péptido de tipo silvestre (Fig. 3). Los datos demuestran que las mutaciones identificadas y priorizadas se pueden utilizar para inducir inmunidad de células T específicas de tumor después de ser utilizadas como vacuna peptídica en ratones sin antígeno.

15 Tabla 2: Listado de secuencias mutadas que indujeron una reactividad de células T específica para el péptido mutado versus el péptido de tipo silvestre. El intercambio de aminoácidos está marcado con un subrayado.

Número	RefSeq ID	Tipo de secuencia	Secuencia del péptido	Reactividad de células T (ratones)
12	NM_00107750, M_010309, NM_201618, NM_201617	Mutada	TPPPEEAMPFEFN <u>G</u> PAQGDHSQPPLQV	5/5
		Tipo silvestre	TPPPEEAMPFEFN <u>E</u> PAQGDHSQPPLQV	0/5

Número	RefSeq ID	Tipo de secuencia	Secuencia del péptido	Reactividad de células T (ratones)
16	NM_008188	Mutada	RVTCNRAGEKHCFSSNEAARDFGGAIQ	3/5
		Tipo silvestre	RVTCNRAGEKHCFISNEAARDFGGAIQ	0/5
20	NM_023279	Mutada	FRRKAFLHWYTGEAMDEMEFTEAESNM	5/5
		Tipo silvestre	FRRKAFLHWYTGE G MDEMEFTEAESNM	1/5
30	NM_197959	Mutada	PSKPSFQEFVDWENVSPELNSTDQPFL	5/5
		Tipo silvestre	PSKPSFQEFVDWE K VSPELNSTDQPFL	1/5
34	NM_145479	Mutada	HLTQQLDITYILKNVAFSRTDKYRQLP	3/5
		Tipo silvestre	HLTQQLDITYILKNEVAFSRTDKYRQLP	0/5
36	NM_133352	Mutada	CGTAFFINFIAIYH H ASRAIPFGTMVA	5/5
		Tipo silvestre	CGTAFFINFIAIYY H ASRAIPFGTMVA	0/5

Ejemplo 3: las mutaciones identificadas pueden proporcionar inmunidad terapéutica antitumoral

Con el fin de validar si las mutaciones identificadas tienen el potencial de conferir inmunidad antitumoral después de la vacunación a ratones sin tratamiento previo, se investigó esta pregunta con el péptido para la mutación número 30 que se demostró que induce una reactividad de células T selectiva de la mutación. Las células B16F10 ($7,5 \times 10^4$) se inocularon subcutáneamente en d0. Los ratones fueron vacunados con el péptido 30 (véase la Tabla 1; 100 µg de péptido + 50 µg de Poli I:C s.c.) el día -4, día +2 y día +9. El grupo de control recibió solo Poli I:C (50 µg s.c.). El crecimiento tumoral se controló cada dos días. En el día +16 se observó que 1 de 5 ratones en el grupo de vacuna peptídica habían desarrollado un tumor, mientras que en el grupo de control, 4 de 5 ratones mostraron crecimiento tumoral.

Los datos demuestran que una secuencia peptídica que incorpora una mutación específica de B16F10 puede conferir inmunidad antitumoral que es capaz de destruir eficazmente las células tumorales (véase la Figura 4). Dado que B16F10 es una línea celular tumoral altamente agresiva, el hallazgo de que la metodología aplicada para identificar y priorizar mutaciones finalmente condujo a la selección de una mutación que por sí sola ya es potente como vacuna es una prueba importante de concepto para todo el proceso.

Ejemplo 4: Datos que respaldan la presentación del antígeno poliepitópico

Las mutaciones validadas de las regiones codificantes de proteínas de un paciente constituyen el conjunto del que se pueden seleccionar candidatos para el ensamblaje de la plantilla de vacuna de polineoepítipo para usar como precursor para la fabricación de GMP de la vacuna de ARN. Ya se han descrito casetes de vectores adecuados como cadena principal de la vacuna (Holtkamp, S. et al., Blood, 108: 4009-4017, 2006; Kreiter, S. et al., Cancer Immunol. Immunother., 56: 1577-1587, 2007; Kreiter, S. et al., J. Immunol., 180: 309-318, 2008). Los casetes de vector preferidos se modifican en sus regiones de codificación y no traducidas (UTR) y aseguran la traducción maximizada de la proteína codificada durante períodos prolongados (Holtkamp, S. et al., Blood, 108: 4009-4017, 2006; Kuhn, AN et al., Gene Ther., 17: 961-971, 2010). Además, la cadena principal del vector contiene módulos de enrutamiento de antígeno para la expansión simultánea de células T citotóxicas y auxiliares (Kreiter, S. et al., Cancer Immunol. Immunother., 56: 1577-1587, 2007; Kreiter, S. et al., J. Immunol., 180: 309-318, 2008; Kreiter, S. et al., Cancer Research, 70 (22), 9031-9040, 2010 (Figura 5). De manera muy importante, se ha demostrado que dicha vacuna de ARN se puede usar para presentar múltiples epitopos al MHC de clase I y de clase II simultáneamente.

Las secuencias de vacuna de ARN de polineoepítipo IVAC se construyen a partir de tramos de hasta 30 aminoácidos que incluyen la mutación en el centro. Estas secuencias están conectadas de la cabeza a la cola a través de enlazadores cortos para formar una vacuna de polineoepítipo que codifica hasta 30 o más mutaciones seleccionadas y sus regiones flanqueantes. Estos insertos personalizados individualmente específicos para el paciente son de codón optimizado con codones y clonados en la cadena principal de ARN descrito anteriormente. El control de calidad de

tales constructos incluye la transcripción y expresión *in vitro* en células para la validación de la transcripción y traducción funcional. El análisis de la traducción se realizará con anticuerpos contra el dominio de direccionamiento del terminal C.

Ejemplo 5: Prueba científica de concepto para el constructo del polineoepítipo de ARN

5 El concepto de polineoepítipo de ARN se basa en un largo ARNm transcrito *in vitro* que consiste en secuencias ordenadas secuencialmente que codifican los péptidos mutados conectados por secuencias enlazadoras (véase Figura 6). Las secuencias de codificación se eligen entre las mutaciones no sinónimas y siempre se forman a partir del codón para el aminoácido mutado flanqueado por regiones de 30 a 75 pares de bases del contexto de secuencia original. La secuencia enlazadora codifica aminoácidos que preferentemente no son procesados por la maquinaria de procesamiento de antígeno celular. Los constructos de transcripción *in vitro* se basan en el vector pST1-A120 que contiene un promotor T7, una secuencia 3' UTR de beta-globina en tándem y una cola de poli(A) de 120 pb, que se ha demostrado que aumentan la estabilidad y la eficiencia traduccional del ARN mejorando de esta manera la capacidad estimuladora de las células T del antígeno codificado (Holtkamp S. et al., Blood 2006; PMID: 16940422). Además, se insertaron un fragmento del péptido señal de MHC de clase I y los dominios transmembrana y citosólicos, incluido el codón de parada (señal de tráfico del MHC de clase I o MITD) que flanquea una secuencia polienlazadora para clonar los epítipos (Kreiter S. et al., J. Immunol., 180: 309-318, 2008). Se ha demostrado que estos últimos aumentan la presentación del antígeno, mejorando así la expansión de las células T CD8+ y CD4+ específicas de antígeno y mejorando las funciones efectoras.

20 Para una primera prueba de concepto, se usaron vectores biepitópicos, es decir, que codifican un polipéptido que contiene dos epítipos mutados. Secuencias optimizadas de codones que codifican para (i) un epítipo mutado de 20 a 50 aminoácidos, (ii) un enlazador rico en glicina/serina, (iii) un segundo epítipo mutado de 20 a 50 aminoácidos y (iv) un enlazador adicional rico en glicina/serina, flanqueado por sitios de reconocimiento adecuados para las endonucleasas de restricción que se clonarán en el constructo a base de pST1 como se describió anteriormente, fue diseñado y sintetizado por un proveedor comercial (Geneart, Regensburg, Alemania). Después de la verificación de la secuencia, estas se clonaron en la cadena principal del vector basado en pST1 para obtener constructos como se representa en la Figura 6.

30 Los plásmidos basados en pST1-A120 como se describieron anteriormente se linealizaron con una endonucleasa de restricción de clase II. Los ADN plasmídicos linealizados se purificaron mediante extracción con cloroformo y fenol y precipitación con etanol. Los vectores de ADN linealizados se cuantificaron espectrofotométricamente y se sometieron a transcripción *in vitro* esencialmente como lo describen Pokrovskaya y Gurevich (1994, Anal. Biochem. 220: 420-423). Se ha añadido un análogo de caperuza a la reacción de transcripción para obtener ARN con las estructuras de caperuza de 5' correspondientemente modificadas. En las reacciones, GTP estaba presente a 1,5 mM, mientras que el análogo de caperuza estaba presente a 6,0 mM. Todos los demás NTP estaban presentes a 7,5 mM. Al final de la reacción de transcripción, el ADN del vector linealizado se digirió con 0,1 U/μL de TURBO DNasa (Ambion, Austin/TX, EE. UU.) Durante 15 minutos a 37 °C. Los ARN se purificaron a partir de estas reacciones utilizando el kit MEGAclean (Ambion, Austin/TX, EE. UU.) de acuerdo con el protocolo del fabricante. La concentración y calidad de ARN se evaluaron mediante espectrofotometría y análisis en un bioanalizador 2100 (Agilent, Santa Clara, CA, EE. UU.).

40 Para demostrar que una secuencia que incorpora un aminoácido mutado y que está flanqueada por la secuencia enlazadora 5' y 3' puede ser procesada, presentada y reconocida por células T específicas de antígeno, se usaron células T de ratones vacunados con péptidos como células efectoras. En un ELISpot de IFNγ se probó si las células T inducidas por la vacunación con péptidos como se describió anteriormente son capaces de reconocer las células objetivo (células dendríticas de médula ósea, BMDC) pulsadas con péptido (2 μg/mL durante 2 horas a 37 °C y 5% CO₂) o transfectadas con ARN (20 μg producidos como se describió anteriormente) por electroporación. Como se ejemplifica en la Figura 7 para la mutación 12 y 30 (véase la tabla 2), se pudo observar que el constructo de ARN puede dar lugar al epítipo reconocido por las células T específicas de mutación.

45 Con los datos proporcionados, se podría demostrar que un polineoepítipo codificado por ARN que incluye un enlazador rico en glicina/serina puede traducirse y procesarse en células presentadoras de antígeno que conducen a la presentación del epítipo correcto que es reconocido por las células T específicas de antígeno .

Ejemplo 6: Diseño de la vacuna de polineoepítipo: La relevancia del enlazador

50 El constructo de ARN de polineoepítipo contiene un constructo de cadena principal en la que se colocan múltiples péptidos que codifican mutaciones somáticas conectadas con una secuencia peptídica enlazadora. Además de la optimización del codón y el aumento de la estabilidad del ARN y la eficiencia de la traducción debido a la cadena principal, una realización de la vacuna de polineoepítipo de ARN contiene enlazadores diseñados para aumentar la presentación de péptidos antigénicos al MHC de clase I y II y disminuir la presentación de epítipos perjudiciales.

55 Enlazador: la secuencia del enlazador se diseñó para conectar múltiples péptidos que contienen mutaciones. El enlazador debería permitir la creación y presentación del epítipo de mutación mientras que dificulta la creación de epítipos perjudiciales, tal como los creados en la sutura de unión entre péptidos adyacentes o entre la secuencia del enlazador y los péptidos endógenos. Estos epítipos de "unión" pueden no solo competir con los epítipos previstos

que se presentarán en la superficie celular, disminuyendo la eficacia de la vacuna, sino que podrían generar una reacción autoinmune no deseada. Por lo tanto, se diseñó la secuencia del enlazador para a) evitar crear péptidos de "unión" que se unan a las moléculas del MHC, b) evitar el procesamiento proteasomal para crear péptidos de "unión", c) ser traducidos y procesados de manera eficiente por el proteasoma.

5 Para evitar la creación de péptidos de "unión" que unen moléculas del MHC, se compararon diferentes secuencias del enlazador. La glicina, por ejemplo, inhibe la unión fuerte en las posiciones de ranura de unión del MHC [Abastado JP. Et al., J. Immunol. 1 de octubre de 1993; 151 (7): 3569-75]. Se examinaron múltiples secuencias de enlazadores y múltiples longitudes de enlazadores y se calculó el número de péptidos de "unión" que se unen a las moléculas del MHC. Se utilizaron las herramientas de software de la Base de datos de epítomos inmunes (IEDB, <http://www.immuneepitope.org/>) para calcular la probabilidad de que una secuencia de péptidos dada contenga un ligando que se unirá a las moléculas del MHC de clase I.

15 En el modelo de B16, se identificaron 102 mutaciones somáticas no sinónimas expresadas que se prevé que se presenten en las moléculas del MHC de clase I. Usando las 50 mutaciones confirmadas, se diseñaron computacionalmente diferentes constructos de vacunas, incluido el uso de no enlazadores o el uso de diferentes secuencias de enlazadores, y se calculó el número de péptidos de "unión" perjudiciales usando el algoritmo IEDB (Figura 8).

20 La Tabla 5 muestra los resultados de varios enlazadores diferentes, diferentes longitudes de enlazadores, y el uso de ningún enlazador y cinco enlazadores. El número de péptidos de unión que se une a al MHC varía de 2 a 91 para las predicciones del epítomo de 9 aa y 10 aa (superior y medio). El tamaño del enlazador influye en el número de péptidos de unión (abajo). Para esta secuencia, se predice la menor cantidad de epítomos de 9 aa para la secuencia del enlazador de 7 aa GGS GGGG.

El Enlazador 1 y el Enlazador 2 usados en los constructos de la vacuna de polineoepítomo de ARN probados experimentalmente (véase más abajo) también tenían un número favorablemente bajo de neoepítomos de unión pronosticados. Esto es válido para las predicciones de 9 mer y 10 mer.

25 Esto demuestra que la secuencia del enlazador es críticamente importante para la creación de epítomos malos de unión al MHC. Además, la longitud de la secuencia del enlazador afecta el número de epítomos malos de unión al MHC. Se encontró que las secuencias que son ricas en G impiden la creación de ligandos de unión al MHC.

30 Tabla 3. Impacto del enlazador (epítomos de 10 aa). El número previsto de epítomos malos definidos como epítomos de unión al MHC de clase I que contienen secuencias de unión, para cada enlazador peptídico. En este documento, se consideran epítomos de 10 aminoácidos. Los enlazadores ricos en glicina tienen la menor cantidad de epítomos de unión

Enlazador	# de epítomos malos (10 aa)
ninguno	14
TSLNALLNAH	54
SIINFEKL	65
SSSSSSSSSS	85
GGGGGGGGGG	6
GGS GGGGSGG (Enlazador 1)	8
GGS GGGGSGG (Enlazador 2)	9

35 Tabla 4. Impacto de la parte del enlazador (epítomos de 9 aa). El número previsto de epítomos malos, definidos como epítomos de unión al MHC de clase I que contienen secuencias de unión, para cada enlazador peptídico. En este documento, se consideran epítomos de 9 aminoácidos. Los enlazadores ricos en glicina tienen la menor cantidad de epítomos de unión.

Enlazador	# de epítomos malos (9 aa)
ninguno	17
TSLNALLNAH	83
SIINFEKL	64

SSSSSSSSSS	33
GGGGGGGGGG	2
GGSGGGGSGG (Enlazador 1)	4
GGSGGGGSGG (Enlazador 2)	3

5 Tabla 5: Impacto de la parte del enlazador. El número previsto de epítomos malos, definidos como epítomos de unión al MHC de clase I que contienen secuencias de unión, para cada enlazador peptídico. En este documento, se consideran epítomos de 9 aminoácidos. Parte superior: el número de epítomos de unión de 9 aa para ningún enlazador y 5 enlazadores diversos. Parte media: el número de epítomos de unión de 10 aa para ningún enlazador y 5 enlazadores diversos. Parte inferior: el número de epítomos de unión de 9 aa para enlazadores similares de diferentes longitudes. Los enlazadores ricos en glicina tienen la menor cantidad de epítomos de unión.

Enlazador	# de epítomos de unión (9 aa)
ninguno	17
TSLNALLNA	91
SIINFEKL	64
SSSSSSSSSS	33
GGGGGGGGGG	2
GGSGGGGSGG	4

Enlazador	# de epítomos de unión (10 aa)
ninguno	14
TSLNALLNA	63
SIINFEKL	65
SSSSSSSSSS	85
GGGGGGGGGG	6
GGSGGGGSGG	9

Enlazador	# de epítomos de unión (9 aa)
GGSGG	5
GGSGGG	4
GGSGGGG	2
GGSGGGGS	7
GGSGGGGSG	4
GGSGGGGSGG	4

10 Para evitar el procesamiento proteasomal que puede crear péptidos de "unión", se exploró el uso de diferentes aminoácidos en el enlazador. Las secuencias ricas en glicina deterioran el procesamiento proteasomal [Hoyt MA et al., (2006) EMBO J 25 (8): 1720-9; Zhang M. y Coffino P. (2004) J Biol Chem 279 (10): 8635-41]. Por lo tanto, las secuencias del enlazador ricas en glicina actúan para minimizar el número de péptidos que contienen enlazadores que el proteasoma puede procesar.

15 El enlazador debería permitir que los péptidos que contienen la mutación sean traducidos y procesados eficientemente por el proteasoma. Los aminoácidos glicina y serina son flexibles [Schlessinger A y Rost B., Proteins. Octubre 1 de

2005; 61 (1): 115-26]; incluirlos en un enlazador da como resultado una proteína más flexible. Se incorporaron glicina y serina en el enlazador para aumentar la flexibilidad de la proteína, lo que debería permitir una traducción y procesamiento más eficientes por el proteasoma, lo que a su vez permite un mejor acceso a los péptidos antigénicos codificados.

- 5 Por lo tanto, el enlazador debería ser rico en glicina para impedir la creación de epítomos malos que se unen al MHC; debería obstaculizar la capacidad del proteasoma para procesar péptidos enlazadores, lo que puede lograrse mediante la inclusión de glicina; y debe ser flexible para aumentar el acceso a péptidos que contienen mutaciones, lo que puede lograrse mediante la combinación de aminoácidos de glicina y serina. Por lo tanto, en una realización del constructo de la vacuna, las secuencias GSGGGGSGG y GSGGGGSGGS se incluyen preferiblemente como secuencias enlazadoras.

Ejemplo 7: Vacuna de polineoepitopo de ARN

Los constructos de la vacuna de polineoepitopo de ARN se basan en el vector pST1-A120 que contiene un promotor T7, una secuencia de 3'UTR beta-globina en tándem y una cola de poli(A) de 120 pb, que se ha demostrado que aumentan la estabilidad y la eficiencia traduccional del ARN mejorando así la capacidad estimuladora de las células T del antígeno codificado (Holtkamp S. et al., Blood 2006; PMID: 16940422). Además, se insertaron un fragmento de péptido señal del MHC de clase I y los dominios transmembrana y citosólicos que incluyen el codón de parada (señal de tráfico del MHC de clase I o MITD) que flanquean una secuencia de polienlazadores para clonar los epítomos (Kreiter S. et al., J. Immunol., 180: 309-318, 2008). Se ha demostrado que estos últimos aumentan la presentación del antígeno, mejorando así la expansión de las células T CD8+ y CD4+ específicas de antígeno y mejorando las funciones efectoras.

Para proporcionar constructos de polineoepitopos de ARN para las 50 mutaciones identificadas y validadas de B16F10, se generaron 3 constructos de ARN. El constructo consiste en secuencias de codones optimizados que codifican (i) un epítopo mutado de 25 aminoácidos, (ii) un enlazador rico en glicina/serina, (iii) repeticiones de la secuencia del epítopo mutado seguido de un enlazador rico en glicina/serina. La cadena de secuencias y enlazadores que contienen epítomos mutados está flanqueada por sitios de reconocimiento adecuados para las endonucleasas de restricción que se clonarán en el constructo basada en pST1 como se describió anteriormente. Los constructos de la vacuna fueron diseñados y sintetizados por GENEART. Después de la verificación de la secuencia, estos se clonaron en la cadena principal del vector basado en pST1 para obtener los constructos de polineoepitopo de la vacuna de ARN.

30 Descripción del enfoque clínico

La aplicación clínica cubrirá las siguientes etapas:

- Los pacientes elegibles deben aceptar el análisis de ADN mediante secuenciación de próxima generación.
- Se obtendrán muestras de tumor obtenidas de procedimientos de diagnóstico de rutina (tejido fijado con formalina embebido en parafina) y células de sangre periférica y se utilizarán para el análisis de mutaciones como se describió.
- Las mutaciones descubiertas serán confirmadas
- Con base en la priorización se diseñará la vacuna. Para las vacunas de ARN, se generará una plantilla de plásmido maestro mediante síntesis y clonación de genes.
- Los plásmidos se utilizarán para la producción de ARN de grado clínico, el control de calidad y la liberación de la vacuna de ARN.
- El producto farmacológico de la vacuna se enviará al centro de prueba correspondiente para su aplicación clínica.
- La vacuna de ARN puede usarse como una vacuna desnuda en regulador de formulación o encapsulada en nanopartículas o liposomas para inyección directa en, por ejemplo, ganglios linfáticos, s.c., i.v., i.m.. Alternativamente, la vacuna de ARN puede usarse para la transfección *in vitro*, por ejemplo, de células dendríticas para transferencia adoptiva.

- 45 Todo el proceso clínico lleva menos de 6 semanas. La "fase de retraso" entre el consentimiento informado del paciente y la disponibilidad del medicamento se abordará cuidadosamente mediante el protocolo del ensayo clínico, lo que incluye permitir que se continúe con el régimen de tratamiento estándar hasta que el producto del medicamento en investigación esté disponible.

Ejemplo 8: Identificación de mutaciones tumorales y su explotación para la vacunación de tumores

- 50 Se aplicó la resecuenciación del exoma NGS para el descubrimiento de mutaciones en la línea celular de melanoma murino B16F10 y se identificaron 962 mutaciones puntuales somáticas no sinónimas, 563 en genes expresados. Las posibles mutaciones conductoras se producen en genes supresores de tumores clásicos (*Pten*, *Trp53*, *Tp63*, *Pml*) y

genes implicados en vías de señalización protooncogénicas que controlan la proliferación celular (por ejemplo, *Mdm1*, *Pdgfra*), la adhesión y migración celular (por ejemplo, *Fdz7*, *Fat1*) o apoptosis (*Casp9*). Además, B16F10 alberga mutaciones en *Aim1* y *Trrap* que se describieron previamente para ser alteradas con frecuencia en el melanoma humano.

- 5 La inmunogenicidad y especificidad de 50 mutaciones validadas se ensayaron usando ratones C57BL/6 inmunizados con péptidos largos que codifican los epítomos mutados. Un tercio (16/50) de ellos demostraron ser inmunogénicos. De estos, el 60% provocó respuestas inmunes dirigidas preferentemente contra la secuencia mutada en comparación con la secuencia de tipo silvestre.

- 10 Se probó la hipótesis en modelos de trasplante de tumores. La inmunización con péptidos confirió el control tumoral *in vivo* en entornos protectores y terapéuticos, calificando epítomos mutados que contenían sustituciones de aminoácidos individuales como vacunas eficaces.

Animales

Los ratones C57BL/6 (Jackson Laboratories) se mantuvieron de acuerdo con las políticas federales y estatales sobre investigación animal en la Universidad de Mainz.

