

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 674 928**

51 Int. Cl.:

C12Q 1/68 (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.11.2013 PCT/AU2013/001275**

87 Fecha y número de publicación internacional: **08.05.2014 WO14066955**

96 Fecha de presentación y número de la solicitud europea: **05.11.2013 E 13850884 (1)**

97 Fecha y número de publicación de la concesión europea: **18.04.2018 EP 2914750**

54 Título: **Métodos para determinar la causa de la mutagénesis somática**

30 Prioridad:

**05.11.2012 AU 2012904826
13.11.2012 AU 2012904940
12.04.2013 AU 2013901253**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
05.07.2018

73 Titular/es:

**GMDX CO PTY LTD (100.0%)
Level 6, 152 Elizabeth Street
Melbourne, Victoria 3000 , AU**

72 Inventor/es:

LINDLEY, ROBYN ALICE

74 Agente/Representante:

ISERN JARA, Jorge

ES 2 674 928 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

Métodos para determinar la causa de la mutagénesis somática

5 Campo de la invención

Esta invención se refiere generalmente a métodos para determinar la probabilidad de que haya ocurrido mutagénesis somática dirigida de una molécula de ácido nucleico por un agente mutagénico, y la probabilidad de que un agente mutagénico sea una causa de mutagénesis somática dirigida de una molécula de ácido nucleico. La invención además se refiere a métodos para diagnosticar cáncer en un sujeto y/o determinar la probabilidad de que un sujeto tenga o desarrollará cáncer, y métodos para tratar sujetos diagnosticados con cáncer o determinados como probables de tener o desarrollar cáncer. En aspectos adicionales, la invención se refiere a métodos para identificar motivos en moléculas de ácido nucleico que son reconocidos o dirigidos por agentes mutagénicos.

15 Aplicaciones relacionadas

Esta solicitud reivindica la prioridad a la Solicitud Provisional australiana N.º 2012904826, titulada "Method for the diagnosis of disease associated with somatic mutagenesis", presentada el 5 de noviembre de 2012; Solicitud Provisional australiana N.º 2012904940, titulada "Database system and method for the diagnosis of disease associated with somatic mutagenesis", presentada el 13 de noviembre de 2012; y la Solicitud Provisional australiana N.º 2013901253, titulada "Method for the diagnosis of disease associated with mutagenesis using small numbers of somatic mutations", presentada el 12 de abril de 2013.

Antecedentes de la invención

25 La progresión de las células normales a células cancerosas puede estar influida por una diversidad de factores, incluyendo cambios en el sistema inmune, estado hormonal, expresión génica y señalización entre tejidos. Un factor particularmente importante en la progresión del cáncer es la mutación somática, la cual juega un papel en los cánceres de la mayoría, si no en todos, los tipos de tejido.

30 La acumulación de mutaciones somáticas en diversos genes parece directamente relacionada con la progresión del cáncer. Esto se ha demostrado usando diversos modelos animales en los que un incremento en la mutagénesis somática resultante de, por ejemplo, corrección de ADN polimerasa o reparación de ADN dañado, estaba asociado con la progresión tumoral acelerada (véase, por ejemplo, Venkatesan y col. (2007). *Mol. Cell. Biol.* 27:7.669-7.682; y Albertson (2009) *Proc. Natl. Acad. Sci. U.S.A.* 106, 17.101-17.104). La mutagénesis somática incrementada de diversos genes también se ha asociado con una diversidad de cánceres. Por ejemplo, las mutaciones somáticas en el gen de TP53 son una de las alteraciones más frecuentes en cánceres humanos. Las mutaciones somáticas de TP53 ocurren en casi todo tipo de cáncer a tasas de 38 % a 50 % en cánceres de ovario, esofágicos, colorrectales, de cabeza y cuello, laringe y pulmón a aproximadamente 5 % en leucemia primaria, sarcoma, cáncer testicular, melanoma maligno y cáncer cervical, y subtipos de cáncer de estadio avanzado o agresivos (tales como los cánceres de mama amplificadas por HER2 o triple negativos) están asociados con una frecuencia incrementada de mutaciones somáticas en TP53 (publicado en Olivier y col. (2010) *Cold Spring Harb Perspect Biol.* 2:a001008). Otros genes asociados con cáncer que acumulan mutaciones somáticas incluyen, por ejemplo, BRAF, HRAS, KRAS2 y NRAS, aunque en la actualidad están incluidos más de 25.000 genes en COSMIC, la base de datos en línea de mutaciones somáticamente adquiridas encontradas en el cáncer humano.