- 15 Células

La línea celular de melanoma B16F10 se adquirió en 2010 de la Colección Americana de Cultivos Tipo (Producto: ATCC CRL-6475, Número de lote: 58078645). Se utilizaron pasajes tempranos (3°, 4°) de células para experimentos tumorales. Las células fueron analizadas rutinariamente para detectar *Mycoplasma*. La reautenticación de las células no se ha realizado desde la recepción.

- 20 Secuenciación de próxima generación

Extracción de ácido nucleico y preparación de muestras: ADN y ARN de células B16F10 a granel y ADN de tejido de cola de C57BL/6 se extrajeron por triplicado usando el kit de Sangre y Tejidos DNeasy de Qiagen (para ADN) y el micro kit RNeasy de Qiagen (para ARN).

- 25 Secuenciación del exoma del ADN: la captura del exoma para la resecuenciación del ADN se realizó por triplicado usando el ensayo de captura basado en la solución de ratón Sure-Select de Agilent (Gnirke A et al., Nat Biotechnol 2009; 27: 182-9), diseñado para capturar todas las regiones de codificación de proteínas de ratón. Se fragmentaron 3 µg de ADN genómico purificado (ADNg) a 150-200 pb usando un dispositivo de ultrasonido Covaris S2. Los fragmentos fueron reparados en los extremos y fosforilados en 5' y adenilados en 3' de acuerdo con las instrucciones del fabricante. Los adaptadores finales emparejados de Illumina se ligaron a los fragmentos de ADNg usando una relación molar 10:1 de adaptador a ADNg. Se agregaron secuencias específicas de células de flujo y captura previa enriquecidas utilizando los cebadores de PCR Illumina PE 1.0 y 2.0 durante 4 ciclos de PCR. Se hibridaron 500 ng de fragmentos de ADNg enriquecidos con PCR ligados con adaptador a cebos de biblioteca de ARN de exoma completo de ratón biotinilado SureSelect de Agilent durante 24 horas a 65 °C. Los complejos de cebo hibridado de ADNg/ARN se eliminaron usando perlas magnéticas recubiertas con estreptavidina, se lavaron y los cebos de ARN se separaron durante la elución en regulador de elución SureSelect. Estos fragmentos de ADNg eluidos se amplificaron en 10 ciclos por PCR después de la captura. Las bibliotecas de ADNg enriquecidas en exoma se agruparon en el cBot usando el kit de agrupación Truseq SR v2.5 usando 7 pM y se secuenciaron 50 pb en el Illumina HiSeq2000 usando el kit Truseq SBS-HS de 50 pb.

- 40 Perfiles del "transcriptoma" de expresión génica de ARN (Seq-ARN): se prepararon bibliotecas de ADNc de Seq-ARNm con código de barras por triplicado, a partir de 5 µg de ARN total (protocolo Illumina Seq-ARNm modificado). El ARNm se aisló usando perlas magnéticas Seramag Oligo (dT) (Thermo Scientific) y se fragmentó usando cationes divalentes y calor. Los fragmentos resultantes (160-220 pb) se convirtieron en ADNc usando cebadores aleatorios y SuperScriptII (Invitrogen) seguido de síntesis de segunda cadena usando ADN polimerasa I y RNasaH. El ADNc se reparó en el extremo, se fosforiló en 5' y se adeniló en 3' de acuerdo con las instrucciones del fabricante. Los adaptadores específicos multiplex del extremo 3' emparejados con salientes T individuales Illumina se ligaron con ADN ligasa T4 usando una relación molar 10:1 de adaptador a inserto de ADNc. Las bibliotecas de ADNc se purificaron y el tamaño se seleccionó en 200-220 pb (gel E-Gel 2% SizeSelect, Invitrogen). El enriquecimiento, la adición de Illumina de seis bases de índice y secuencias específicas de células de flujo se realizó por PCR usando ADN polimerasa Phusion (Finnzymes). Todas las limpiezas hasta esta etapa se realizaron con un volumen de 1,8x de perlas magnéticas AgencourtAMPure XP. Todos los controles de calidad se realizaron usando el ensayo Qubit HS de Invitrogen y el tamaño del fragmento se determinó usando el ensayo de ADN HS con 2100 Bioanalyzer Agilent. Las bibliotecas de Seq-ARN con código de barras se agruparon y secuenciaron como se describió anteriormente.

- 55 Análisis de datos de NGS, expresión génica: los resultados de las lecturas de secuencia 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 (Waterston RH et al., Nature 2002; 420: 520-62) con Bowtie (versión 0.12.5) (Langmead B et al., Genome Biol 2009; 10: R25). Para las alineaciones del genoma, se permitieron dos desajustes y solo se registró la mejor alineación ("mejor v2"); para las

alineaciones de transcriptomas se utilizaron los parámetros predeterminados. Las lecturas que no se pueden alinear con la secuencia genómica se alinearon con una base de datos de todas las posibles secuencias de unión exón-exón de transcritos RefSeq (Pruitt KD et al., *Nucleic Acids Res* 2007; 35: D61-D65). Los valores de expresión se determinaron intersectando las coordenadas de lectura con las de los transcritos de RefSeq, contando las lecturas superpuestas de exón y unión, y normalizando a unidades de expresión RPKM (Reads which map per Kilobase of exon model per million mapped reads) (Mortazavi A et al., *Nat Methods* 2008; 5: 621-8).

Análisis de datos de NGS, descubrimiento de mutaciones somáticas: las mutaciones somáticas se identificaron como se describe en el Ejemplo 9. 50 nucleótidos (nt), las lecturas de un solo extremo se alinearon con el genoma del ratón de referencia mm9 usando bwa (opciones predeterminadas, versión 0.5.8c) (Li H y Durbin R, *Bioinformatics* 2009; 25: 1754-60). Se eliminaron el mapeo de lecturas ambiguas para múltiples ubicaciones del genoma. Las mutaciones se identificaron utilizando tres programas de software: SAMtools (versión 0.1.8) (Li H, *Bioinformatics* 2011; 27: 1157-8), GATK (versión 1.0.4418) (McKenna A et al, *Genome Res* 2010; 20: 1297-303) y SomaticSniper (<http://genome.wustl.edu/software/somaticsniper>) (Ding L et al., *Hum Mol Genet* 2010; 19: R188-R196). A las posibles variaciones identificadas en todos los triplicados de B16F10 se les asignó un valor de confianza de "tasa de descubrimiento falso" (FDR) (véase el Ejemplo 9).

Selección de mutación, validación y función

Selección: las mutaciones tenían que cumplir los siguientes criterios para ser seleccionados: (i) presente en todos los B16F10 y ausente en todos los triplicados de C57BL/6, (ii) $FDR \leq 0,05$, (iii) homogéneo en C57BL/6, (iv) ocurrir en un transcrito de RefSeq y (v) causar que los cambios no sinónimos se califiquen como una mutación auténtica. La selección para la validación y las pruebas de inmunogenicidad requirieron que las mutaciones sean genes expresados (mediana de las réplicas a través de RPKM > 10).

Validación: las mutaciones derivadas de ADN se clasificaron como validadas si se confirmaban mediante la secuenciación de Sanger o las lecturas de Seq-ARN de B16F10. Todas las variantes seleccionadas se amplificaron a partir de 50 ng de ADN de células B16F10 y tejido de cola de C57BL/6 usando cebadores flanqueantes, productos visualizados (sistema QIAxcel, Qiagen) y purificados (kit de purificación por PCR QIAquick, Qiagen). El amplicón del tamaño esperado se escindió del gel, se purificó (kit de extracción en gel QIAquick, Qiagen) y se sometió a secuenciación de Sanger (Eurofins MWG Operon, Ebersberg, Alemania) con el cebador directo utilizado para la amplificación por PCR.

Impacto funcional: los programas SIFT (Kumar P et al., *Nat Protoc* 2009; 4: 1073-81) y POLYPHEN-2 (Adzhubei IA et al., *Nat Methods* 2010; 7: 248-9), que predicen la importancia funcional de un aminoácido en la función de la proteína basada en la ubicación de dominios de proteínas y la conservación de la secuencia de especies cruzadas, se emplearon para evaluar el impacto de mutaciones seleccionadas. Se utilizaron herramientas de Ingenuity IPA para inferir la función del gen.

Péptidos sintéticos y adyuvantes

Todos los péptidos, incluyendo ovoalbúmina clase I (OVA₂₅₈₋₂₆₅), clase II (OVA clase II₃₃₀₋₃₃₈), nucleoproteína de influenza (Inf-NP₃₆₆₋₃₇₄), nucleoproteína del virus de vesiculostomatitis (VSV-NP₅₂₋₅₉) y la proteína 2 relacionada con tirosinasa (Trp₂₁₈₀₋₁₈₈) se adquirió a través de Jerini Peptide Technologies (Berlín, Alemania). Los péptidos sintéticos tenían una longitud de 27 aminoácidos con el aminoácido mutado (MUT) o de tipo silvestre (WT) en la posición 14. Poliinosínico: se utilizó ácido policitidílico (poli (I: C), InvivoGen) como adyuvante inyectado por vía subcutánea. El pentámero de MHC específico para el péptido Inf-NP₃₆₆₋₃₇₄ se adquirió a través de Prolimmune Ltd..

Inmunización de ratones

Se inyectaron por vía subcutánea ratones C57BL/6 hembras de edad similar con 100 µg de péptido y 50 µg de poli (I: C) formulado en PBS (200 µL de volumen total) en el flanco lateral (5 ratones por grupo). Cada grupo se inmunizó el día 0 y el día 7 con dos péptidos codificadores de mutación diferentes, un péptido por flanco. Doce días después de la inyección inicial se sacrificaron los ratones y se aislaron los esplenocitos para pruebas inmunológicas.

Alternativamente, ratones hembra C57BL/6 de edad similar se inyectaron por vía intravenosa con 20 µg de ARN 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). Todos los grupos 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 las pruebas inmunológicas. Las secuencias de ADN que representan una (Monoepítipo), dos (Biepítipo) o 16 mutaciones (Poliepítipo) se construyeron utilizando 50 aminoácidos (aa) con la mutación en la posición 25 (Biepítipo) o 27 aa con la mutación en la posición 14 (Mono y Poliepítipo), se separaron mediante un enlazador de glicina/serina de 9 aa y se clonaron en la cadena principal pST1-2BgUTR-A120 (Holtkamp et al., *Blood* 2006; 108: 4009-17). La transcripción *in vitro* de esta plantilla y la purificación se describieron previamente (Kreiter et al., *Cancer Immunol Immunother* 2007; 56: 1577-87).

Ensayo por inmunoadsorción ligado a enzimas de puntos

El ensayo por inmunoadsorción ligado a enzimas de puntos (ELISPOT) (Kreiter S et al., *Cancer Res* 2010; 70: 9031-

40) y la generación de células dendríticas singénicas derivadas de médula ósea (BMDC) como estimuladores se describieron previamente (Lutz MB et al., J Immunol Methods 1999; 223: 77-92). Las BMDC se pulsaron con péptidos (2 µg/mL) o se transfectaron con ARN transcrito *in vitro* (IVT) que codifica la mutación indicada o para el ARN de control (ARN-eGFP). Las secuencias que representan dos mutaciones, cada una de las cuales comprende 50 aminoácidos con la mutación en la posición 25 y separadas por un enlazador de glicina/serina de 9 aa, se clonaron en la cadena principal pST1-2BgUTR-A120 (Holtkamp S et al., Blood 2006; 108: 4009-17). La transcripción *in vitro* de esta plantilla y la purificación se describieron previamente (Kreiter S et al., Cancer Immunol Immunother 2007; 56: 1577-87). Para el ensayo, se incubaron conjuntamente 5×10^4 BMDC modificadas con péptido o ARN 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 citoquinas con un anticuerpo anti-IFN- γ (clon R4-6A2; Mabtech). Los números de puntos se contaron y analizaron con el analizador ImmunoSpot® S5 Versa ELISPOT, 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 cuando la prueba produjo un valor de $p < 0,05$ y los números medios de puntos fueron > 30 puntos/ 5×10^5 células efectoras. Las reactividades se clasificaron por números medios de puntos (-: < 30 ; +: > 30 ; ++: > 50 ; +++: > 200 puntos/pozo).

Ensayo de citoquinas intracelulares

Las alícuotas de los esplenocitos preparados para el ensayo ELISPOT se sometieron a análisis de la producción de citoquinas mediante citometría de flujo intracelular. Con este fin, se sembraron en placa 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 pozos. Las células de cada animal se reestimularon durante 5 h a 37 °C con 2×10^5 BMDC pulsadas con péptido. 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 minutos 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 Cytotfix/Cytoperm (BD Bioscience) durante 20 minutos a 4 °C para la permeabilización de la membrana externa. Después de la permeabilización, las células se lavaron con regulador de Permeabilización/Lavado (BD Bioscience), se resuspendieron en 50 µL/muestra en regulador de Permeabilización/Lavado y se tiñeron intracelularmente con los siguientes anticuerpos anti-ratón durante 30 minutos a 4 °C: PE anti-IFN- γ , PE-Cy7 anti-TNF- α , APC anti-IL2 (BD Pharmingen). Después de lavar con regulador de Permeabilización/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).

Modelo de tumor de melanoma B16

Para los experimentos de vacunación tumoral, se inocularon $7,5 \times 10^4$ células de melanoma B16F10 s.c. en los flancos de ratones C57BL/6. En el contexto profiláctico, la inmunización con péptido específico de mutación se realizó 4 días antes y los días 2 y 9 después de la inoculación del tumor. Para el experimento terapéutico, la vacuna peptídica se administró los días 3 y 10 después de la inyección del tumor. Los tamaños del tumor se midieron cada tres días y los ratones se sacrificaron cuando el diámetro del tumor alcanzó 15 mm.

Alternativamente, para los experimentos de vacunación tumoral se inocularon 1×10^5 células de melanoma B16F10 s.c. en los flancos de ratones C57BL/6 hembra de la misma edad. La vacunación con péptidos se realizó los días 3, 10 y 17 después de la inoculación del tumor con 100 µg de péptido y 50 µg de poli (I: C) formulado en PBS (200 µL de volumen total) inyectado por vía subcutánea en el flanco lateral. Las inmunizaciones de ARN se realizaron usando 20 µg de ARN codificador de mutación transcrito *in vitro* formulado con 20 µL de Lipofectamine^{MR} RNAiMAX (Invitrogen) en PBS en un volumen de inyección total de 200 µL. Como control, se inyectó un grupo de animales con RNAiMAX (Invitrogen) en PBS. Los animales fueron inmunizados los días 3, 6, 10, 17 y 21 después de la inoculación del tumor. Los tamaños del tumor se midieron cada tres días usando un calibrador y los ratones se sacrificaron cuando el diámetro del tumor alcanzó 15 mm.

Identificación de mutaciones no sinónimas en melanoma de ratón B16F10

Nuestro objetivo fue identificar mutaciones puntuales somáticas potencialmente inmunogénicas en el melanoma de ratón B16F10 por NGS y evaluar su inmunogenicidad *in vivo* mediante la vacunación con péptidos de ratones que miden las respuestas de células T provocadas por el ensayo ELISPOT (Figura 9A). Se secuenciaron los exomas del genoma de fondo de tipo silvestre de C57BL/6 y de las células B16F10, cada una con extracciones y capturas por triplicado. Para cada muestra, se generaron más de 100 millones de lecturas de 50 nt de extremo único. De estos 80%, se alinean exclusivamente con el genoma mm9 del ratón y el 49% se alinean en el objetivo, lo que demuestra un enriquecimiento objetivo exitoso y resulta en una cobertura de más del 20% para el 70% de los nucleótidos objetivo en cada una de las muestras por triplicado. La Seq-ARN de las células B16F10, también perfilado por triplicado, generó una media de 30 millones de lecturas de 50 nt de extremo único, de las cuales el 80% se alinean con el transcriptoma del ratón.

Se analizaron las lecturas de ADN (captura de exoma) de B16F10 y C57BL/6 para identificar mutaciones somáticas. El análisis de variación del número de copias (Sathirapongsasuti JF et al., Bioinformatics 2011; 27: 2648-54) demostró

amplificaciones y eliminaciones de ADN en B16F10, incluida la eliminación homocigótica del supresor tumoral *Cdkn2a* (inhibidor 2A de quinasa dependiente de ciclina, p16Ink4A). Centrándose en mutaciones puntuales para identificar posibles mutaciones inmunogénicas, se identificaron 3570 mutaciones puntuales somáticas a FDR ≤ 0,05 (Figura 9B). La clase más frecuente de mutaciones fueron las transiciones C > T/G > A, típicamente resultantes de la luz ultravioleta (Pfeifer GP et al., *Mutat Res* 2005; 571: 19-31). De estas mutaciones somáticas, 1392 ocurren en transcritos, con 126 mutaciones en regiones no traducidas. De las 1266 mutaciones en las regiones de codificación, 962 causan cambios de proteínas no sinónimas y 563 de estos ocurren en genes expresados (Figura 9B).

Asignación de mutaciones identificadas a genes portadores y validación

Cabe destacar que muchos de los genes mutados (962 genes que contienen mutaciones puntuales somáticas no sinónimas) se han asociado previamente con los fenotipos del cáncer. Se encontraron mutaciones en genes supresores tumorales establecidos, incluidos *Pten*, *Trp53* (también llamado p53) y *Tp63*. En *Trp53*, el supresor tumoral mejor establecido (Zilfou JT et al., *Cold Spring Harb Perspect Biol* 2009; 1:a001883), la mutación de asparagina a ácido aspártico en la posición de proteína 127 (p.N127D) se localiza en el dominio de unión al ADN y SIFT predice que alterará la función. *Pten* contenía dos mutaciones (p.A39V, p.T131P), y se pronostica que ambas tendrán un impacto perjudicial sobre la función de la proteína. La mutación p.T131P es adyacente a una mutación (p.R130M) que disminuye la actividad de fosfatasa (Dey N et al., *Cancer Res* 2008; 68: 1862-71). Además, se encontraron mutaciones en genes asociados con vías de reparación del ADN, tales como *Brca2* (cáncer de mama 2, inicio temprano), *Atm* (ataxia telangiectasia mutada), *Ddb1* (proteína 1 de unión al ADN específica de daño) y *Rad9b* (homólogo B de RAD9). Además, se producen mutaciones en otros genes asociados a tumores, incluidos *Aim1* (supresor tumoral "Ausente en el melanoma 1"), *Flt1* (oncogén *Vegfr1*, tirosina quinasa 1 relacionada con aletas), *Pml* (supresor tumoral "leucemia promielocítica"), *Fat1* ("FAT supresor tumoral homólogo 1"), *Mdm1* (proteína nuclear de unión a TP53), *Mta3* (familia 1 asociada a metástasis, miembro 3) y *Alk* (receptor de tirosina quinasa del linfoma anaplásico). Se encontró una mutación en p.S144F en *Pdgfra* (receptor del factor de crecimiento derivado de plaquetas, polipéptido alfa), un receptor de tirosina quinasa unido a la membrana celular de la vía MAPK/ERK, previamente identificado en tumores (Verhaak RG et al., *Cancer Cell* 2010; 17: 98-110). Se produce una mutación en p.L222V en *Casp9* (caspasa 9, cisteína peptidasa relacionada con la apoptosis). CASP9 escinde proteolíticamente la polimerasa poli(ADP-ribosa) (PARP), regula la apoptosis y se ha relacionado con varios tipos de cáncer (Hajra KM et al., *Apoptosis* 2004; 9: 691-704). La mutación que se encontró puede afectar potencialmente la señalización de PARP y apoptosis. Lo más interesante es que no se encontraron mutaciones en *Braf*, *c-Kit*, *Kras* o *Nras*. Sin embargo, se identificaron mutaciones en *Rassf7* (proteína asociada a RAS) (p.S90R), *Ksr1* (supresor de quinasa de ras 1) (p.L301V) y *Atm* (vía PI3K) (p.K91T), todos los cuales se predicen que tienen un impacto significativo en la función proteica. *Trrap* (proteína asociada al dominio de transformación/transcripción) se identificó a principios de este año en muestras de melanoma humano como un nuevo objetivo potencial de melanoma (Wei X et al., *Nat Genet* 2011; 43: 442-6). En B16F10, se produce una mutación *Trrap* en p.K2783R y se predice que perturba el dominio FAT de quinasa relacionado con fosfatidilinositol quinasa (PIK) superpuesto.

De las 962 mutaciones no sinónimas identificadas usando NGS, se seleccionaron 50 mutaciones, incluidas 41 con FDR < 0,05, para validación basada en PCR y pruebas de inmunogenicidad. Los criterios de selección fueron la ubicación en un gen expresado (RPKM > 10) y la inmunogenicidad prevista. Cabe destacar que se pudieron validar las 50 mutaciones (Tabla 6, Figura 9B).

Tabla 6: Mutaciones seleccionadas para validación. Desde la izquierda: ID asignada, símbolo del gen, sustitución y posición de aminoácidos, nombre del gen, localización y tipo subcelular predicho (Ingenuity).

ID	Símbolo	Cambio	Entrada del nombre del gen	Localización subcelular	Tipo
MUT1	<i>Fzd7</i>	p.G304A	receptor 7 de la familia rizada	Membrana plasmática	Receptor acoplado a la proteína G
MUT2	<i>Xpot</i>	p-I830S	exportina, ARNt (receptor de exportación nuclear para ARNt)	Núcleo	otro
MUT3	<i>Ranbp2</i>	p.Q2871H	proteína 2 de unión a RAN	Núcleo	enzima
MUT4	<i>Dnajb12</i>	p.P54T	homólogo de DnaJ (Hsp40), subfamilia B, miembro 12	Citoplasma	otro

ES 2 746 233 T3

ID	Símbolo	Cambio	Entrada del nombre del gen	Localización subcelular	Tipo
MUT5	<i>Eef2</i>	p.G795A	factor 2 de alargamiento de la traducción eucariota	Citoplasma	regulador de traducción
MUT6	<i>Ptrf</i>	p.D382G	polimerasa I y factor de liberación del transcrito	Núcleo	regulador de transcripción
MUT7	<i>Trp53</i>	p.N128D	protein tumoral p53	Núcleo	regulador de transcripción
MUT8	<i>Ddx23</i>	p.V602A	polipéptido 23 de la caja DEAD (Asp-Glu-Ala-Asp)	Núcleo	enzima
MUT9	<i>Golgb1</i>	p.E2855D	golgina B1	Citoplasma	otro
MUT10	<i>Pcdhga11</i>	p.G82R	protocadherina gamma subfamilia A, 11	Membrana plasmática	otro
MUT11	<i>Snx15</i>	p.E211G	clasificación de nexina 15	Citoplasma	transportador
MUT12	<i>Gnas</i>	p.S112G	locus del complejo GNAS (proteína de unión al nucleótido guanina, estimulación alfa)	Membrana plasmática	enzima
MUT13	<i>Fndc3b</i>	p.C561W	dominio tipo III de fibronectina que contiene 3B	Citoplasma	otro
MUT14	<i>Sbno1</i>	p.P309T	homólogo 1 de muesca de fresa (<i>Drosophila</i>)	desconocido	enzima
MUT15	<i>Pi4k2b</i>	p.R344Q	fosfatidilinositol 4-quinasa tipo 2 beta	Citoplasma	quinasa
MUT16	<i>Thumpd3</i>	p.T243S	dominio THUMP que contiene 3	desconocido	otro
MUT17	<i>Tnpo3</i>	p.G504A	transportina 3	Citoplasma	otro
MUT18	<i>Numa1</i>	p.Q447K	proteína 1 del aparato mitótico nuclear	Núcleo	otro
MUT19	<i>Wwp2</i>	p.E742K	dominio WW que contiene proteina ligasa 2 ubiquitina E3	Citoplasma	enzima
MUT20	<i>Tubb3</i>	p.G402A	tubulina, beta 3	Citoplasma	otro
MUT21	<i>Atp11a</i>	p.R522S	ATPasa, clase VI, tipo 11A	Membrana plasmática	transportador
MUT22	<i>Asf1b</i>	p.A141P	homólogo B de la función antisilenciamiento de ASF1 (<i>S. cerevisiae</i>)	Núcleo	otro
MUT23	<i>Wdr82</i>	p.I221L	dominio 82 de la repetición WD	Núcleo	otro
MUT24	<i>Dag1</i>	p.P425A	distroglicano 1 (glicoproteína 1 asociada a distrofina)	Membrana plasmática	receptor transmembrana
MUT25	<i>Plod2</i>	p.F530V	procolágeno-lisina, 2-oxoglutarato 5-dioxigenasa 2	Citoplasma	enzima

ES 2 746 233 T3

ID	Símbolo	Cambio	Entrada del nombre del gen	Localización subcelular	Tipo
MUT26	<i>Orc2</i>	p.F278V	complejo de reconocimiento de origen, subunidad 2	Núcleo	otro
MUT27	<i>Obsl1</i>	p.T1764M	obscurina tipo 1	desconocido	otro
MUT28	<i>Ppp1r7</i>	p.L170P	proteína fosfatasa 1, subunidad reguladora 7 (inhibidora)	Núcleo	fosfatasa
MUT29	<i>Mthfd11</i>	p.F294V	tipo metileno tetrahidrofolato deshidrogenasa (dependiente de NADP+) 1	Citoplasma	enzima
MUT30	<i>Kif18b</i>	p.K739N	miembro 18B de la familia de las quinesinas	desconocido	otro
MUT31	<i>Ascc2</i>	p.A59G	subunidad 2 del complejo cointegrador de activación de la señal	desconocido	otro
MUT32	<i>Itsn2</i>	p.S1551R	intersectina 2	Citoplasma	otro
MUT33	<i>Pbk</i>	p.V145D	quinasa de unión a PDZ	Citoplasma	quinasa
MUT34	<i>Klhl22</i>	p.F179V	tipo kelch 22 (<i>Drosophila</i>)	desconocido	otro
MUT35	<i>Ddb1</i>	p.L438I	proteína 1 de unión a ADN específica del daño, 127kDa	Núcleo	otro
MUT36	<i>Tm9sf3</i>	p.Y382H	miembro 3 de la superfamilia transmembrana 9	Citoplasma	transportador
MUT37	<i>Dpf2</i>	p.F275V	D4, familia 2 de dedos PHD de zinc dobles	Núcleo	otro
MUT38	<i>Atrn</i>	p.5745N	atractina	Espacio extracelular	otro
MUT39	<i>Snx5</i>	p.R373Q	clasificación de nexina 5	Citoplasma	transportador
MUT40	<i>Armc1</i>	p.585I	repetición de armadillo que contiene 1	Citoplasma	otro
MUT41	<i>Ash1l</i>	p.L632I	tipo ash1 (ausente, pequeña, u homeótica) (<i>Drosophila</i>)	Núcleo	regulador de transcrito
MUT42	<i>S100a13</i> <i>2510039O18</i>	p.S18C	proteína A13 de unión al calcio de S100	Citoplasma	otro
MUT43	<i>Rik</i>	p.E391K	KIAA2013	desconocido	otro
MUT44	<i>Cpsf3l</i>	p.D314N	similar al factor 3 específico de escisión y poliadenilación	Núcleo	otro
MUT45	<i>Mkrn1</i>	p.N346Y	proteína 1 del dedo anular de makorina	desconocido	otro
MUT46	<i>Actn4</i>	p.F835V	actinina, alfa 4	Citoplasma	otro
MUT47	<i>Rpl13a</i>	p.A24G	proteína ribosomal L13a	Citoplasma	otro

ID	Símbolo	Cambio	Entrada del nombre del gen	Localización subcelular	Tipo
MUT48	<i>Def8</i>	p.R255G	diferencialmente expresada en el homólogo de FDCP 8 (ratón)	desconocido	otro
MUT49	<i>Fat1</i>	p.I1940M	Dominio sema del homólogo 1 supresor tumoral FAT (<i>Drosophila</i>), dominio de inmunoglobulina (Ig), dominio básico corto,	Membrana plasmática	otro
MUT50	<i>Sema3b</i>	p.L663V	secretado, (semaforina) 3B	Espacio extracelular	otro

La Figura 9C muestra las ubicaciones de los cromosomas B16F10, densidad de genes, expresión génica, mutaciones y mutaciones filtradas (anillos internos).

Pruebas *in vivo* de pruebas de inmunogenicidad con péptidos largos que representan mutaciones

5 Para proporcionar antígenos para la prueba de inmunogenicidad de estas mutaciones, se emplean péptidos largos que tienen muchas ventajas sobre otros péptidos para la inmunización (Melief CJ y van der Burg SH, *Nat Rev Cancer* 2008; 8: 351-60). Los péptidos largos son capaces de inducir células T CD8+ así como CD4+ específicas de antígeno (Zwaveling S et al., *Cancer Res* 2002; 62: 6187-93; Bijker MS et al., *J Immunol* 2007; 179: 5033-40). Además, los péptidos largos requieren que el procesamiento se presente en las moléculas del MHC. Tal captación se realiza de manera más eficiente por las células dendríticas, que son óptimas para preparar una potente respuesta de células T. 10 Los péptidos adecuados, por el contrario, no requieren recorte y se cargan de manera exógena en todas las células que expresan moléculas del MHC, incluidas las células B y T no activadas, lo que conduce a la inducción de tolerancia y fratricida (Toes RE et al., *J Immunol* 1996; 156 : 3911-8; Su MW et al., *J Immunol* 1993; 151: 658-67). Para cada una de las 50 mutaciones validadas, se diseñaron péptidos de 27 aminoácidos de longitud con el aminoácido mutado o de tipo silvestre colocado centralmente. Por lo tanto, cualquier epítipo potencial del MHC de clase I y clase II de 8 a 14 aminoácidos de longitud que porte la mutación podría procesarse a partir de este péptido precursor. Como adyuvante para la vacunación con péptidos, se utilizó poli (I:C) que se sabe que promueve la presentación cruzada y aumenta la eficacia de la vacuna (Datta SK et al., *J Immunol* 2003; 170: 4102-10; Schulz O et al., *Nature* 2005; 433: 887-92). Las 50 mutaciones se probaron *in vivo* en ratones para la inducción de células T. De manera impresionante, se encontró 20 que 16 de 50 péptidos que codifican mutaciones provocan respuestas inmunes en ratones inmunizados. Las células T inducidas mostraron diferentes patrones de reactividad (Tabla 7).

25 Tabla 7: Resumen de las reactividades de células T determinadas en forma consecutiva a la vacunación con péptido codificador de mutación. 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, cuando cualquiera de las pruebas produjo un valor de $p < 0,05$ y la media del números de puntos fue > 30 puntos/ 5×10^5 células efectoras. Las reactividades se clasificaron por la media del números de punto: < 30 ; +: > 30 ; ++: > 50 ; +++ > 200 puntos/pozo.

Mutación	Símbolo del gen	Reactividad contra mutación	Reactividad contra WT	Mutación	Símbolo del gen	Reactividad contra mutación	Reactividad contra WT
MUT01	<i>Fzd7</i>	-	-	MUT26	<i>Orc2</i>	-	-
MUT02	<i>Xpot</i>	-	-	MUT27	<i>Obsl1</i>	-	-
MUT03	<i>Ranbp2</i>	-	-	MUT28	<i>Ppp1r7</i>	+	+
MUT04	<i>Dnajb12</i>	-	-	MUT29	<i>Mthfd11</i>	+	-
MUT05	<i>Eef2</i>	+++	+++	MUT30	<i>Kif18b</i>	+++	-
MUT06	<i>Ptrf</i>	-	-	MUT31	<i>Ascc2</i>	-	-

Mutación	Símbolo del gen	Reactividad contra mutación	Reactividad contra WT	Mutación	Símbolo del gen	Reactividad contra mutación	Reactividad contra WT
MUT07	<i>Trp53</i>	-	-	MUT32	<i>Itsn2</i>	-	-
MUT08	<i>Ddx23</i>	-	-	MUT33	<i>Pbk</i>	-	-
MUT09	<i>Golgb1</i>	-	-	MUT34	<i>Klhl22</i>	-	-
MUT10	<i>Pcdhga11</i>	-	-	MUT35	<i>Ddb1</i>	-	-
MUT11	<i>Snx15</i>	-	-	MUT36	<i>Tm9sf3</i>	+	-
MUT12	<i>Gnas</i>	+	-	MUT37	<i>Dpf2</i>	-	-
MUT13	<i>Fndc3b</i>	-	-	MUT38	<i>Atrn</i>	-	-
MUT14	<i>Sbno1</i>	-	-	MUT39	<i>Snx5</i>	-	-
MUT15	<i>Pi4k2b</i>	-	-	MUT40	<i>Armc1</i>	-	-
MUT16	<i>Thumpd3</i>	-	-	MUT4 1	<i>Ash1l</i>	-	-
MUT17	<i>Tnpo3</i>	+++	++	MUT42	<i>S100a13</i>	-	-
MUT18	<i>Numa1</i>	-	-	MUT43	<i>Rik</i>	-	-
MUT19	<i>Wwp2</i>	-	-	MUT44	<i>Cpsf31</i>	+++	++
MUT20	<i>Tubb3</i>	+++	-	MUT45	<i>Mkru1</i>	++	++
MUT21	<i>Atp11a</i>	-	-	MUT46	<i>Actn4</i>	++	+
MUT22	<i>Asf1b</i>	++	++	MUT47	<i>Rpl13a</i>	-	-
MUT23	<i>Wdr82</i>	-	-	MUT48	<i>Def8</i>	++	++
MUT24	<i>Dag1</i>	++	+	MUT49	<i>Fat1</i>	-	-
MUT25	<i>Plod2</i>	+++	++	MUT50	<i>Sema3b</i>	+++	++

5 Once péptidos indujeron una respuesta inmune que reconoce preferentemente el epítipo mutado. Esto se ejemplifica para ratones inmunizados con las mutaciones 30 (MUT30, *Kif18b*) y 36 (MUT36, *Plod2*) (Figura 10A). Las pruebas ELISPOT revelaron fuertes respuestas inmunes específicas de mutaciones sin reactividad cruzada contra el péptido de tipo silvestre o un péptido de control no relacionado (VSV-NP). Con cinco péptidos, incluidas las mutaciones 05 (MUT05, *Eef2*) y 25 (MUT25, *Plod2*) (Figura 10A), se obtuvieron respuestas inmunitarias con reconocimiento comparable tanto del péptido mutado como del péptido de tipo silvestre. La mayoría de los péptidos mutados no fueron capaces de inducir respuestas significativas de células T como se ejemplifica en las mutaciones 01 (MUT01, *Fzd7*), 02 (MUT02, *Xpot*) y 07 (MUT07, *Trp53*). Las respuestas inmunitarias inducidas por varias de las mutaciones descubiertas estaban bien en el intervalo de inmunogenicidad (500 puntos/5x10⁵ células) generadas inmunizando ratones como control positivo con un epítipo descrito del MHC de clase I de la proteína 2 relacionada con tirosinasa del antígeno tumoral del melanoma murino (Trp2180-188, Figura 10A) (Bloom MB et al., Exp Med 1997; 185: 453-9; Schreurs MW et al. Cancer Res 2000; 60: 6995-7001). Para péptidos seleccionados que inducen una fuerte respuesta de células T específicas de mutación, se confirmó el reconocimiento inmunitario mediante un enfoque independiente.

10 En lugar de péptidos largos, se usó ARN transcrito *in vitro* (ARN IVT) que codifica los fragmentos de péptidos mutados

15

MUT17, MUT30 y MUT44 para la lectura inmunológica. Las BMDC transfectadas con ARN que codifica la mutación o ARN irrelevante sirvieron como células presentadoras de antígeno (APC) en un ensayo ELISPOT, mientras que las células del bazo de ratones inmunizados sirvieron como población de células efectoras. Las BMDC transfectadas con ARNm que codifica MUT17, MUT30 y MUT44 fueron reconocidas específica y fuertemente por los esplenocitos de ratones inmunizados con los péptidos largos respectivos (Figura 10B). Se registró una reactividad significativamente menor frente a las BMDC transfectadas con ARN de control, lo que probablemente se deba a la activación inespecífica de las BMDC por el ARN monocatenario (prueba t de Student; MUT17: $p = 0,0024$, MUT30: $p = 0,0122$, MUT44: $p = 0,0075$). Estos datos confirman que las células T específicas de mutación inducida en efecto reconocen epítopos procesados endógenamente. Dos mutaciones que inducen un reconocimiento preferido de epítopos mutados están en los genes *Actn4* y *Kif18b*. La mutación somática en ACTN4 (actinina, alfa 4) está en p.F835V en el dominio de la proteína "mano EF" de unión a calcio. Si bien tanto SIFT como POLYPHEN predicen un impacto significativo de esta mutación en la función de la proteína, el gen no es un oncogén establecido. Sin embargo, las células T específicas de mutación contra ACTN4 se han asociado recientemente con un resultado positivo para el paciente (Echchakir H et al., Cancer Res 2001; 61: 4078-83). KIF18B (miembro 18B de la familia de quinesina) es una quinesina con actividad motora de microtúbulos y unión a ATP y nucleótidos que está involucrada en la regulación de la división celular (Lee YM et al., Gene 2010; 466: 16-25) (Figura 10C). La secuencia de ADN en la posición que codifica p.K739 es homogénea en la referencia C57BL/6, mientras que las lecturas de ADN de B16F10 revelan una mutación somática heterocigótica. Ambos nucleótidos se detectaron en las lecturas de Seq-ARN de B16F10 y se validaron mediante secuenciación de Sanger. KIF18B no se ha asociado previamente con un fenotipo de cáncer. La mutación p.K739N no está localizada en un dominio de proteína funcional o conservado conocido (Figura 10C, abajo) y, por lo tanto, lo más probable es que sea una mutación pasajera y no una conductora. Estos ejemplos sugieren una falta de correlación entre la capacidad de inducir una respuesta inmune que reconoce la mutación y una relevancia funcional o inmunológica.

Evaluación *in vivo* de la actividad antitumoral de las candidatas a vacunas

Para evaluar si las respuestas inmunes provocadas *in vivo* se traducen en efectos antitumorales en ratones portadores de tumores, se eligieron MUT30 (mutación en *Kif18b*) y MUT44 como ejemplos. Se ha demostrado que estas mutaciones inducen una fuerte reacción inmune preferentemente contra el péptido mutado y se procesan endógenamente (Figura 10A, B). El potencial terapéutico de la vacunación con péptidos mutados se exploró inmunizando ratones con MUT30 o MUT44 y adyuvante 3 y 10 días después del injerto con $7,5 \times 10^5$ B16F10. El crecimiento de tumores fue inhibido por ambas vacunas de péptidos en comparación con el grupo de control (Figura 11A). Como B16F10 es un tumor de crecimiento muy agresivo, también se probaron respuestas inmunes protectoras. Los ratones fueron inmunizados con péptido MUT30, se inocularon s.c. con $7,5 \times 10^5$ células B16F10 4 días después y se reforzaron con MUT30 2 y 9 días después de la exposición al tumor. Se observó una protección completa del tumor y la supervivencia del 40% de los ratones tratados con MUT30, mientras que todos los ratones del grupo tratado con control murieron dentro de los 44 días (Figura 11B a la izquierda). En esos ratones, que desarrollaron tumores a pesar de la inmunización con MUT30, el crecimiento de los tumores fue más lento, lo que resultó en un alargamiento de la supervivencia media de 6 días en comparación con el grupo de control (Figura 11B derecha). Estos datos implican que la vacunación contra una sola mutación puede conferir efectos antitumorales.

Inmunización con ARN codificadores de mutación

Las 50 mutaciones validadas de la línea celular de melanoma B16F10 se usaron para construir diferentes vacunas de ARN. Las secuencias de ADN que representan una (Monoepítipo), dos (Biepítipo), o 16 mutaciones diferentes (Poliepítipo), se construyeron utilizando 50 aminoácidos (aa) con la mutación en la posición 25 (Biepítipo) o 27 aa con la mutación en la posición 14 (Mono y Poliepítipo) y se separaron mediante un enlazador de glicina/serina de 9aa. Estos constructos se clonaron en la cadena principal de pST1-2BgUTR-A120 para la transcripción *in vitro* de ARNm (Holtkamp et al., Blood 2006; 108: 4009-17).

Para probar la capacidad *in vivo* de inducir respuestas de células T contra las diferentes vacunas de ARN se inmunizaron grupos de tres ratones C57BL/6 mediante la formulación del ARN con lipofectamina de ARNiMAX y la posterior inyección intravenosa. Después de 5 inmunizaciones, se sacrificaron los ratones y se analizaron los esplenocitos para detectar respuestas de células T específicas de mutación usando tinción intracelular de citoquinas y análisis ELISPOT de IFN- γ después de la reestimulación con el péptido codificador de mutación correspondiente o péptido de control (VSV-NP).

La Figura 12 muestra un ejemplo para cada diseño de vacuna. En la fila superior, los ratones se vacunaron con el ARN monoepítipo que codifica MUT30 (mutación en *Kif18b*), que induce células T CD4+ específicas de MUT30 (véase el ejemplo del gráfico de FACS). En la fila central, el gráfico y el diagrama de FACS muestran la inducción de células T CD4+ específicas de MUT08 (mutación en *Ddx23*) después de la inmunización con el Biepítipo que codifica MUT33 y MUT08. En la fila inferior, los ratones se inmunizaron con un poliepítipo que codifica 16 mutaciones diferentes que incluyen MUT08, MUT33 y MUT27 (véase la Tabla 8). El gráfico y el diagrama de FACS ilustran que las células T reactivas a MUT27 son de un fenotipo CD8.

Tabla 8. Resumen de mutaciones y nombres de genes codificados por las vacunas de ARN mono, bi y poliepítipo

Constructo	Mutación codificada	Anotación Gene
Monoepítopo	MUT30	<i>Kif18b</i>
Biepítopo	MUT33	<i>Pbk</i>
	MUT08	<i>Ddx23</i>
Poliepítopo	MUT01	<i>Fzd7</i>
	MUT02	<i>Xpot</i>
	MUT03	<i>Ranbp2</i>
	MUT04	<i>Dnajb12</i>
	MUT05	<i>Eef2</i>
	MUT06	<i>Ptrf</i>
	MUT07	<i>Trp53</i>
	MUT08	<i>Ddx23</i>
	MUT26	<i>Orc2</i>
	MUT27	<i>Obsl1</i>
	MUT28	<i>Ppp1r7</i>
	MUT29	<i>Mthfd1l</i>
	MUT30	<i>Kif18b</i>
	MUT31	<i>Ascc2</i>
	MUT32	<i>Itsn2</i>
MUT33	<i>Pbk</i>	

5 Se usó el mismo Poliepítopo para generar los datos mostrados en la Figura 13. El gráfico muestra los datos de ELISPOT después de la reestimulación de esplenocitos con péptidos de control (VSV-NP), MUT08, MUT27 y MUT33, lo que demuestra que la vacuna de Poliepítopo puede inducir respuestas específicas de células T contra varias mutaciones diferentes.

Tomados en conjunto, los datos muestran la posibilidad de inducir células T específicas de mutación usando mono, bi y poliepítopos codificados por ARN. Además, los datos muestran la inducción de células T CD4+ y CD8+ y la inducción de varias especificidades diferentes a partir de un constructo.

Inmunización con epítomos modelo

10 Para caracterizar aún más el diseño de la vacuna de ARN poliepitépica, se construyó una secuencia de ADN, que incluía cinco epítomos de modelos conocidos diferentes, incluyendo un epítopo del MHC de clase II (ovoalbúmina clase I (SIINFEKL), clase II (OVA clase II), nucleoproteína de influenza (Inf-NP), nucleoproteína del virus de la vesiculostomatitis (VSV-NP) y la proteína 2 relacionada con la tirosinasa (Trp2)). Los epítomos se separaron con el mismo enlazador de glicina/serina de 9 aa utilizado para el Poliepítopo de mutación. Estos constructos se clonaron en
15 la cadena principal de pST1-2BgUTR-A120 para la transcripción *in vitro* de ARNm.

El ARN transcrito *in vitro* se usó para vacunar a cinco ratones C57BL/6 mediante inmunización intranodal (cuatro inmunizaciones con 20 µg de ARN en los ganglios linfáticos inguinales). Cinco días después de la última inmunización, se tomaron muestras de sangre y esplenocitos de los ratones para su análisis. La Figura 14A muestra el análisis ELISPOT de IFN-γ de los esplenocitos reestimulados con los péptidos indicados. Se puede ver claramente que los
20 tres epítomos del MHC de clase I (SIINFEKL, Trp2 y VSV-NP) inducen un número muy alto de células T CD8+ específicas de antígeno. También el epítopo del MHC de clase II OVA de clase II induce una fuerte respuesta de células T CD4+. El cuarto epítopo del MHC de clase I se analizó mediante tinción de células T CD8+ específicas de

Inf-NP con un complejo de MHC-péptido pentamérico marcado con fluorescencia (Pentámero) (Figura 14B).

Estos datos demuestran que el diseño de poliepitopo que usa el enlazador de glicina/serina para separar diferentes epítomos inmunogénicos del MHC de clase I y clase II es capaz de inducir células T específicas contra cada epítomo codificado, independientemente de su inmunodominancia.

5 Respuesta antitumoral después de la terapia con una vacuna de ARN poliepitópico que codifica la mutación

El mismo Poliepitopo que se analizó en la Figura 13 para determinar la inmunogenicidad se usó para investigar la actividad antitumoral de los ARN que codifican la mutación contra las células tumorales B16F10. En detalle, se inocularon subcutáneamente grupos de ratones C57BL/6 ($n = 10$) con 1×10^5 células de melanoma B16F10 en el flanco. Los días 3, 6, 10, 17 y 21, los ratones se inmunizaron con el ARN poliepitópico usando un reactivo de transfección liposomal. El grupo de control se inyectó con liposomas solos.

La Figura 21 muestra las curvas de supervivencia de los grupos, revelando una supervivencia media fuertemente mejorada de 27 días con 1 de 10 ratones sobreviviendo sin tumor en comparación con la supervivencia media de 18,5 días en el grupo control.

Respuesta antitumoral después de la terapia con una combinación de péptido mutado y normal

15 La actividad antitumoral de las mutaciones validadas se evaluó mediante un experimento terapéutico en un tumor *in vivo* usando el MUT30 como una vacuna peptídica. En detalle, se inocularon subcutáneamente grupos de ratones C57BL/6 ($n = 8$) con 1×10^5 células de melanoma B16F10 en el flanco. En los días 3, 10 y 17, los ratones se inmunizaron usando poli I:C como adyuvante con MUT30, proteína 2 relacionada con tirosinasa (Trp2₁₈₀₋₁₈₈) o una combinación de ambos péptidos. Trp2 es un epítipo CD8+ conocido expresado por las células de melanoma B16F10.

20 La Figura 15A muestra el crecimiento tumoral medio de los grupos. Se puede ver claramente que hasta el día 28 el crecimiento tumoral se inhibe casi por completo en el grupo que se inmunizó con la combinación del epítipo conocido de células T CD8+ y el MUT30 inductor de células T CD4+. El epítipo conocido de Trp2 por sí solo no es suficiente para proporcionar un buen efecto antitumoral en este contexto, pero ambos grupos de terapia única (MUT30 y Trp2) aún proporcionan una inhibición del crecimiento tumoral en comparación con el grupo no tratado al comienzo del experimento hasta día 25. Estos datos se ven reforzados por las curvas de supervivencia que se muestran en la Figura 15 B. Claramente, la supervivencia media se incrementa por los ratones inyectados con los péptidos individuales, con 1/8 ratones sobreviviendo en el grupo con la vacuna de Trp2. Además, el grupo tratado con ambos péptidos muestra una supervivencia media incluso mejor con la supervivencia de 2/8 ratones.

Tomados juntos, ambos epítomos actúan de manera sinérgica para proporcionar un fuerte efecto antitumoral.

30 Ejemplo 9: Marco para la detección y aplicación de mutaciones somáticas basadas en la confianza en células de melanoma B16-F10

NGS es imparcial ya que permite un descubrimiento de alto rendimiento de variaciones dentro de un genoma completo o regiones específicas, como los exones codificadores de proteínas.

35 Sin embargo, si bien es revolucionaria, la plataforma de NGS todavía es propensa a errores que conducen a llamadas de variación erróneas. Además, la calidad de los resultados depende de parámetros de diseño experimentales y metodologías de análisis. Si bien las llamadas de variación generalmente incluyen puntajes diseñados para diferenciar las variaciones verdaderas de los errores, la utilidad de estos puntajes no se comprende completamente, ni su interpretación con respecto a la optimización de los experimentos. Esto es particularmente cierto cuando se comparan estados de tejido, tal como la comparación de tejido tumoral y normal para mutaciones somáticas. Como consecuencia, los investigadores se ven obligados a confiar en la experiencia personal para determinar parámetros experimentales y umbrales de filtrado arbitrarios para seleccionar mutaciones. Nuestro estudio tiene como objetivo a) establecer un marco para comparar parámetros y métodos para identificar mutaciones somáticas y b) asignar un valor de confianza a las mutaciones identificadas. Se secuenciaron muestras por triplicado de ratones C57BL/6 y la línea celular de melanoma B16-F10. Con estos datos, se formuló la tasa de descubrimiento falso de mutaciones somáticas detectadas, una medida que luego se usó para evaluar el software de descubrimiento de mutaciones y los protocolos de laboratorio existentes.

40 Varios factores experimentales y algorítmicos contribuyen a la tasa de falsos positivos para las variaciones encontradas por NGS [Nothnagel, M. et al., Hum. Gineta. 23 de febrero de 2011 [Epub antes de la impresión]]. Las fuentes de error incluyen artefactos de PCR, sesgos en el cebado [Hansen, K.D., et al., Nucleic. Acids Res. 38, e131 (2010); Taub, M.A. et al., Genome Med. 2, 87 (2010)] y enriquecimiento dirigido [Bainbridge, M.N. et al., Genome Biol. 11, R62 (2010)], efectos de secuencia [Nakamura, K. et al., Acids Res. (2011) publicado por primera vez en línea el 16 de mayo de 2011 doi: 10.1093/nar/gkr344], llamadas de base que causan errores de secuencia [Kircher, M et al., Genome Biol. 10, R83 (2009). Epub 14 de agosto de 2009] y la alineación leída [Lassmann, T. et al., Bioinformatics 27, 130-131 (2011)], lo que causa una variación en la cobertura y errores de secuencia que influyen en el análisis posterior, por ejemplo, llamado de variante invocando indels [Li, H., Bioinformatics 27, 1157-1158 (2011)].

No se ha descrito ningún modelo estadístico general para describir el impacto de diferentes fuentes de error en las llamadas de mutación somática; solo se cubren los aspectos individuales sin eliminar todo sesgo. Los métodos informáticos recientes para medir la cantidad esperada de llamadas de mutación positivas falsas incluyen la utilización de la relación transición/transversión de un conjunto de variaciones [Zhang, Z., Gerstein, M., *Nucleic Acids Res* 31, 5338-5348 (2003); DePristo, MA et al., *Nature Genetics* 43, 491-498 (2011)], aprendizaje automático [DePristo, MA et al., *Nature Genetics* 43, 491-498 (2011)] y errores de herencia al trabajar con genomas familiares [Ewen KR et al., *Am. J. Hum. Genet.* 67, 727-736 (2000)] o muestras agrupadas [Druley, T.E. et al., *Nature Methods* 6, 263-265 (2009); Bansal, V., *Bioinformatics* 26, 318-324 (2010)]. Para fines de optimización, Druley et al. [Druley, T.E. et al., *Nature Methods* 6, 263- 265 (2009)] se basaron en fragmentos de secuencia de plásmidos cortos, que sin embargo podrían no ser representativos de la muestra. Para un conjunto de variaciones de un solo nucleótido (SNV) y experimentos seleccionados, es factible una comparación con las SNV identificadas por otras técnicas [Van Tassell, C.P. et al., *Nature Methods* 5, 247-252 (2008)] pero es difícil de evaluar en términos de mutaciones somáticas novedosas.

Usando un proyecto de secuenciación de exoma como ejemplo, se propone el cálculo de una tasa de descubrimiento falso (FDR) basada únicamente en datos de NGS. El método no solo es aplicable a la selección y priorización de objetivos de diagnóstico y terapéuticos, sino que también admite el desarrollo de algoritmos y métodos al permitir definir recomendaciones basadas en la confianza para experimentos similares.

Para descubrir mutaciones, el ADN del tejido de la cola de tres ratones C57BL/6 (black6) (compañeros de camada) y el ADN de las células de melanoma B16-F10 (B16), por triplicado, se enriquecieron individualmente en los exones codificadores de proteínas (exoma completo de ratón Sure Select de Agilent), dando como resultado 6 muestras. El ARN se extrajo de las células B16 por triplicado. Las lecturas de extremo único de 50 nt (1 x 50 nt) y de extremo emparejado de 100 nt (2 x 100 nt) se generaron en un Illumina HiSeq 2000. Cada muestra se cargó en un carril individual, lo que resultó en un promedio de 104 millones de lecturas por carril. Las lecturas de ADN se alinearon con el genoma de referencia del ratón usando bwa [Li, H. Durbin, R., *Bioinformatics* 25, 1754-1760 (2009)] y las lecturas de ARN se alinearon con Bowtie [Langmead, B. et al., *Genome Biol.* 10, R25 (2009)]. Se logró una cobertura media de 38 veces para el 97% de las regiones objetivo para las bibliotecas de 1 x 50 nt, mientras que el experimento de 2 x 100 nt arrojó una cobertura promedio de 165 veces para el 98% de las regiones objetivo.

Las variaciones somáticas se identificaron de forma independiente utilizando los paquetes de software SAMtools [Li, H. et al., *Bioinformatics* 25, 2078-2079 (2009)], GATK [DePristo, MA et al., *Nature Genetics* 43, 491-498 (2011)] y SomaticSNIper [Ding, L. et al., *Hum. Genet* (2010) que se publicó por primera vez en línea el 15 de septiembre de 2010] (Fig. 16) comparando las variaciones de un solo nucleótido encontradas en las muestras de B16 con los loci correspondientes en las muestras de black6 (las células B16 se derivaron originalmente de un ratón black6). Las posibles mutaciones se filtraron de acuerdo con las recomendaciones de los respectivos autores de software (SAMtools y GATK) o seleccionando un umbral inferior apropiado para la puntuación somática de SomaticSNIper, respectivamente.

Para crear una tasa de descubrimiento falso (FDR) para el descubrimiento de mutaciones, primero se intersecan los sitios de mutación y se obtienen 1.355 mutaciones somáticas de alta calidad como consenso entre los tres programas (Fig. 17). Sin embargo, las diferencias observadas en los resultados de las herramientas de software aplicadas son sustanciales. Para evitar conclusiones erróneas, se desarrolló un método para asignar una FDR a cada mutación usando las réplicas. Las repeticiones técnicas de una muestra deben generar resultados idénticos y cualquier mutación detectada en esta "misma comparación versus la misma comparación" es un falso positivo. Por lo tanto, para determinar la tasa de descubrimiento falso para la detección de mutaciones somáticas en una muestra de tumor en relación con una muestra normal ("comparación de tumores"), se puede usar una repetición técnica de la muestra normal como referencia para estimar el número de falsos positivos.

La Figura 18A muestra ejemplos de variaciones encontradas en los datos de black6/B16, que incluyen una mutación somática (izquierda), una variación no somática de la referencia (centro) y posibles falsos positivos (derecha). Cada mutación somática se puede asociar con una puntuación de calidad Q. El número de falsos positivos en la comparación de tumores indica un número de falsos positivos en la misma comparación versus la misma comparación. Por lo tanto, para una mutación dada con un puntaje de calidad Q detectado en la comparación de tumores, se estimó la tasa de descubrimiento falso calculando la relación de las mismas mutaciones frente a las mismas mutaciones con una puntuación de Q o mejor al número total de mutaciones encontradas en la comparación de tumores con una puntuación de Q o mejor.

Un desafío surge en la definición de Q ya que la mayoría de los marcos de detección de mutaciones calculan múltiples puntuaciones de calidad. Aquí, se aplica un clasificador de bosque aleatorio [Breiman, L., *Statist. Sci.* 16, 199-231 (2001)] para combinar múltiples puntajes en un solo puntaje de calidad P. Revisar la sección de métodos para obtener detalles sobre los detalles del puntaje de calidad y el cálculo de FDR.

Un sesgo potencial en la comparación de métodos es la cobertura diferencial; por lo tanto, se normaliza la tasa de descubrimiento falso para la cobertura:

$$FDR(Q) = \frac{\# \text{ de SNV iguales vs. iguales con puntuación } \geq Q}{\# \text{ comparación de tumor de cobertura común}}$$

SNV tumorales con puntuación $\geq Q$ # comparación de iguales vs iguales de cobertura común

Se calculó la cobertura común contando todas las bases del genoma de referencia que están cubiertas tanto por el tumor como por la muestra normal o por ambas muestras "iguales versus iguales", respectivamente.

5 Al estimar el número de falsos positivos y positivos en cada FDR (véase Métodos), se generaron curvas características operativas del receptor (ROC) y se calculó el AUC (área bajo la curva) para cada método de descubrimiento de mutaciones, permitiendo así una comparación de estrategias para el descubrimiento de mutaciones (Fig. 18B).

10 Además, la selección de los datos de referencia podría influir en el cálculo de las FDR. Usando los datos disponibles de black6/B16 es posible crear 18 tripletes (combinaciones de black6 vs. black6 y black6 vs. b16). Al comparar las distribuciones de FDR resultantes para los conjuntos de mutaciones somáticas, los resultados son consistentes (Fig. 18B).

15 Usando esta definición de una tasa de descubrimiento falso, se ha establecido un marco genérico para evaluar la influencia de numerosos parámetros experimentales y algorítmicos en el conjunto resultante de mutaciones somáticas. A continuación, aplicamos este marco para estudiar la influencia de las herramientas de software, la cobertura, la secuencia final pareada y el número de repeticiones técnicas en la identificación de mutaciones somáticas.

20 Primero, la elección de la herramienta de software tiene un claro impacto en las mutaciones somáticas identificadas (Fig. 19A). En los datos probados, SAMtools produce el mayor enriquecimiento de verdaderos positivos en un conjunto de mutaciones somáticas clasificadas por la FDR. Sin embargo, se observa que todas las herramientas ofrecen muchos parámetros y puntuaciones de calidad para las mutaciones individuales. En el presente documento, se ha utilizado la configuración predeterminada de acuerdo con lo especificado por los desarrolladores de algoritmos; se espera que los parámetros puedan optimizarse y se enfatiza que el marco de la FDR definido aquí está diseñado para ejecutar y evaluar dicha optimización.

25 Para el experimento de secuenciación de B16 descrito, se secuencia cada muestra en un carril de celda de flujo individual y se logra una cobertura de base media de la región objetivo de 38 veces para las muestras individuales. Sin embargo, esta cobertura podría no ser necesaria para obtener un conjunto igualmente bueno de mutaciones somáticas, posiblemente reduciendo los costos. Además, el impacto de la profundidad de la cobertura en la detección de las SNV del genoma completo se ha discutido recientemente [Ajay, S.S. et al., Genome Res. 21, 1498-1505 (2011)]. Para estudiar el efecto de la cobertura en los datos de captura del exón, se redujo el número de lecturas de secuencia alineadas para cada biblioteca de 1 x 50 nt para generar una cobertura aproximada de 5, 10 y 20 veces, respectivamente, y luego se volvieron a aplicar los algoritmos de llamada de mutación. Como se esperaba, una mayor cobertura resulta en un mejor conjunto de mutaciones somáticas (es decir, menos falsos positivos), aunque la mejora de la cobertura de 20 veces hasta el máximo es marginal (Fig. 19B).

35 Es sencillo simular y clasificar diferentes configuraciones experimentales utilizando los datos y el marco disponibles. Al comparar duplicados con triplicados, los triplicados no ofrecen un beneficio en comparación con los duplicados (Fig. 19C), mientras que los duplicados ofrecen una clara mejora en comparación con un estudio sin ninguna réplica. En términos de la proporción de mutaciones somáticas en los conjuntos dados, se observa un enriquecimiento en una FDR de 5% a partir del 24,2% para un procedimiento sin réplicas hasta 71,2% para duplicados y 85,8% para triplicados. A pesar del enriquecimiento, el uso de la intersección de triplicados elimina más mutaciones con una FDR baja que aquellos con una FDR alta, como lo indica el AUC de la ROC más baja y el desplazamiento de la curva hacia la izquierda (Fig. 19C): la especificidad aumenta ligeramente a costa de una menor sensibilidad.

40 La biblioteca adicionalmente secuenciada de 2 x 100 nt se usó para simular una biblioteca de 1 x 100, dos bibliotecas de 2 x 50 y dos de 1 x 50 nt, respectivamente, mediante la eliminación *in silico* de la segunda lectura y/o los extremos 3' y 5' de las lecturas, dando como resultado un total de 5 bibliotecas simuladas. Estas bibliotecas se compararon usando las FDR calculadas de mutaciones predichas (Fig. 19D). A pesar de la cobertura media mucho más alta (más de 77 frente a 38), las mutaciones somáticas encontradas usando las bibliotecas de 2 x 50 5' y 1 x 100 nt tienen un AUC de la ROC más baja y, por lo tanto, una distribución de FDR peor que la biblioteca de 1 x 50 nt. Este fenómeno resulta de la acumulación de mutaciones altas de FDR en regiones de baja cobertura, ya que los conjuntos de mutaciones bajas de FDR encontrados son muy similares. La consecuencia es que la longitud de secuenciación óptima es pequeña, de modo que las bases secuenciadas se concentran alrededor de las secuencias de la sonda de captura (sin embargo, posiblemente pierden información sobre el estado somático de las mutaciones en regiones no cubiertas) o deben estar cerca de la longitud del fragmento (2 x 100 nt = 200 nt de longitud total para ~250 nt fragmentos en nuestro caso), llenando efectivamente los vacíos de cobertura. Esto también es compatible con el AUC de la ROC de la biblioteca de 2 x 50 nt 3' (simulado usando solo los extremos 3' de la biblioteca de 2 x 100 nt) que es más alto que el de la biblioteca de 2 x 50 nt 5' (simulado usando solo los extremos 5' de la biblioteca de 2 x 100 nt) a pesar de la baja calidad base de los extremos de lectura de 3'.

55 Estas observaciones nos permiten definir procedimientos de mejores prácticas para el descubrimiento de mutaciones somáticas. En todos los parámetros evaluados, una cobertura de 20 veces en ambas muestras y el uso de un duplicado técnico logra resultados cercanos a los óptimos en estas muestras relativamente homogéneas, al tiempo que

considera los costos. Una biblioteca de 1x50 nt que genera aproximadamente 100 millones de lecturas parece ser la opción más pragmática para lograr esta cobertura. Esto sigue siendo cierto en todos los posibles emparejamientos de conjuntos de datos. Se aplican retrospectivamente esas configuraciones de parámetros, no se usa filtrado adicional de las llamadas de variantes sin procesar y se calcularon las FDR para 50 mutaciones seleccionadas de la intersección de los tres métodos como se muestra en la Figura 17. Todas las mutaciones fueron confirmadas por una combinación de resecuenciación de Sanger y las lecturas de la secuencia de ARN-Seq de B16. Cuarenta y cuatro de esas mutaciones se habrían encontrado utilizando un corte de FDR del 5% (Fig. 20). Como control negativo, se volvieron a secuenciar los loci de 44 mutaciones predichas con FDR altas (> 50%) y se examinaron las secuencias respectivas en los datos de Seq-ARN. Se encontró que 37 de estas mutaciones no estaban validadas, mientras que los restantes siete loci de mutaciones potenciales no estaban cubiertos por las lecturas de Seq-ARN y producían una reacción de no secuenciación.

Si bien se mostró la aplicación del marco a cuatro preguntas específicas, de ninguna manera se limita a estos parámetros, sino que puede aplicarse para estudiar la influencia de todos los parámetros experimentales o algorítmicos, por ejemplo, la influencia del software de alineación, la elección de una métrica de mutación o la elección del proveedor para la selección del exoma.

Se llevaron a cabo todos los experimentos en una serie de experimentos de células de melanoma B16; sin embargo, el método no está restringido a estos datos. El único requisito es la disponibilidad de un conjunto de datos de referencia 'igual frente a igual', lo que significa que se debe realizar al menos una sola repetición técnica de una muestra no tumoral para cada nuevo protocolo. Si bien estos experimentos indican que el método es robusto con respecto a la elección de la repetición técnica dentro de ciertos límites, por lo que no se requiere necesariamente una repetición en cada experimento. Sin embargo, el método requiere que las diversas medidas de calidad sean comparables entre el conjunto de datos de referencia y los conjuntos de datos restantes.

Dentro de esta contribución, hemos sido pioneros en un marco estadístico para una detección de mutaciones somáticas impulsada por la tasa de descubrimiento falso. Este marco no solo es aplicable para la selección de objetivos de diagnóstico o terapéuticos, sino que también permite una comparación genérica de las etapas de protocolo experimental y computacional en datos de verdad cuasi fundamentados generados. En este documento, se aplicó esta idea para tomar decisiones de protocolo con respecto a las herramientas de software, la cobertura, las réplicas y la secuencia final emparejada.

Métodos

Captura y secuenciación de bibliotecas

Secuenciación de próxima generación, secuenciación del ADN: la captura del exoma para la secuenciación del ADN se realizó usando el ensayo de captura basado en la solución Sure-Select de Agilent [Gnirke, A., et al., Nat. Biotechnol. 27, 182-189 (2009)], en este caso diseñado para capturar todos los exones de ratón conocidos. Se fragmentaron 3 µg de ADN genómico purificado hasta 150-200 nt usando un dispositivo de ultrasonido Covaris S2. Los fragmentos de ADN_g se repararon finalmente usando ADN polimerasa T4, ADN polimerasa Klenow y se fosforilaron en 5' usando polinucleótido quinasa T4. Los fragmentos de ADN_g de extremos romos se adenilaron en 3' usando el fragmento de Klenow (3' a 5' exo menos). Los adaptadores finalmente emparejados del extremo 3' con salientes T individuales Illumina se ligaron a los fragmentos de ADN_g usando una relación molar 10:1 de adaptador a inserto de ADN genómico usando ADN ligasa T4. Los fragmentos de ADN_g ligados al adaptador se enriquecieron antes de la captura y se agregaron secuencias específicas de células de flujo usando los cebadores de PCR Illumina PE 1.0 y 2.0 y la polimerasa Herculasa II (Agilent) usando 4 ciclos de PCR.

Se hibridaron 500 ng de fragmentos de ADN_g enriquecidos por PCR ligados al adaptador con cebos de biblioteca de ARN de exoma completo de ratón biotinilado SureSelect de Agilent durante 24 horas a 65 °C. Los complejos de cebo hibridados de ADN_g/ARN se eliminaron usando perlas magnéticas recubiertas con estreptavidina. Los complejos de cebo de ADN_g/ARN se lavaron y los cebos de ARN se escindieron durante la elución en regulador de elución SureSelect, dejando los fragmentos de ADN_g enriquecidos por PCR ligados al adaptador capturado. Los fragmentos de ADN_g se amplificaron por PCR después de la captura usando ADN polimerasa Herculasa II (Agilent) y cebadores de PCR GA SureSelect durante 10 ciclos.

Las limpiezas se realizaron usando un volumen 1,8x de perlas magnéticas AMPure XP (Agencourt). Para los controles de calidad, se utilizó el ensayo Qubit HS de Invitrogen y el tamaño del fragmento se determinó usando el ensayo HS de ADN con el 2100 Bioanalyzer de Agilent.

Las bibliotecas de ADN_g enriquecidas en el exoma se agruparon en el cBot usando el kit de agrupamiento Truseq SR v2.5 usando 7 pM y secuenciadas en el Illumina HiSeq2000 usando el kit Truseq SBS.

Análisis de datos del exoma

Las lecturas de secuencia se alinearon usando bwa (versión 0.5.8c) [Li, H. Durbin, R., Bioinformatics 25, 1754-1760 (2009)] usando opciones predeterminadas para el ensamblaje del genoma de ratón de referencia mm9 [Mouse Genome Sequencing Consortium, Nature 420, 520-562 (2002)]. Se eliminaron las lecturas ambiguas, esas lecturas

que se asignan a múltiples ubicaciones del genoma de acuerdo con lo provisto por la salida de bwa. Las alineaciones restantes se clasificaron, indexaron y convirtieron a un formato binario y comprimido (BAM) y los puntajes de calidad de lectura se convirtieron del puntaje de calidad estándar Illumina phred+64 a puntajes de calidad estándar de Sanger utilizando programas de instrucciones de algoritmos.

5 Para cada carril de secuenciación, se identificaron mutaciones usando tres programas de software: acumulación de SAMtools (versión 0.1.8) [Li, H. et al., *Bioinformatics* 25, 2078-2079 (2009)], GATK (versión 1.0.4418) [DePristo, MA et al., *Nature Genetics* 43, 491-498 (2011)], y SomaticSniper [Ding, L. et al., *Hum. Mol. Genet.* (2010) publicado por primera vez en línea el 15 de septiembre de 2010]. Para SAMtools, se utilizaron las opciones recomendadas por el autor y los criterios de filtro (http://sourceforge.net/apps/mediawiki/SAMtools/index.php?title=SAM_FAQ; consultado en septiembre de 2011), incluido el filtrado de primera ronda, cobertura máxima 200. Para el filtrado de segunda ronda de SAMtools, el puntaje mínimo de calidad indel fue de 50, la calidad mínima de mutación puntual fue de 30. Para las llamadas de mutación GATK, se siguieron las pautas de mejores prácticas diseñadas por el autor que se presentan en el manual del usuario de GATK (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit; consultado en octubre de 2010). Para cada muestra, se realizó un realineamiento local alrededor de los sitios indel seguido de una recalibración de calidad base. El módulo UnifiedGenotyper se aplicó a los archivos de datos de alineación resultantes. Cuando fue necesario, los polimorfismos conocidos de dbSNP [Sherry, S.T. et al., *Nucleic Acids Res.* 29, 308-311 (2009)] (versión 128 para mm9) se suministraron a las etapas individuales. La etapa de recalibración de puntuación variante se omitió y se reemplazó por la opción de filtrado duro. Para las llamadas de mutación SomaticSniper, se utilizaron las opciones predeterminadas y solo se consideraron adicionalmente las mutaciones predichas con una "puntuación somática" de 30 o más. Además, para cada locus potencialmente mutado, se requirió una cobertura distinta de cero en el tejido normal y se eliminaron todas las mutaciones ubicadas en secuencias repetitivas de acuerdo con lo definido por la pista RepeatMasker del navegador de genoma UCSC para el ensamblaje del genoma del ratón mm9 [Fujita, P.A. et al., *Nucleic Acids Res.* 39, 876-882 (2011)].

Seq-ARN

25 Las bibliotecas de Seq-ARNmADNc con código de barras se prepararon a partir de 5 µg de ARN total usando una versión modificada del protocolo de Seq-ARNm de Illumina. El ARNm se aisló usando perlas magnéticas SeramagOligo (dT) (Thermo Scientific). El ARNm aislado se fragmentó usando cationes divalentes y calor, dando como resultado fragmentos que oscilan entre 160 y 200 pb. El ARNm fragmentado se convirtió en ADNc usando cebadores aleatorios y SuperScriptII (Invitrogen) seguido de la síntesis de la segunda cadena usando ADN polimerasa I y RNasaH. El ADNc se reparó con ADN polimerasa T4, ADN polimerasa Klenow y se fosforiló en 5' con polinucleótido quinasa T4. Los fragmentos de ADNc con extremos romos se adenilaron en 3' usando el fragmento de Klenow (3' a 5' exo menos). Se ligaron adaptadores específicos multiplex del extremo 3' emparejados con salientes T individuales Illumina en los fragmentos de ADNc usando ADN ligasa T4. Las bibliotecas de ADNc se purificaron y el tamaño se seleccionó a 300 pb usando el gel E-Gel SizeSelect al 2% (Invitrogen). El enriquecimiento, la adición del índice se seis bases de Illumina y secuencias específicas de células de flujo se realizó por PCR usando ADN polimerasa Phusion (Finnzymes). Todas las limpiezas se realizaron con un volumen 1,8x de perlas magnéticas Agencourt AMPure XP.

Las bibliotecas de Seq-ARN con código de barras se agruparon en el cBot usando el kit de clúster Truseq SR v2.5 usando 7 pM y secuenciadas en el Illumina HiSeq2000 usando el kit Truseq SBS.

40 Los datos de salida sin procesar del HiSeq se procesaron de acuerdo con el protocolo estándar Illumina, incluida la eliminación de lecturas de baja calidad y desmultiplexación. Las lecturas de secuencia se alinearon luego con la secuencia del genoma de referencia [Mouse Genome Sequencing Consortium, *Nature* 420, 520-562 (2002)] usando Bowtie [Langmead, B. et al., *Genome Biol.* 10, R25 (2009)]. Las coordenadas de alineación se compararon con las coordenadas de exón de los transcritos de RefSeq [Pruitt, K.D. et al., *Nucleic Acids Res.* 33, 501-504 (2005)] y para cada transcrito se registraron los recuentos de alineamientos superpuestos. Las lecturas de secuencia que no se alinean con la secuencia genómica se alinearon con una base de datos de todas las posibles secuencias de unión exón-exón de los transcritos de RefSeq [Pruitt, K.D. et al., *Nucleic Acids Res.* 33, 501-504 (2005)]. Las coordenadas de alineación se compararon con las coordenadas de unión y exón de RefSeq, se contaron las lecturas y se normalizaron a RPKM (número de lecturas que mapean por kilobase de nucleótidos de transcrito por millón de lecturas mapeadas [Mortazavi, A. et al., *Nat. Methods* 5, 621- 628 (2008)]) para cada transcrito.

50 Validación de SNV

Se seleccionaron SNV para la validación por resecuenciación de Sanger y ARN, se identificaron las SNV que fueron predichas por los tres programas, no sinónimos, y se encontraron en transcritos que tienen un mínimo de 10 RPKM. De estos, se seleccionaron los 50 con los puntajes de calidad de SNP más altos de acuerdo con lo previsto por los programas. Como control negativo, se seleccionaron 44 SNV que tienen una FDR del 50% o más, están presentes en una sola muestra de línea celular y se predicen por un solo programa de llamada de mutación. Usando ADN, las variantes seleccionadas se validaron mediante amplificación por PCR de las regiones usando 50 ng de ADN, seguido por secuenciación de Sanger (Eurofins MWG Operon, Ebersberg, Alemania). Las reacciones fueron exitosas para 50 y 32 loci de controles positivos y negativos, respectivamente. La validación también se realizó mediante el examen de las lecturas tumorales de Seq-ARN.

Cálculo de FDR y aprendizaje automático

5 Cálculo del puntaje de la calidad aleatorio del bosque: algoritmos de llamada de mutación comúnmente utilizados (DePristo, MA et al., Nature Genetics 43, 491-498 (2011), Li, H. et al., Bioinformatics 25, 2078-2079 (2009), Ding, L. et al., Hum. Mol. Genet (2010), publicado por primera vez en línea el 15 de septiembre de 2010) arrojaron múltiples
 10 puntajes, todos los cuales son potencialmente influyentes para la calidad de la llamada de mutación. Estos incluyen, entre otros, la calidad de la base de interés asignada por el instrumento, la alineación de calidad para esta posición, el número de lecturas que cubren esta posición o una puntuación para la diferencia entre los dos genomas en comparación con esta posición. Para el cálculo de la tasa de descubrimiento falso, se requiere un orden de las mutaciones, sin embargo, esto no es directamente factible para todas las mutaciones, ya que se podría tener información contradictoria de los diversos puntajes de calidad.

Se utilizó la siguiente estrategia para lograr una clasificación completa. En una primera etapa, se aplicó una definición muy rigurosa de superioridad al suponer que una mutación tiene mejor calidad que otra si y solo si es superior en todas las categorías. Entonces un conjunto de propiedades de calidad $S = (s_1, \dots, s_n)$ es preferible a $T = (t_1, \dots, t_n)$, denotado por $S > T$, si $s_i > t_i$ para todo $i = 1, \dots, n$. Se define una FDR intermedia (IFDR) de la siguiente manera

15
$$\text{IFDR}(Q) = \frac{\# \text{ de SNV iguales vs. iguales con puntuación } > T}{\# \text{ SNV tumorales con puntuación } S > T} \times \frac{\# \text{ comparación de tumor de cobertura común}}{\# \text{ comparación de iguales vs iguales de cobertura común}}$$

20 Sin embargo, se considera la IFDR solo como una etapa intermedia, ya que en muchos casos estrechamente relacionados, no es posible realizar ninguna comparación y, por lo tanto, no se obtiene un beneficio de la gran cantidad de datos disponibles. Por lo tanto, se aprovecha la buena propiedad de generalización de la regresión forestal aleatoria [Breiman, L., Statist. Sci. 16, 199-231 (2001)] y entrenar un bosque aleatorio como se implementa en R (R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Viena, Austria, 2010, Liaw, A., Wiener, M., R News 2, 18-22 (2002)).

25 Para m mutaciones de entrada con n propiedades de calidad cada una, se determinó el intervalo de valores para cada propiedad y se muestrearon hasta p valores con un espaciado uniforme fuera de este intervalo; cuando el conjunto de valores para una propiedad de calidad era menor que p , este conjunto se usaba en lugar del conjunto muestreado. Luego se crea cada combinación posible de valores de calidad muestreados o seleccionados, lo que da como resultado un máximo de p^n puntos de datos en el espacio de calidad n dimensional. Se usó una muestra aleatoria del 1% de estos puntos y los valores IFDR correspondientes como predictor y respuesta, respectivamente, para el entrenamiento forestal aleatorio.

30 La puntuación de regresión resultante es la puntuación de calidad generalizada Q ; se puede considerar como una combinación ponderada localmente de los puntajes de calidad individuales. Permite la comparación directa de un solo valor de cualquiera de las dos mutaciones y el cálculo de la tasa real de descubrimiento falso:

$$\text{FDR}(Q) = \frac{\# \text{ de SNV iguales vs. iguales con puntuación } \geq Q}{\# \text{ SNV tumorales con puntuación } \geq Q} \times \frac{\# \text{ comparación de tumor de cobertura común}}{\# \text{ comparación de iguales vs iguales de cobertura común}}$$

35 Para el entrenamiento del modelo de bosque aleatorio utilizado para crear los resultados para este estudio, se calculó la IFDR de la muestra sobre las mutaciones somáticas de todas las muestras antes de seleccionar el subconjunto aleatorio del 1%. Esto garantiza la asignación de todo el espacio de calidad disponible a los valores de FDR. Se utilizan las propiedades de calidad "calidad SNP", "profundidad de cobertura", "calidad de consenso" y "calidad de mapeo RMS" (SAMtools, $p = 20$); "calidad SNP", "profundidad de cobertura", "confianza variante/profundidad sin filtrar" y "calidad de mapeo RMS" (GATK, $p = 20$); o calidad SNP, "profundidad de cobertura", "calidad de consenso", "calidad de mapeo RMS" y "puntaje somático" (SomaticSNiPer, $p = 12$), respectivamente. Los diferentes valores de p aseguran un tamaño establecido de magnitud comparable.

45 Cálculo de cobertura común: el número de posibles llamadas de mutación puede introducir un sesgo importante en la definición de una tasa de descubrimiento falsa. Solo si tenemos el mismo número de ubicaciones posibles para que ocurran mutaciones para nuestra comparación de tumores y para nuestra comparación de igual frente a igual, el número de mutaciones llamadas es comparable y puede servir como base para un cálculo falso de la tasa de descubrimiento. Para corregir este sesgo potencial, se utiliza el índice de cobertura común. Como cobertura común, se define el número de bases con cobertura de al menos una en ambas muestras que se utilizan para la llamada de mutación. Se calcula la cobertura común individualmente para la comparación de tumores, así como para la
 50 comparación de igual frente a igual.

Estimación de la ROC

55 Las curvas características de funcionamiento del receptor (ROC) y el área correspondiente bajo la curva (AUC) son útiles para organizar clasificadores y visualizar su rendimiento [Fawcett, T., Pattern Recognition. Lett. 27, 861-874 (2006)]. Se extiende este concepto para evaluar el desempeño de los procedimientos experimentales y computacionales. Sin embargo, el trazado de gráficos de ROC requiere el conocimiento de todos los ejemplos de

verdadero y falso positivo (TP y FP) en un conjunto de datos, información que generalmente no se proporciona y es difícil de establecer para datos de alto rendimiento (tal como los datos de NGS). Por lo tanto, se utilizaron las FDR calculadas para estimar las tasas respectivas de TP y FP y trazar un gráfico de ROC y calcular un AUC. La idea central es que la FDR de una sola mutación en el conjunto de datos suministre la proporción de cuánto contribuye esta mutación a la suma de mutaciones TP/FP, respectivamente. Además, para una lista de asignaciones aleatorias para TP y FP, el AUC de la ROC resultante será igual a 0,5 con este método, lo que indica una predicción completamente aleatoria. Se comienza con dos condiciones:

$$FDR = \frac{FPR}{FPR + TPR} \quad [1]$$

y

$$FPR + TPR = 1 \quad [2]$$

siendo FPR y TPR las relaciones positivas falsas positivas verdaderas necesarias, respectivamente, para la mutación dada, definiendo el punto correspondiente en el espacio de ROC. [1] y [2] se pueden reorganizar en

$$TPR = 1 - FPR \quad [3]$$

y

$$FPR = FDR \quad [4]$$

Para obtener una curva ROC estimada, las mutaciones en el conjunto de datos se ordenan por FDR y para cada mutación se traza un punto en los valores acumulativos de TPR y FPR hasta esta mutación, dividido por la suma de todos los valores de TPR y TPR, respectivamente. El AUC se calcula sumando las áreas de todos los trapecios consecutivos entre la curva y el eje x.

20 Listado de secuencias

<110> BioNTech AG y col.

<120> VACUNAS INDIVIDUALIZADAS PARA EL CÁNCER

<130> 674-75PCT

<150> PCT/EP2011/002576

25 <151> 2011-05-24

<150> PCT/EP2012/000006

<151> 2012-01-02

<160> 39

<170> PatentIn versión 3.5

30 <210> 1

<211> 15

<212> PRT

<213> Secuencia artificial

<220>

35 <223> Secuencia del enlazador

<220>

<221> REPETICIÓN

<222> (1) .. (3)

40 <223> Porción de secuencia repetida a veces, en la que a es independientemente un número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20

ES 2 746 233 T3

<220>
 <221> REPETICIÓN
 <222> (4) .. (6)
 5 <223> Porción de secuencia repetida b veces, en la que b es independientemente un número seleccionado de 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> (7) .. (9)
 10 <223> Porción de secuencia repetida c veces, en la que c es independientemente un número seleccionado de 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> (10) .. (12)
 15 <223> Porción de secuencia repetida d veces, en la que d es independientemente un número seleccionado de 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)
 20 <223> Porción de secuencia repetida e veces, en la que e es independientemente un número seleccionado de 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 o 20
 <220>
 <221> CARACTERÍSTICA NUEVA
 <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
 25 <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
 30 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia del enlazador
 <400> 2

35

gly gly ser gly gly gly gly ser gly
 1 5

<210> 3
 <211> 8
 <212> PRT
 40 <213> Secuencia artificial
 <220>
 <223> Secuencia del epítipo

ES 2 746 233 T3

<400> 3

Ser Ile Ile Asn Phe Glu Lys Leu
1 5

5 <210> 4
<211> 30
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia de péptidos mutada

10 <400> 4

Ser Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Asn Val
1 5 10 15

Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu Pro Ser
20 25 30

15 <210> 5
<211> 30
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia de péptidos de tipo silvestre

<400> 5

20

Ser Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Lys Val
1 5 10 15

Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu Pro Ser
20 25 30

25 <210> 6
<211> 27
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia peptídica mutada

30 <400> 6

Thr Pro Pro Pro Glu Glu Ala Met Pro Phe Glu Phe Asn Gly Pro Ala
1 5 10 15

Gln Gly Asp His Ser Gln Pro Pro Leu Gln Val
20 25

35 <210> 7
<211> 27
<212> PRT
<213> Secuencia artificial

<220>
<223> secuencia de péptidos de tipo silvestre

<400> 7

ES 2 746 233 T3

Thr Pro Pro Pro Glu Glu Ala Met Pro Phe Glu Phe Asn Glu Pro Ala
1 5 10 15

Gln Gly Asp His Ser Gln Pro Pro Leu Gln Val
20 25

5 <210> 8
<211> 27
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia peptídica mutada

<400> 8

10

Arg Val Thr Cys Asn Arg Ala Gly Glu Lys His Cys Phe Ser Ser Asn
1 5 10 15

Glu Ala Ala Arg Asp Phe Gly Gly Ala Ile Gln
20 25

15 <210> 9
<211> 27
<212> PRT
<213> secuencia de péptidos de tipo silvestre

<400> 9

Arg Val Thr Cys Asn Arg Ala Gly Glu Lys His Cys Phe Thr Ser Asn
1 5 10 15

Glu Ala Ala Arg Asp Phe Gly Gly Ala Ile Gln
20 25

20 <210> 10
<211> 27
<212> PRT
<213> Secuencia artificial

25 <220>
<223> Secuencia peptídica mutada

<400> 10

Phe Arg Arg Lys Ala Phe Leu His Trp Tyr Thr Gly Glu Ala Met Asp
1 5 10 15

Glu Met Glu Phe Thr Glu Ala Glu Ser Asn Met
20 25

30 <210> 11
<211> 27
<212> PRT
<213> Secuencia artificial

35 <220>
<223> Secuencia de péptidos de tipo silvestre

<400> 11

ES 2 746 233 T3

Phe Arg Arg Lys Ala Phe Leu His Trp Tyr Thr Gly Glu Gly Met Asp
 1 5 10 15

Glu Met Glu Phe Thr Glu Ala Glu Ser Asn Met
 20 25

- <210> 12
- <211> 27
- 5 <212> PRT
- <213> Secuencia artificial
- <220>
- <223> Secuencia peptídica mutada
- <400> 12

Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Asn Val Ser
 1 5 10 15

Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu
 20 25

- <210> 13
- <211> 27
- 15 <212> PRT
- <213> Secuencia artificial
- <220>
- <223> secuencia de péptidos tipo silvestre
- <400> 13

Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Lys Val Ser
 1 5 10 15

Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu
 20 25

- <210> 14
- <211> 27
- <212> PRT
- <213> Secuencia artificial
- 25 <220>
- <223> Secuencia peptídica mutada
- <400> 14

His Leu Thr Gln Gln Leu Asp Thr Tyr Ile Leu Lys Asn Val Val Ala
 1 5 10 15

Phe Ser Arg Thr Asp Lys Tyr Arg Gln Leu Pro
 20 25

- <210> 15
- <211> 27
- <212> PRT
- <213> Secuencia artificial
- <220>
- 35 <223> secuencia de péptidos de tipo silvestre
- <400> 15

ES 2 746 233 T3

His Leu Thr Gln Gln Leu Asp Thr Tyr Ile Leu Lys Asn Phe Val Ala
1 5 10 15

Phe Ser Arg Thr Asp Lys Tyr Arg Gln Leu Pro
20 25

5 <210> 16
<211> 27
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia peptídica mutada

<400> 16

10 Cys Gly Thr Ala Phe Phe Ile Asn Phe Ile Ala Ile Tyr His His Ala
1 5 10 15

Ser Arg Ala Ile Pro Phe Gly Thr Met Val Ala
20 25

15 <210> 17
<211> 27
<212> PRT
<213> Secuencia artificial

<220>
<223> secuencia de péptidos de tipo silvestre

<400> 17

Cys Gly Thr Ala Phe Phe Ile Asn Phe Ile Ala Ile Tyr Tyr His Ala
1 5 10 15

Ser Arg Ala Ile Pro Phe Gly Thr Met Val Ala
20 25

20 <210> 18
<211> 7
<212> PRT
<213> Secuencia artificial

25 <220>
<223> Secuencia del enlazador

<400> 18

Gly Gly Ser Gly Gly Gly Gly
1 5

30 <210> 19
<211> 10
<212> PRT
<213> Secuencia artificial

35 <220>
<223> Secuencia del enlazador

<400> 19

Thr Ser Leu Asn Ala Leu Leu Asn Ala His
1 5 10

ES 2 746 233 T3

<210> 20
 <211> 10
 <212> PRT
 <213> Secuencia artificial
 5 <220>
 <223> Secuencia del enlazador
 <400> 20
 Ser
 1 5 10
 10 <210> 21
 <211> 10
 <212> PRT
 <213> Secuencia artificial
 <220>
 15 <223> Secuencia del enlazador
 <400> 21
 Gly
 1 5 10
 20 <210> 22
 <211> 10
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia del enlazador
 25 <400> 22
 Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
 1 5 10
 30 <210> 23
 <211> 10
 <212> PRT
 <213> Secuencia artificial
 <220>
 <223> Secuencia del enlazador
 <400> 23
 35 Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly
 1 5 10
 <210> 24
 <211> 9
 <212> PRT
 40 <213> Secuencia artificial
 <220>
 <223> Secuencia del enlazador
 <400> 24
 Thr Ser Leu Asn Ala Leu Leu Asn Ala
 1 5

ES 2 746 233 T3

<210> 25
<211> 9
<212> PRT
<213> Secuencia artificial

5 <220>
<223> Secuencia del enlazador

<400> 25

10 Ser Ser Ser Ser Ser Ser Ser Ser Ser
1 5

<210> 26
<211> 9
<212> PRT
<213> Secuencia artificial

15 <220>
<223> Secuencia del enlazador

<400> 26

Gly Gly Gly Gly Gly Gly Gly Gly Gly -
1 5

20 <210> 27
<211> 9
<212> PRT
<213> Secuencia artificial

<220>
25 <223> Secuencia del enlazador

<400> 27

Gly Gly ser Gly Gly Gly ser Gly Gly
1 5

30 <210> 28
<211> 5
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia del enlazador

35 <400> 28

Gly Gly ser Gly Gly
1 5

40 <210> 29
<211> 6
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia del enlazador

<400> 29

ES 2 746 233 T3

Gly Gly ser Gly Gly Gly
1 5

5 <210> 30
<211> 7
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia del enlazador

<400> 30

10

Gly Gly ser Gly Gly Gly Gly
1 5

15 <210> 31
<211> 8
<212> PRT
<213> Secuencia artificial

<220>
<223> Secuencia del enlazador

<400> 31

20

Gly Gly ser Gly Gly Gly Gly ser
1 5

<210> 32
<211> 10
<212> PRT
<213> Secuencia artificial

25 <220>
<223> Secuencia del enlazador

<400> 32

30

Gly Gly ser Gly Gly Gly ser Gly Gly ser
1 5 10

<210> 33
<211> 14
<212> ADN
<213> Secuencia artificial

<220>
35 <223> oligonucleótido

<400> 33

ggaaactttt tccc 14

40

<210> 34
<211> 21
<212> ADN
<213> Secuencia artificial

<220>
<223> oligonucleótido

45

<400> 34
ccgggaaac ttttccag t 21

ES 2 746 233 T3

<210> 35
 <211> 21
 <212> ADN
 <213> Secuencia artificial

5 <220>
 <223> oligonucleótido

<400> 35
 ccggggaaac gtttccag t 21

10 <210> 36
 <211> 60
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Constructo artificial

15 <400> 36

Asn Thr Thr Phe Asn Val Ser Glu Glu Ser Pro Ser Lys Pro Ser Phe
 1 5 10 15
 Gln Glu Phe Val Asp Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr
 20 25 30
 Asp Gln Pro Phe Leu Pro Ser Ala Pro Val Phe Ile Phe Thr Lys Gly
 35 40 45
 Arg Lys Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
 50 55 60

20 <210> 37
 <211> 60
 <212> PRT
 <213> Secuencia artificial

<220>
 25 <223> Constructo artificial

<400> 37

Tyr Phe Val Leu Tyr Lys Pro Pro Pro Lys Asp Asn Ile Pro Ala Leu
 1 5 10 15
 Val Glu Glu Tyr Leu Glu Arg Gly Asn Phe Val Ala Asn Asp Leu Asp
 20 25 30
 Trp Leu Leu Ala Leu Pro His Asp Lys Phe Trp Cys Gln Val Ile Phe
 35 40 45
 Asp Glu Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
 50 55 60

30 <210> 38
 <211> 60
 <212> PRT
 <213> Secuencia artificial

<220>
 <223> Constructo artificial

35 <400> 38

ES 2 746 233 T3

Phe Glu Gln Leu Ser Glu Ser Ala Lys Glu Glu Leu Ile Asn Phe Lys
 1 5 10 15

Arg Lys Arg Val Ala Ala Phe Gln Lys Asn Leu Ile Glu Met Ser Glu
 20 25 30

Leu Glu Ile Lys His Ala Arg Asn Asn Val Ser Leu Leu Gln Ser Cys
 35 40 45

Ile Asp Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
 50 55 60

<210> 39

<211> 60

5 <212> PRT

<213> Secuencia artificial

<220>

<223> Constructo artificial

<400> 39

10

Thr Gly Ala Ser Pro Gly Leu Gly Ala Tyr Thr Pro Pro Pro Glu Glu
 1 5 10 15

Ala Met Pro Phe Glu Phe Asn Gly Pro Ala Gln Gly Asp His Ser Gln
 20 25 30

Pro Pro Leu Gln Val Pro Asp Leu Ala Pro Gly Gly Pro Glu Ala Leu
 35 40 45

Val Pro Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
 50 55 60

REIVINDICACIONES

1. Una vacuna individualizada contra el cáncer para usar en un método de tratamiento de un paciente con cáncer, cuyo método comprende las etapas:
- (a) proporcionar dicha vacuna individualizada contra el cáncer mediante un método que comprende las etapas:
- 5 (aa) identificar mutaciones somáticas específicas de cáncer en una muestra de tumor del paciente con cáncer para proporcionar una firma de mutación de cáncer del paciente; y
- (ab) proporcionar una vacuna de ARN con la firma de mutación de cáncer obtenida en la etapa (aa), en la que la vacuna de ARN con la firma de mutación del paciente comprende ARN que codifica un polipéptido poliepitópico recombinante que comprende neoepítomos basados en mutación; y
- 10 (b) administrar dicha vacuna individualizada contra el cáncer al paciente.
2. La vacuna individualizada contra el cáncer para uso de acuerdo con la reivindicación 1, en la que la etapa de identificar mutaciones somáticas específicas de cáncer comprende identificar la firma de mutación del cáncer del exoma de una o más células cancerosas.
3. La vacuna individualizada contra el cáncer para su uso de acuerdo con la reivindicación 1 o 2, en la que la etapa de identificar mutaciones somáticas específicas del cáncer comprende la secuenciación de células individuales de una o más células cancerosas.
- 15 4. La vacuna individualizada contra el cáncer para uso de acuerdo con la reivindicación 3, en la que las células cancerosas son células tumorales en circulación.
5. La vacuna individualizada contra el cáncer para su uso de acuerdo con una cualquiera de las reivindicaciones 1 a 4, en la que la etapa de identificar mutaciones somáticas específicas del cáncer implica el uso de secuenciación de próxima generación (NGS).
- 20 6. La vacuna individualizada contra el cáncer para su uso de acuerdo con una cualquiera de las reivindicaciones 1 a 5, en la que la etapa de identificar mutaciones somáticas específicas de cáncer comprende secuenciar el ADN y/o ARN genómico de la muestra de tumor.
- 25 7. La vacuna individualizada contra el cáncer para su uso de acuerdo con una cualquiera de las reivindicaciones 1 a 6, en la que el polipéptido poliepitópico recombinante comprende hasta 30 neoepítomos basados en mutaciones.
8. La vacuna individualizada contra el cáncer para uso de acuerdo con la reivindicación 7, en la que el polipéptido poliepitópico recombinante comprende además epítomos que no contienen mutaciones somáticas específicas del cáncer que se expresan por células cancerosas.
- 30 9. La vacuna individualizada contra el cáncer para uso de acuerdo con la reivindicación 7 u 8, en la que los epítomos están en su contexto de secuencia natural para formar una secuencia de vacuna, en la que la secuencia de vacuna comprende preferiblemente 20 o más aminoácidos.
10. La vacuna individualizada contra el cáncer para su uso de acuerdo con una cualquiera de las reivindicaciones 1 a 9, en la que los neoepítomos, los epítomos y/o las secuencias de vacuna están alineados de un extremo al otro y/o están separados por enlazadores.
- 35

Fig. 1

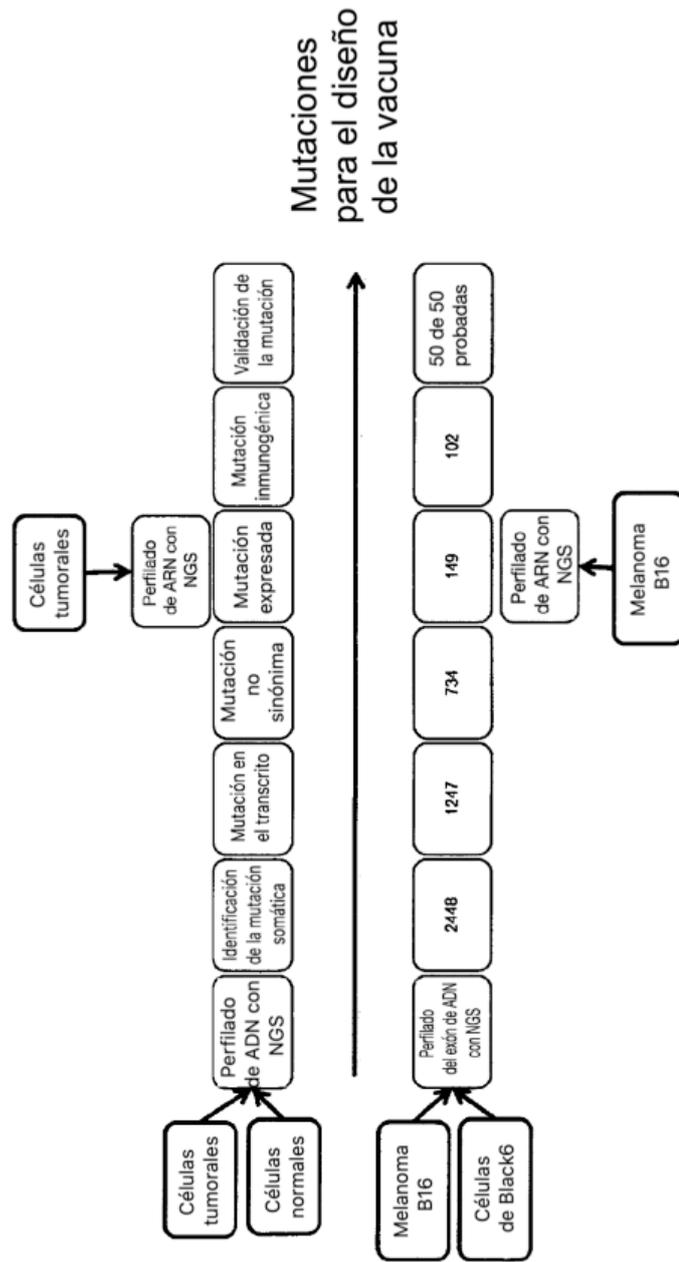
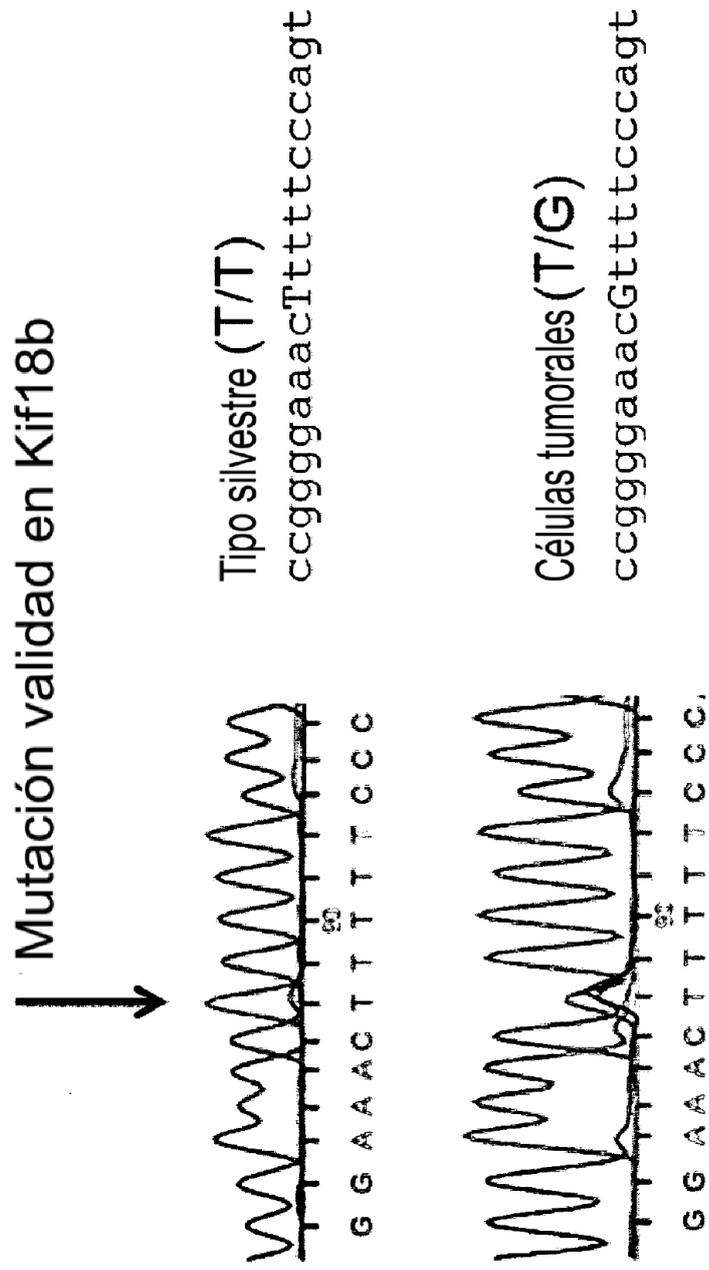


Fig. 2



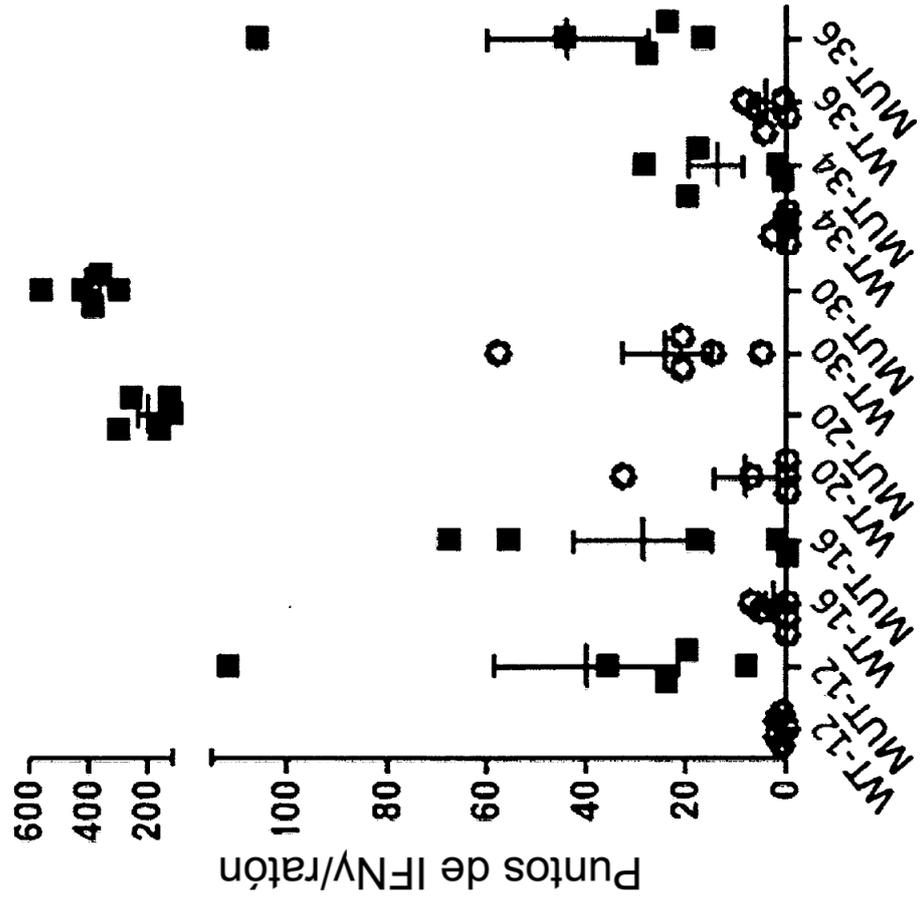


Fig. 3

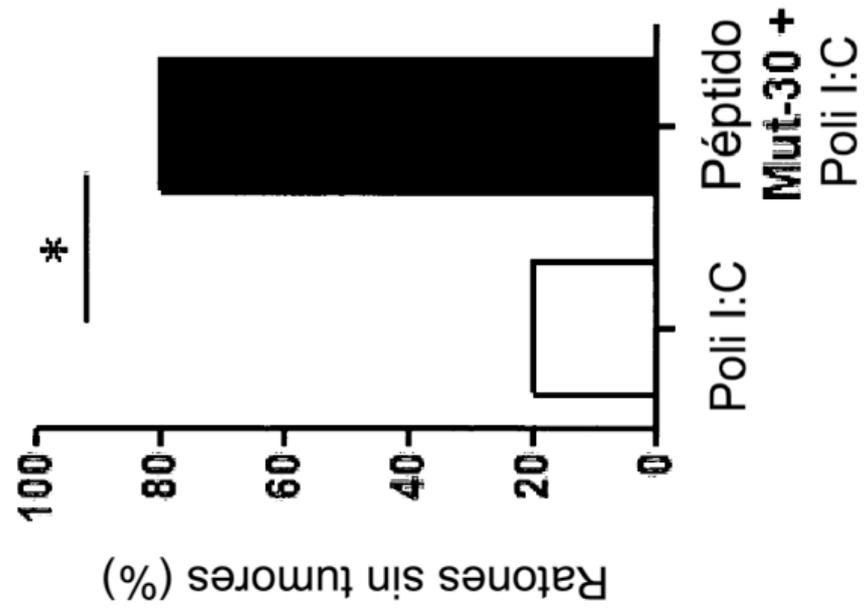


Fig. 4

Fig. 5

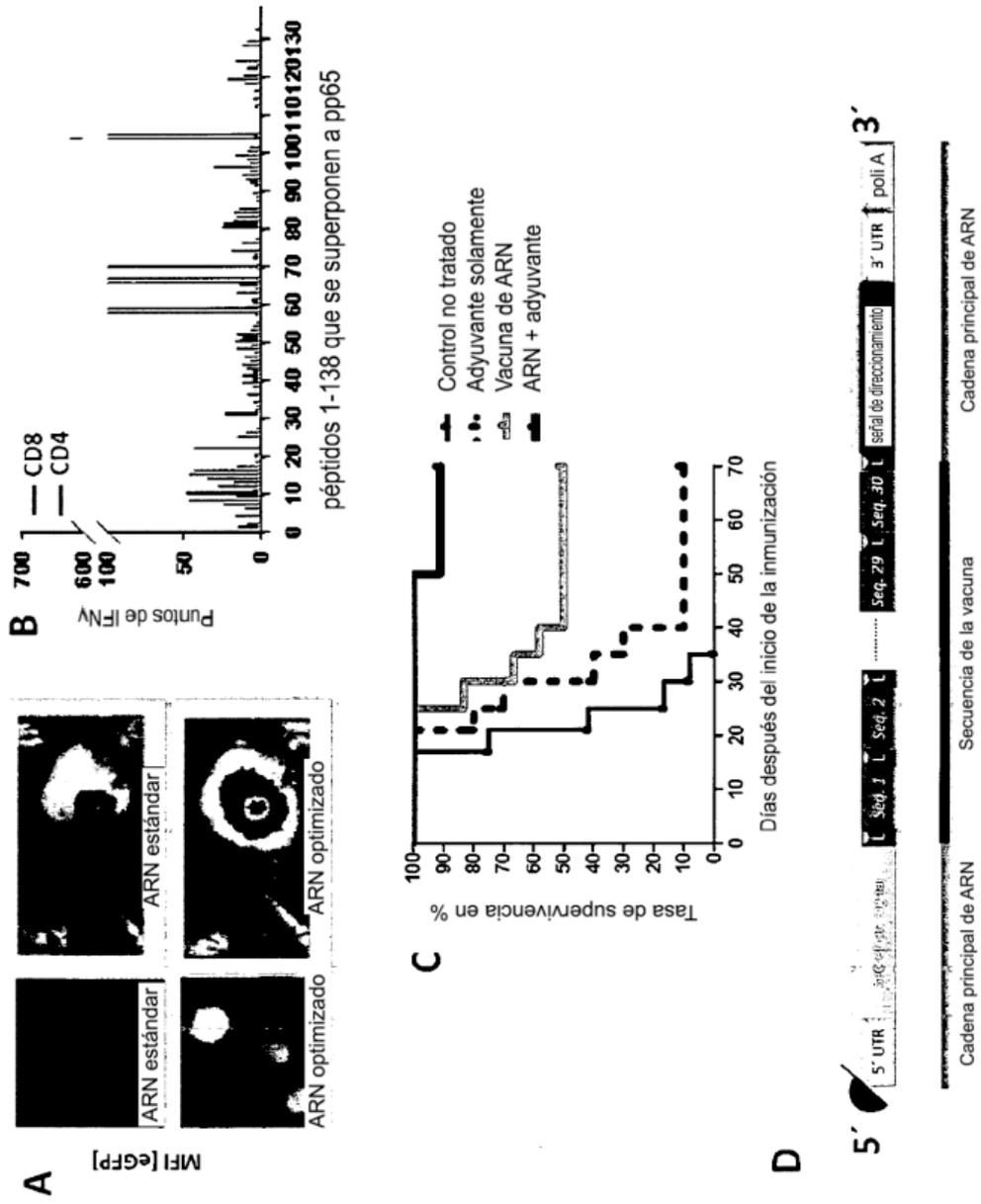


Fig. 6

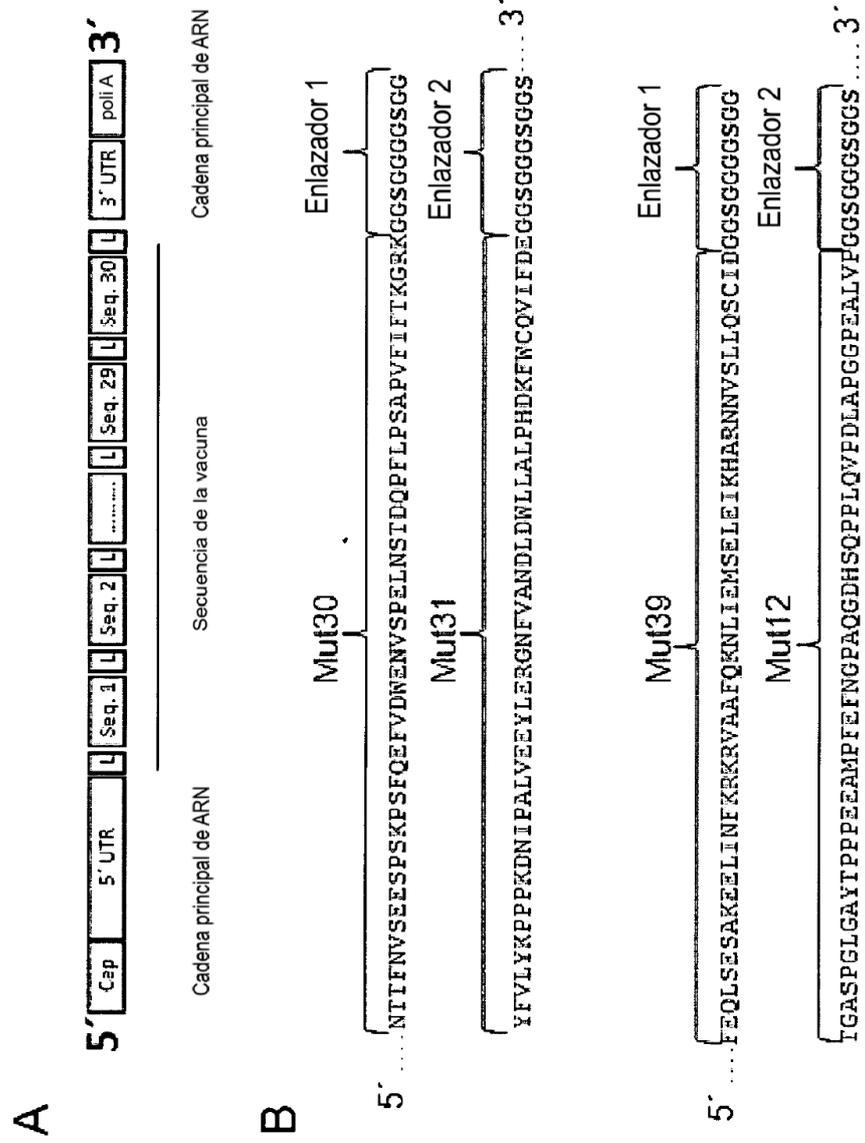


Fig. 7

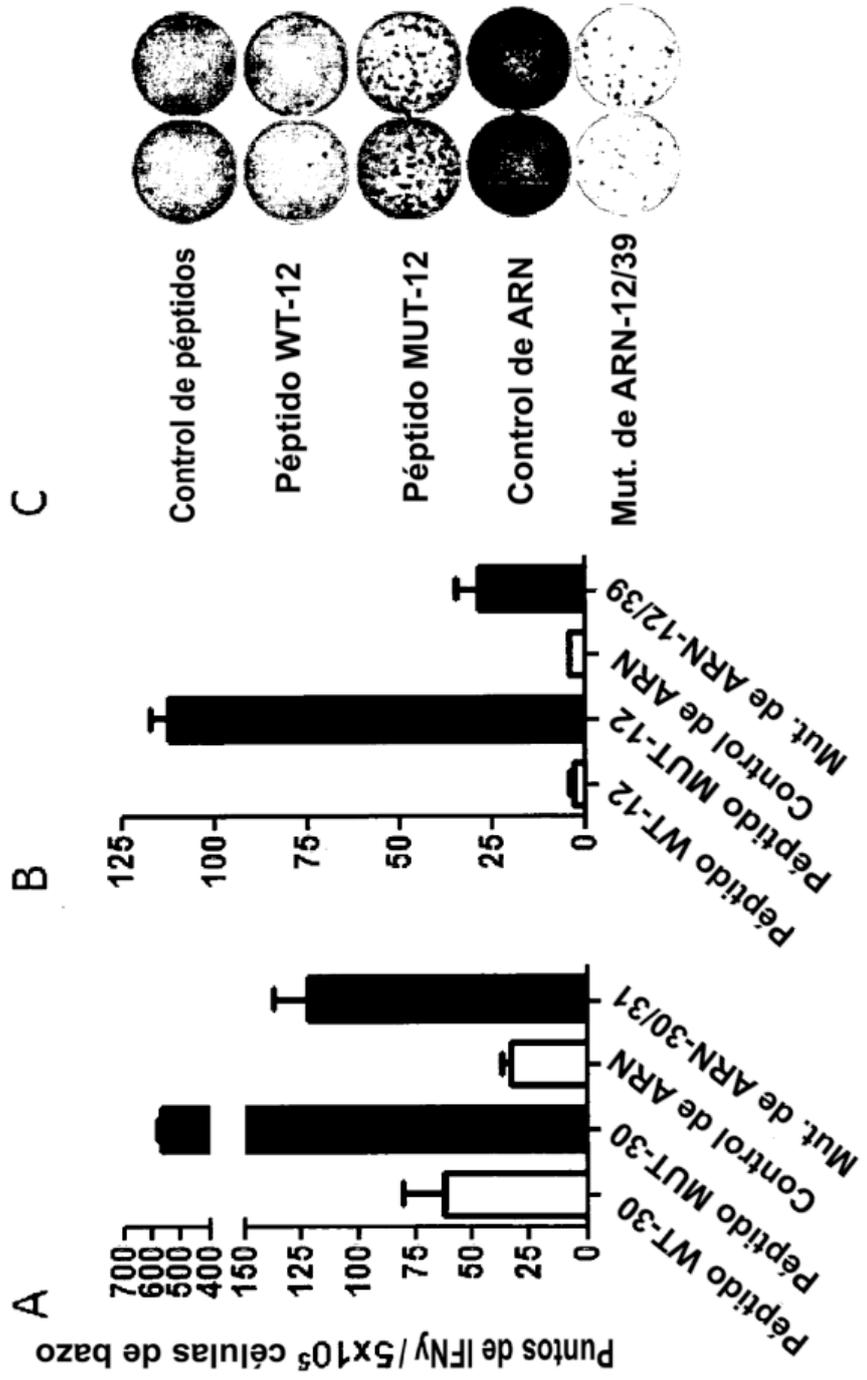


Fig. 8

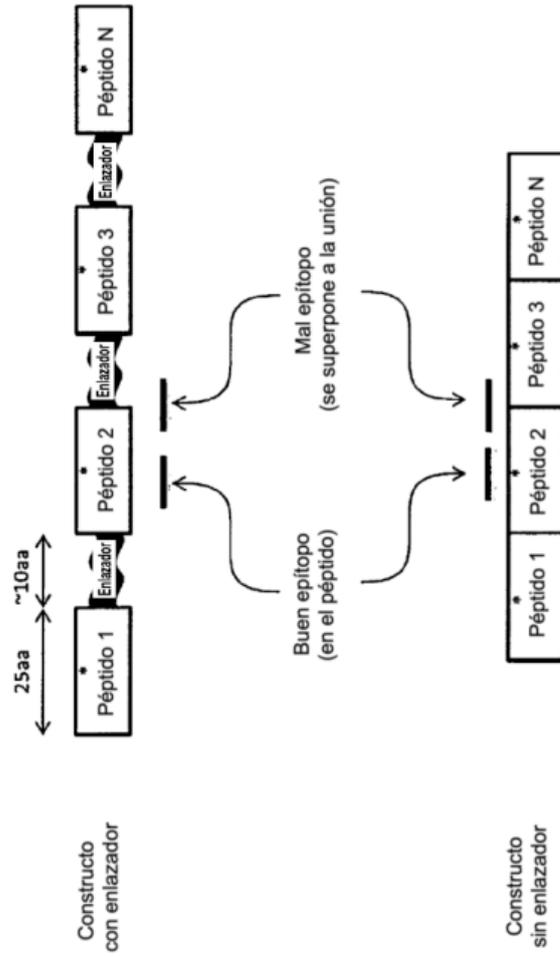


Fig. 9

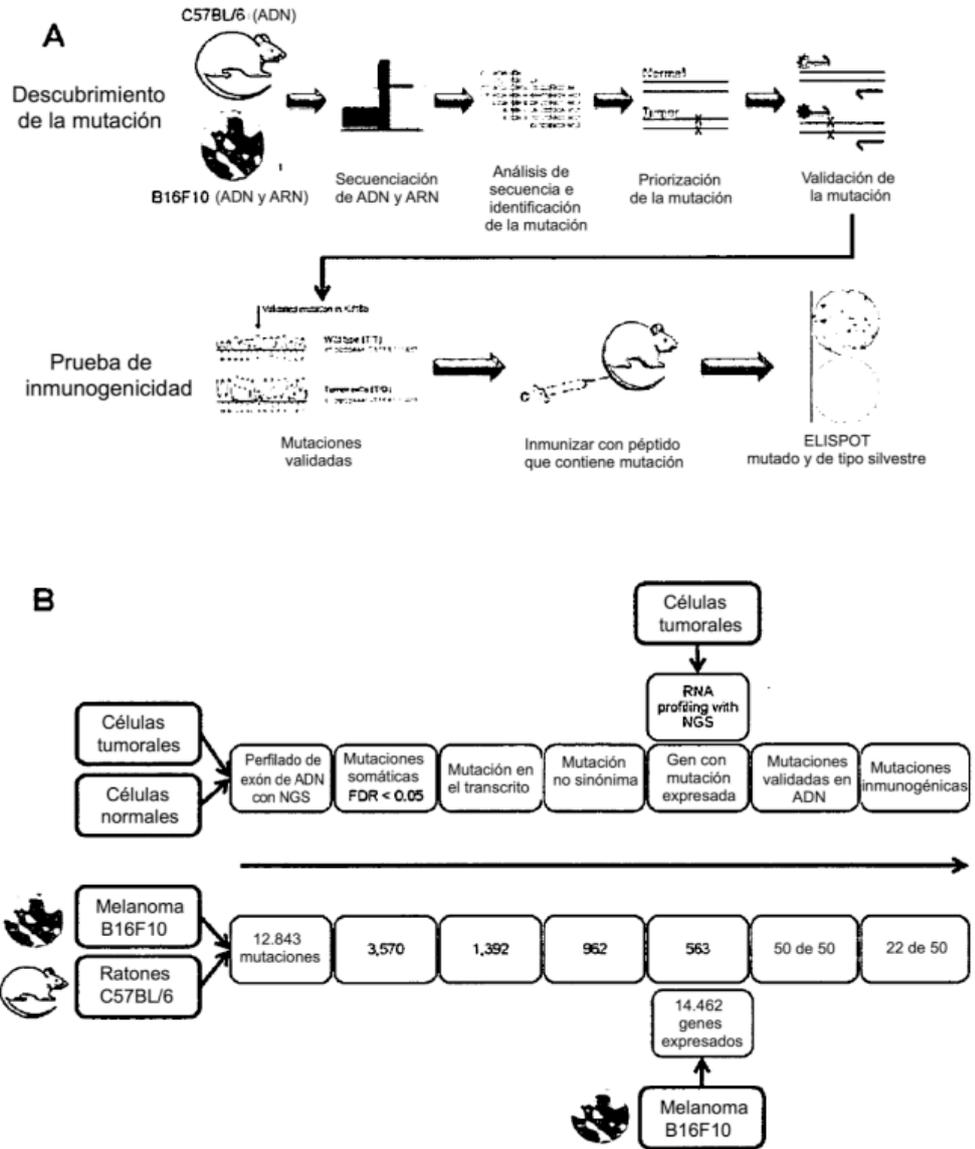


Fig. 9

C

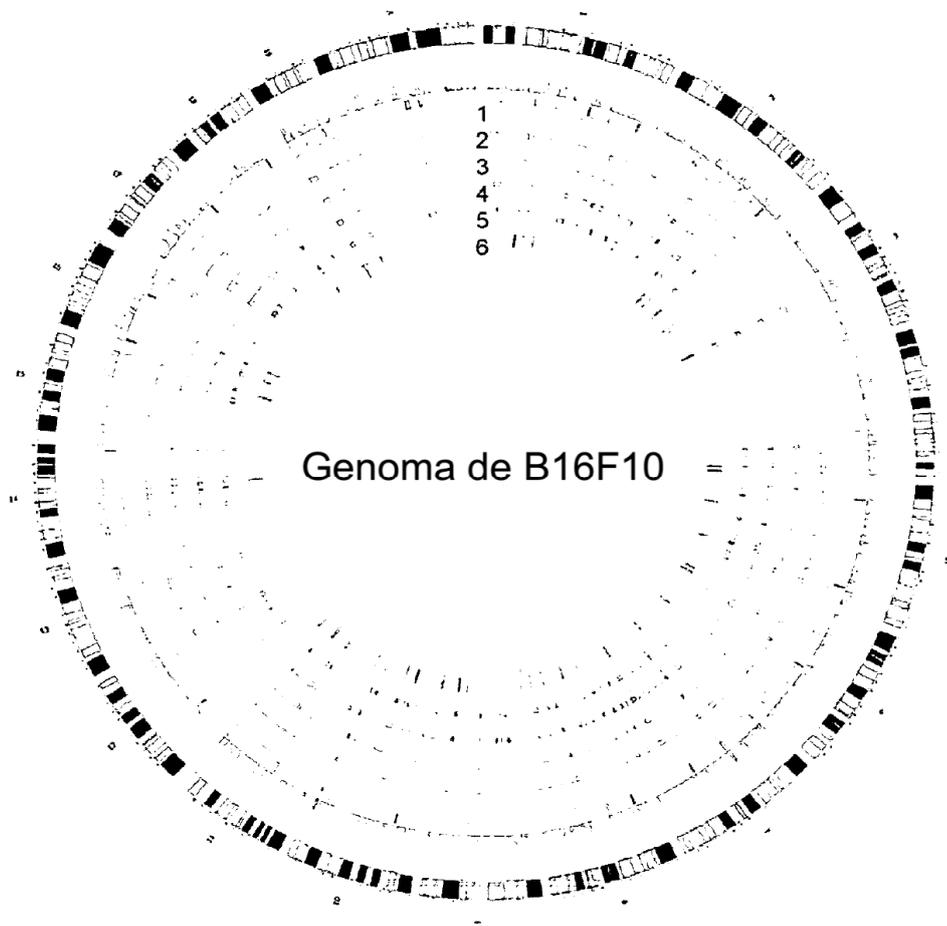


Fig. 10

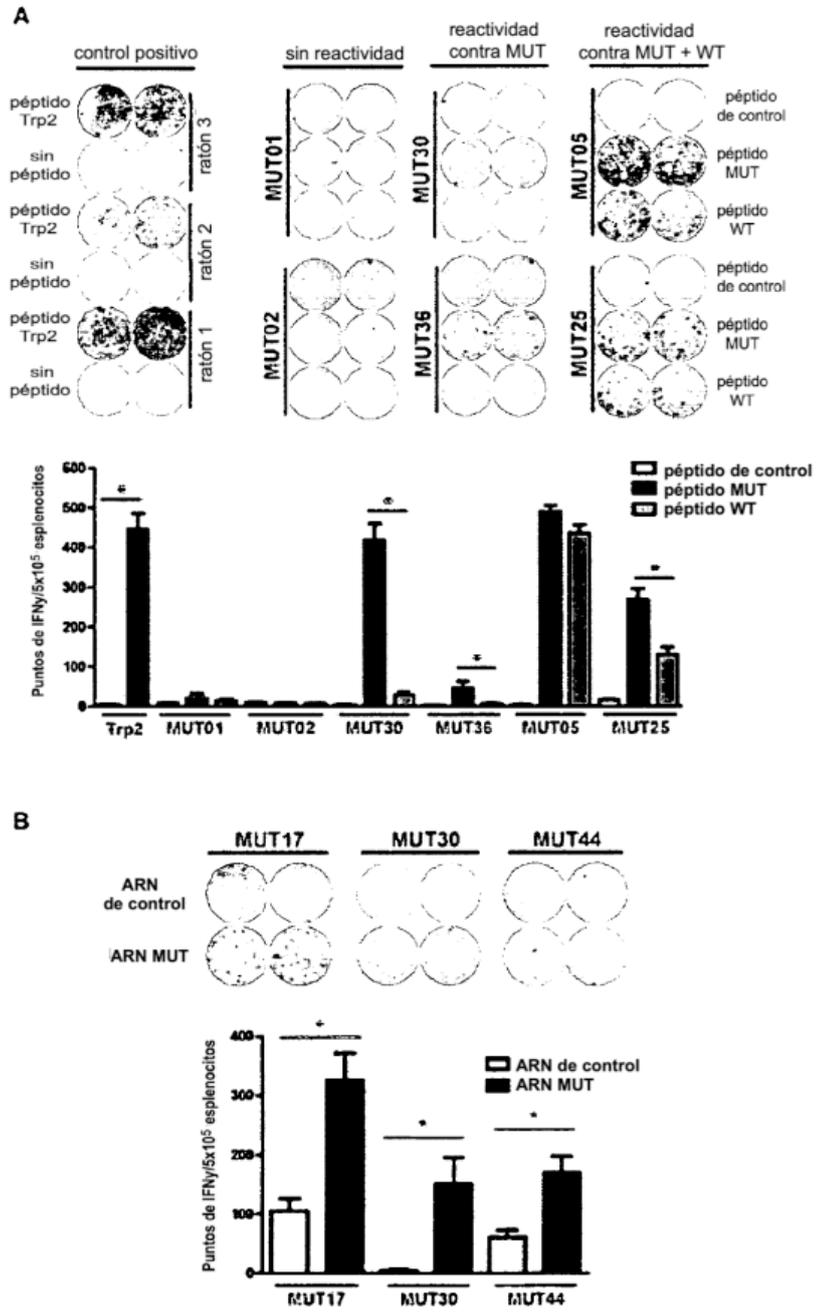


Fig. 10

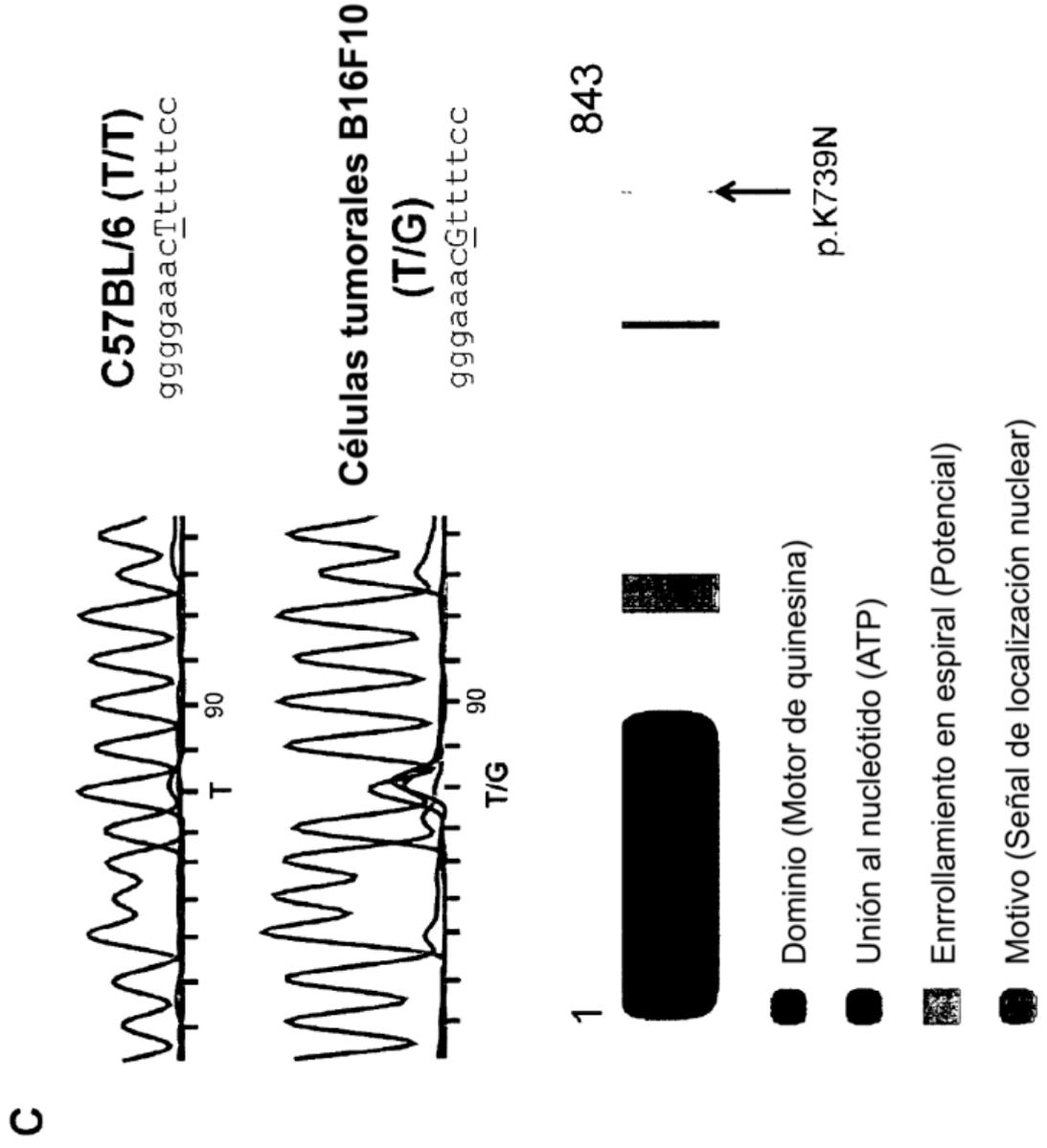


Fig. 11

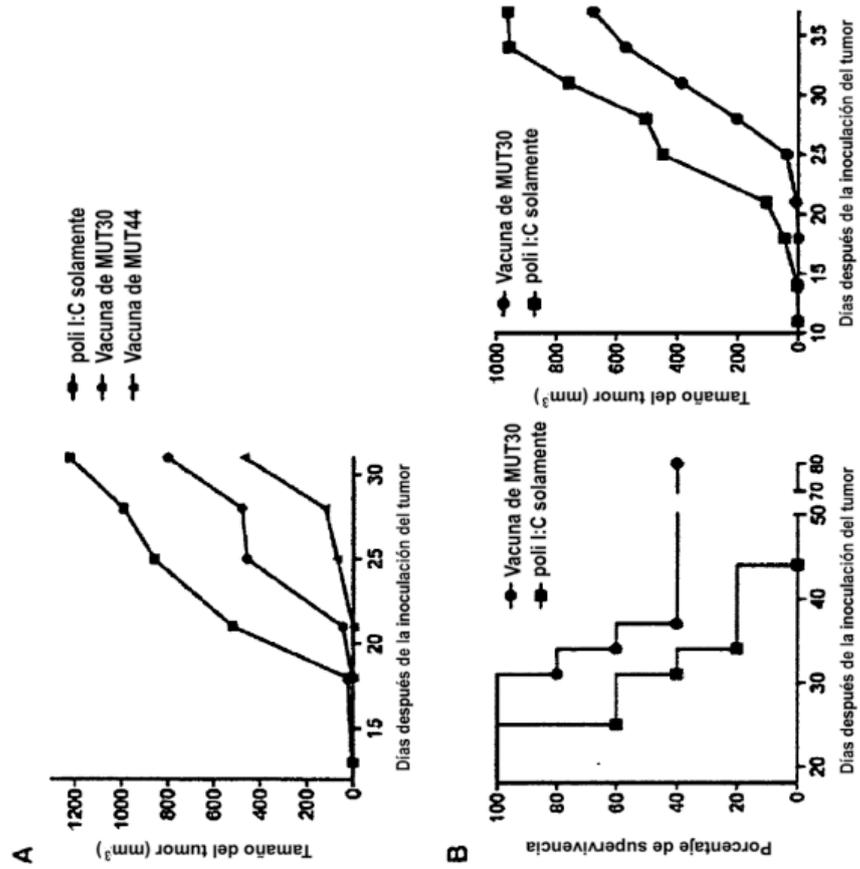


Fig. 12

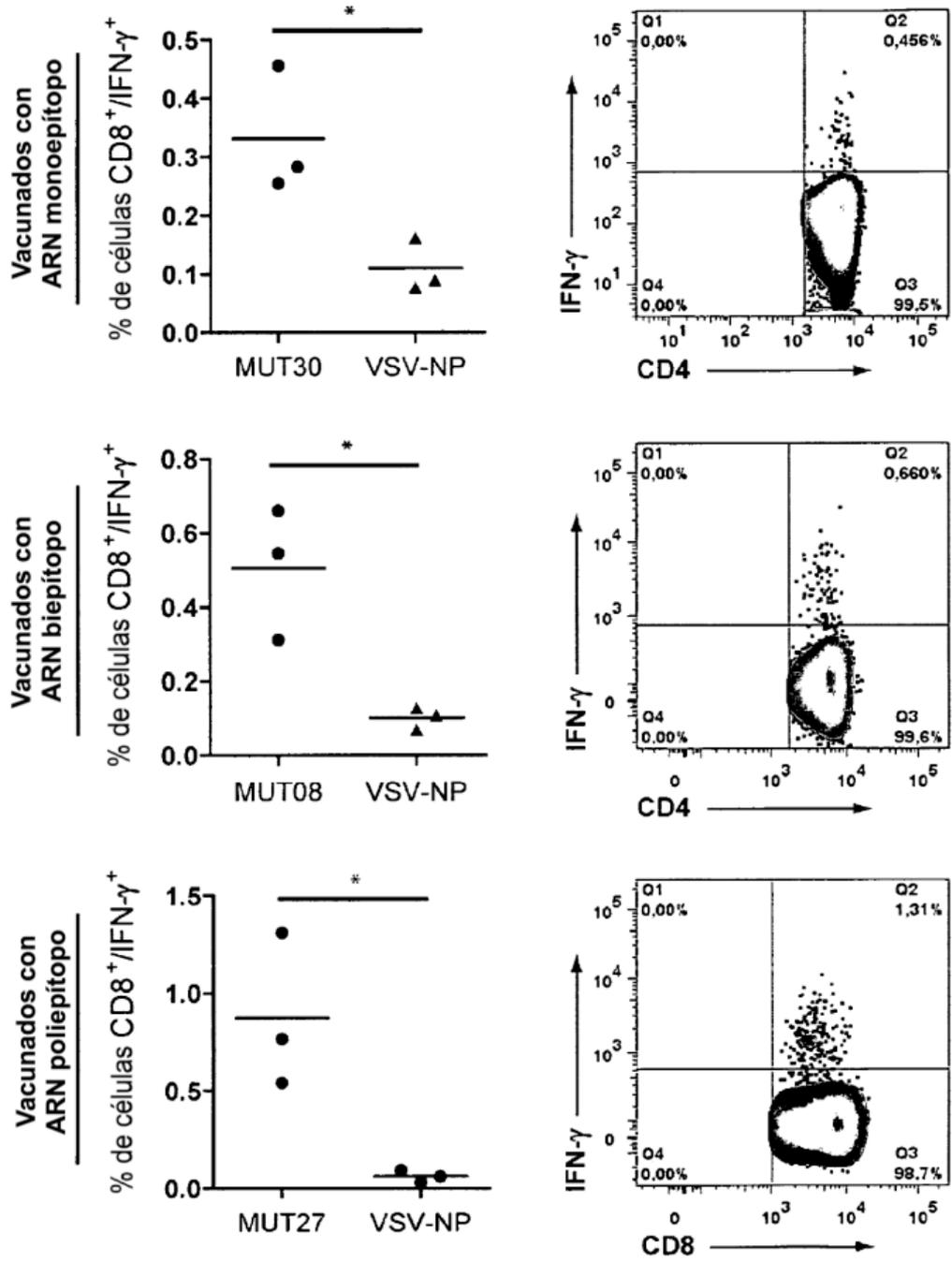


Fig. 13

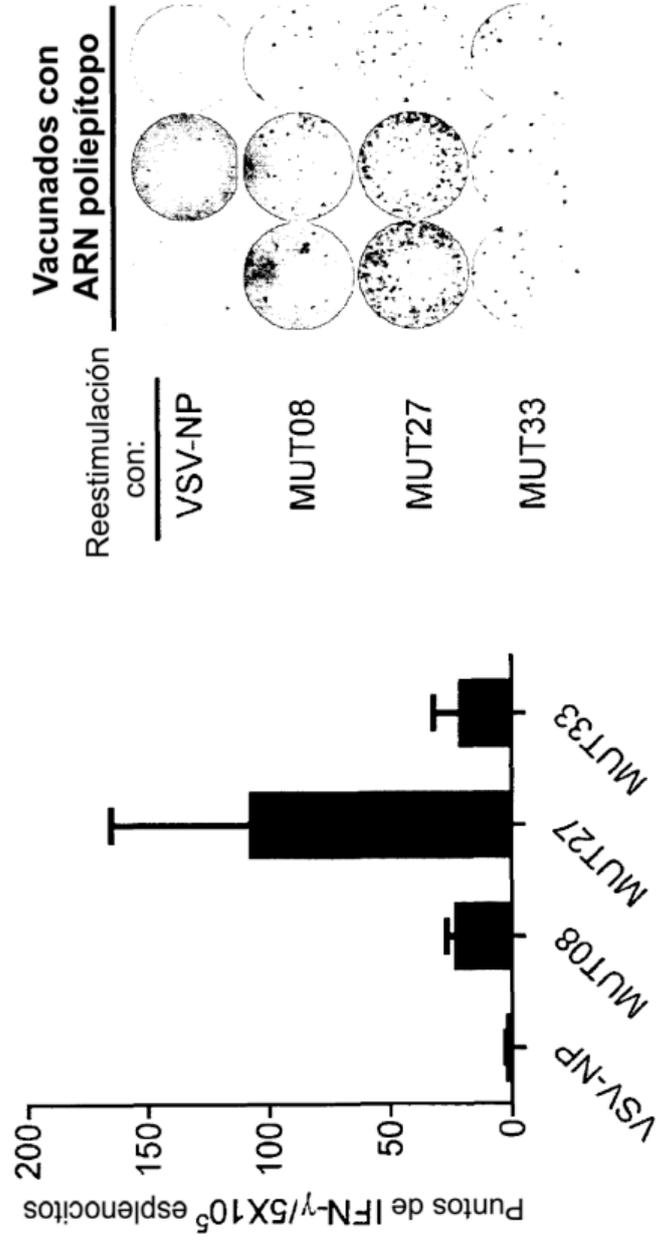


Fig. 14

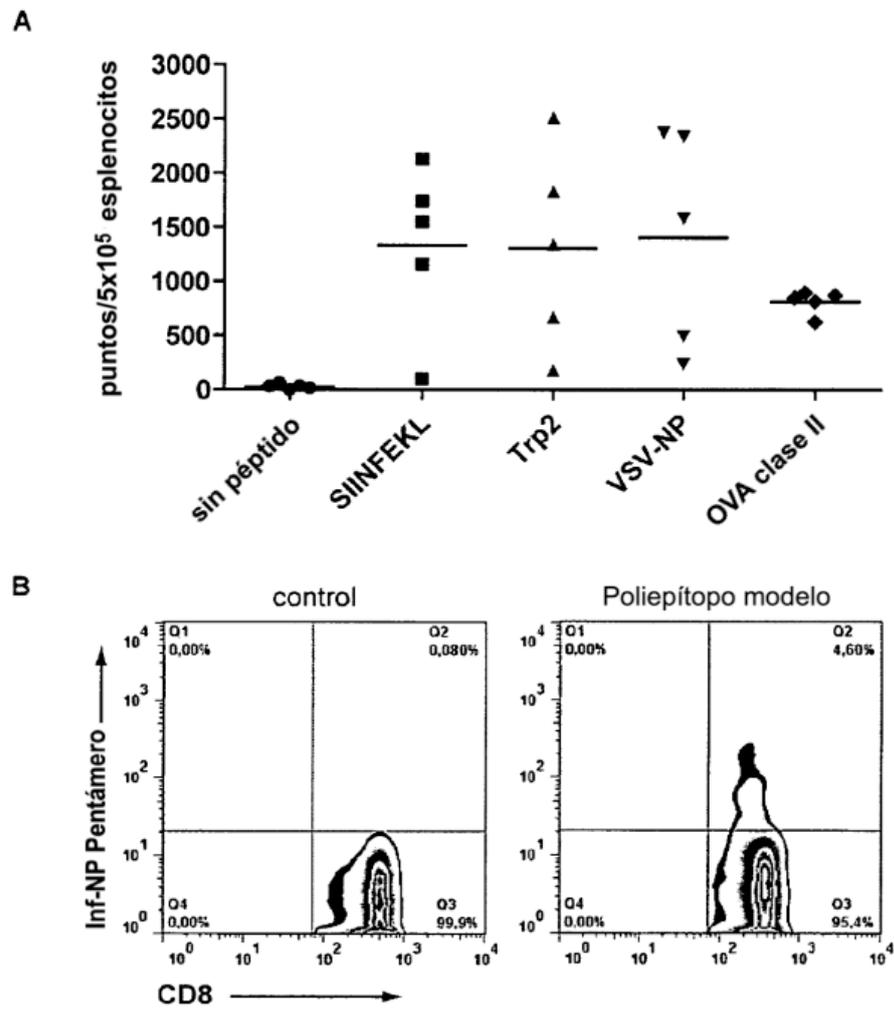
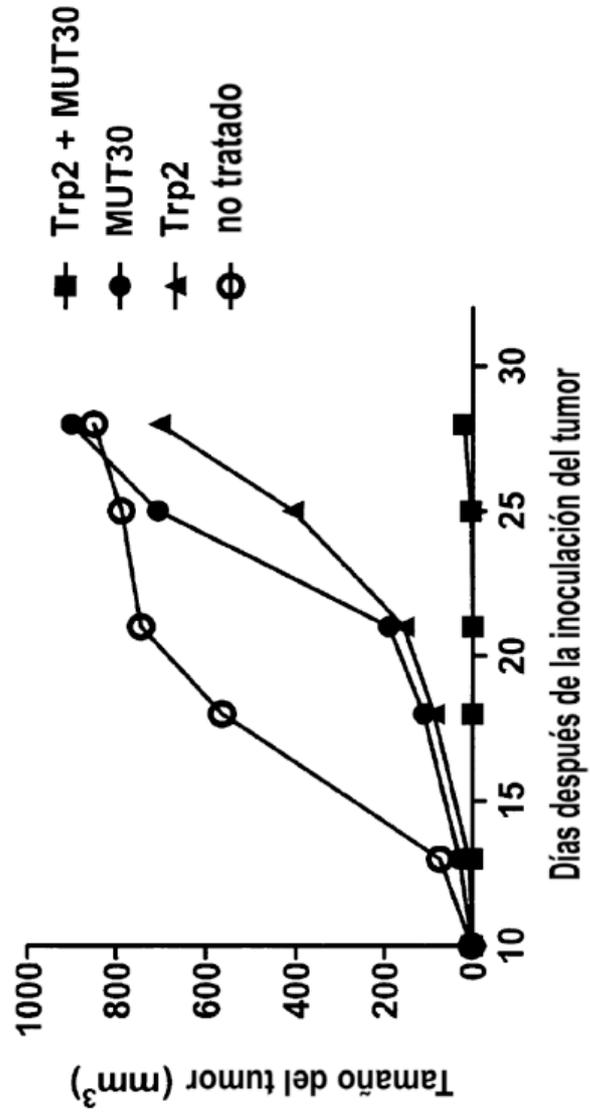


Fig. 15 A



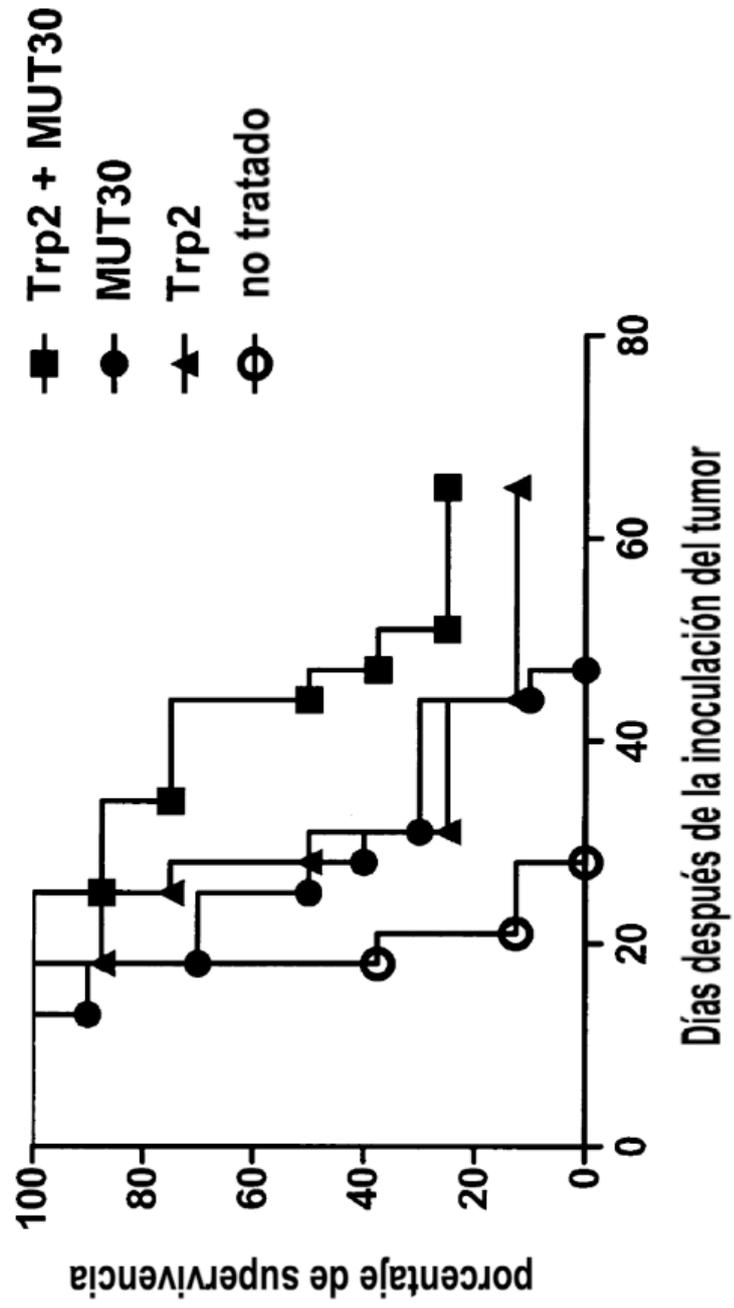


Fig. 15 B

Fig. 16

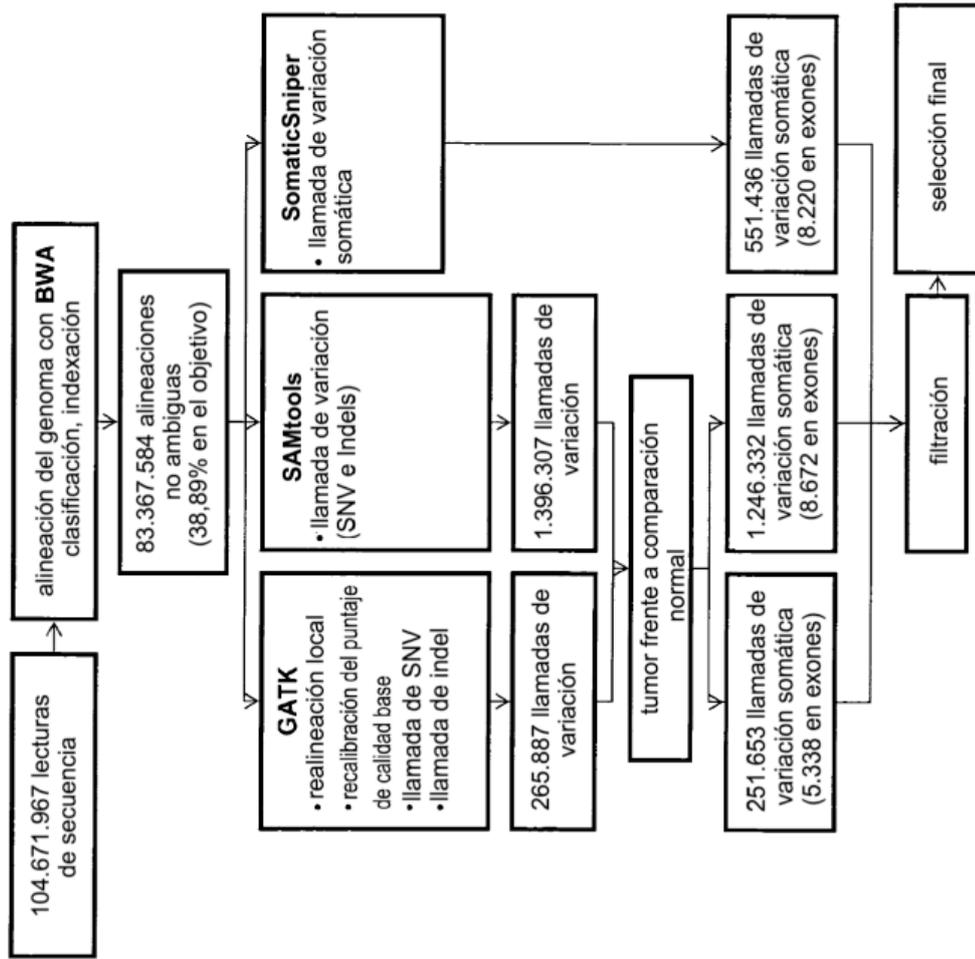


Fig. 17

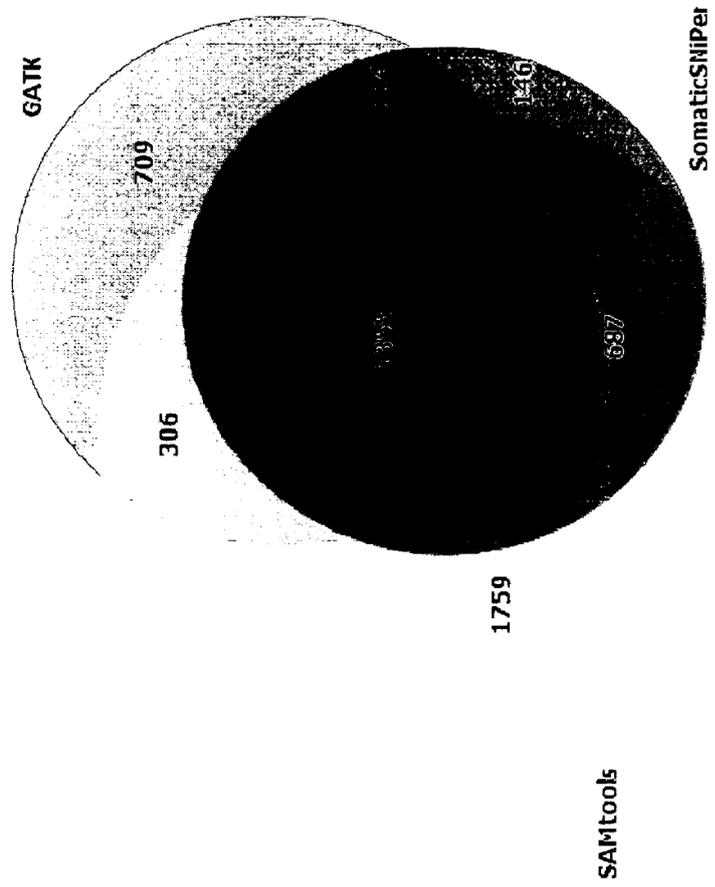


Fig. 18

A

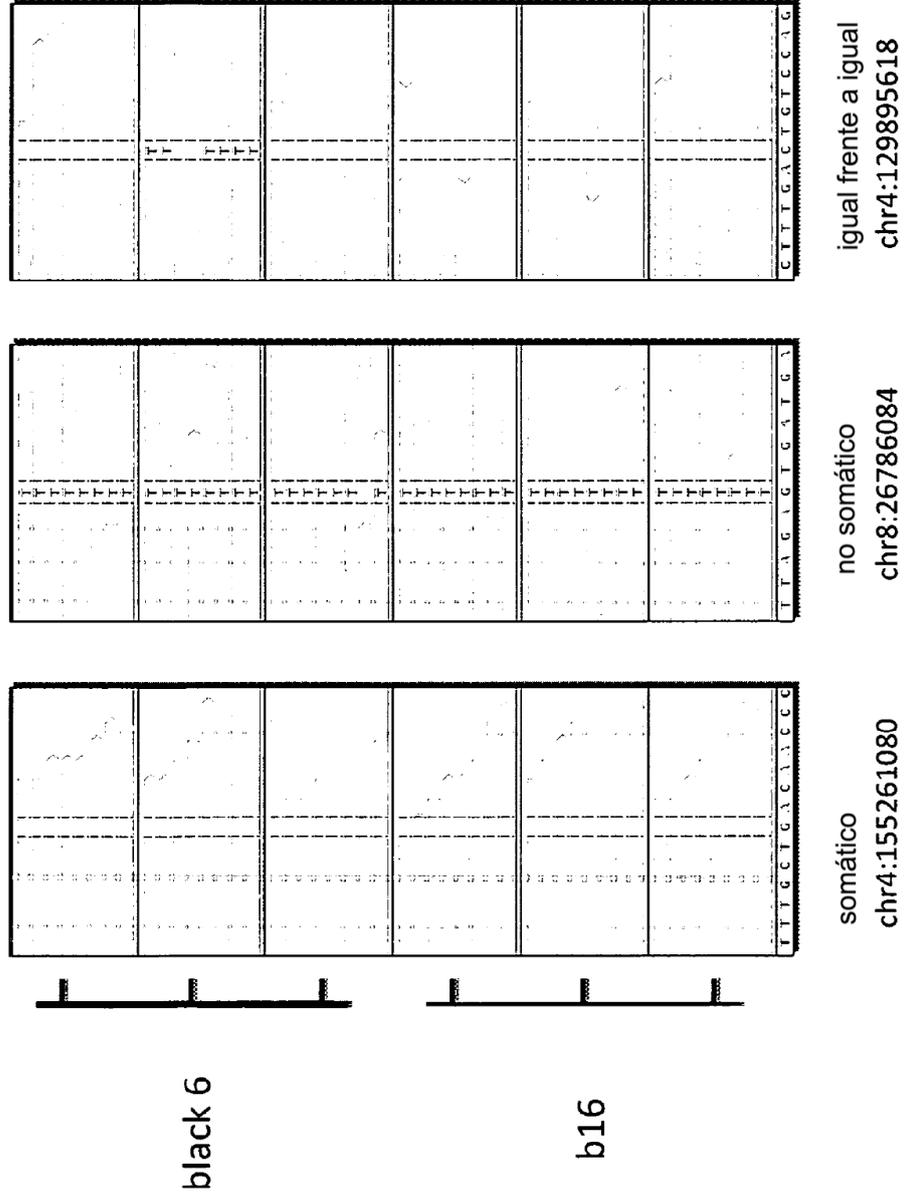


Fig. 18

B

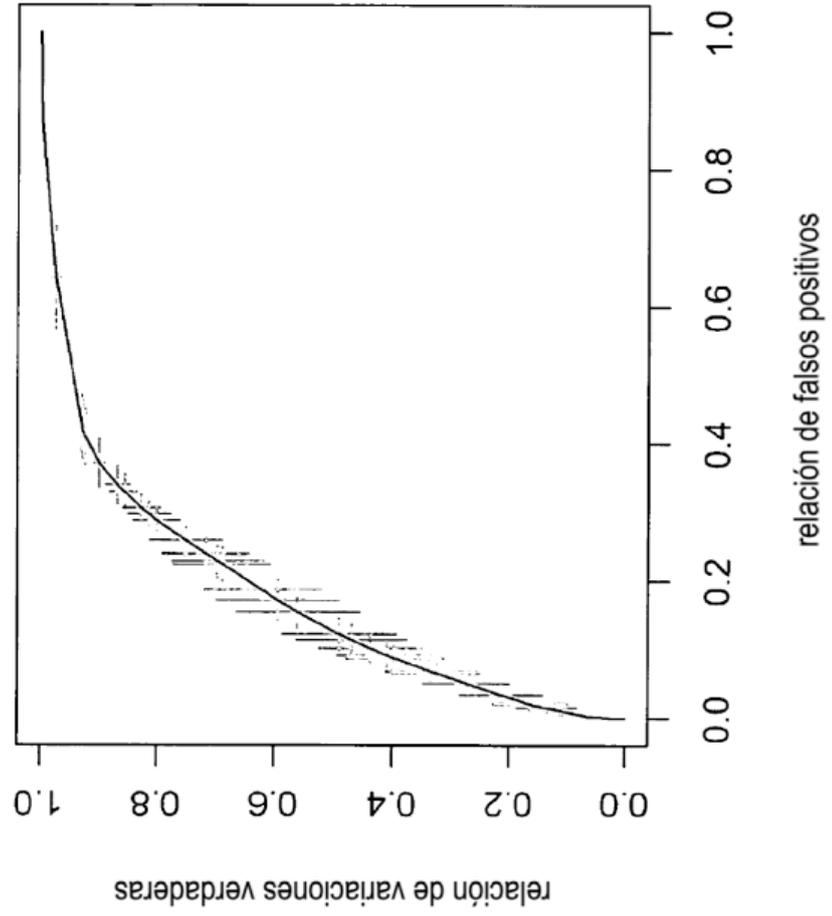


Fig. 19

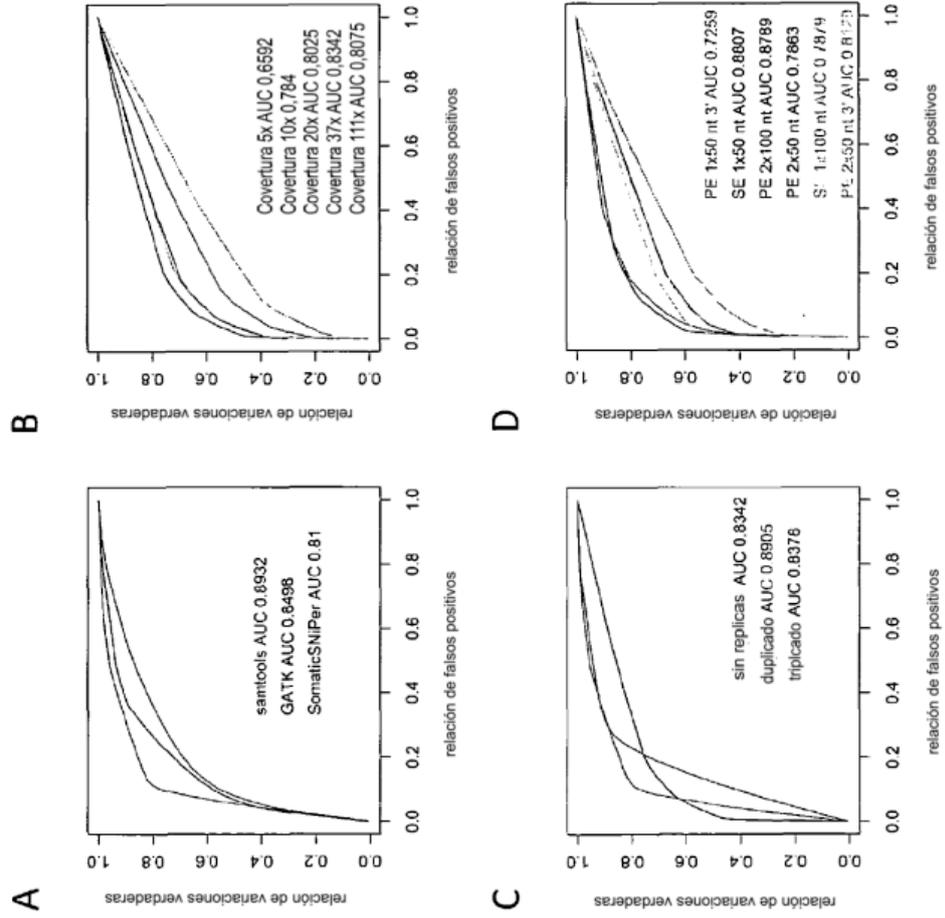


Fig. 20

A

Cromosoma	Posición	Alelo de referencia	Alelo(s) de muestra	Rango relativo [%]	FDR [%]
8	110078987	G	R	3,2	0,6
1	59540714	G	S	7,6	0,7
5	124854313	G	K	7,8	0,7
10	59352802	C	M	8,3	0,7
16	36919828	A	M	9,8	0,7
2	144078227	C	Y	10,1	0,7
8	12834637	G	S	10,9	0,7
19	6121411	T	Y	11,4	0,7
1	58533360	A	M	12,1	0,7
15	98478052	A	R	13,6	0,7

B

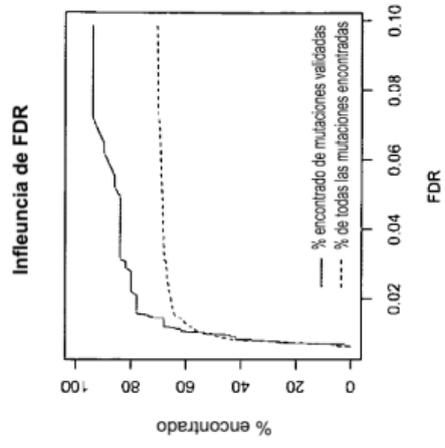


Fig. 21