50 La mutagénesis somática puede estar causada por factores ambientales, tales como el humo de cigarrillos, la luz y la radiación UV, y/o factores o procesos biológicos, tales como translocación de cromosoma, mala reparación o no reparación de ADN, e hipermutación somática iniciada por enzima (HMS). Determinar la causa y la extensión de la mutagénesis somática en células no solamente puede ayudar en el diagnóstico de las afecciones asociadas con la mutagénesis somática o predecir el riesgo de desarrollar tales afecciones, sino también puede ayudar en el desarrollo de los protocolos de tratamiento o prevención más apropiados. Por tanto, hay una necesidad de métodos precisos para determinar la presencia de mutagénesis somática e identificar qué agente mutagénico o agentes son responsables de la mutagénesis somática en un sujeto. En la técnica existente, Beale y col. describe el uso de pruebas bioquímicas para analizar las especificidades de ADN diana de la citidina desaminasa inducida por la activación de agentes mutagénicos (AID) y citidina desaminasa similar al polipéptido catalítico de enzima editora del ARNm de la apolipoproteína B (APOBEC), lo cual revela una importancia crítica de los nucleótidos inmediatamente 5' de la C dirigida para la especificidad (Beale y col. (2004). *J. Mol. Biol.* 337(3):585-96). Nik-Zainal y col. describe el análisis de mutaciones somáticas causadas por APOBEC en un contexto de secuencia en muestras de cáncer de mama en las que se calcula la frecuencia de mutación en diferentes motivos (Nik-Zainal y col. (2012) *Cell* 149(5):979-93).

Compendio de la invención

65 La presente invención se define en la reivindicación 1 independiente, y sus ciertas características opcionales se definen en las reivindicaciones dependientes. En la medida en que los términos "invención", "ejemplo" y "realización"

se usan en el presente documento, esto se interpretará de tal modo que la única protección debería ser para la invención como se reivindica. La presente invención se basa en parte en la determinación de que hay una propensión hacia la mutagénesis somática mediante diversos agentes mutagénicos en motivos cuando los motivos están presentes en un contexto de codón particular dentro de la molécula de ácido nucleico. Por tanto, aunque

5 previamente se entendió que algunos agentes mutagénicos se dirigen a motivos, como se describe en el presente documento, la mutagénesis en estos motivos ocurre predominantemente cuando los motivos están dentro de un contexto de codón particular, un proceso calificado en el presente documento como mutagénesis somática dirigida. Al identificar este requerimiento adicional para el contexto de codón del motivo, los presentes inventores han desarrollado métodos para determinar la probabilidad de que haya ocurrido este tipo de mutagénesis somática

10 dirigida, y la probabilidad de que uno o más agentes mutagénicos particulares sean la causa de la mutagénesis somática dirigida. También se ha desarrollado métodos generalizados para identificar motivos dirigidos por agentes mutagénicos a través de la evaluación de casos de mutación somática dirigida y están descritos en el presente documento.

15 Ya que la acumulación de mutaciones somáticas está asociada con el desarrollo y la progresión del cáncer, también se han desarrollado métodos para diagnosticar cáncer en un sujeto y determinar la probabilidad de que un sujeto tenga o desarrollará cáncer. Al identificar el agente mutagénico causativo y/o diagnosticar el cáncer o la probabilidad de desarrollar cáncer, se pueden desarrollar protocolos de tratamiento apropiados y específicos para inhibir o reducir la actividad del agente mutagénico, y/o tratar o prevenir el cáncer.

20 Por tanto, en un aspecto, la presente invención se dirige a métodos para detectar o determinar la probabilidad de que haya ocurrido mutagénesis somática dirigida de una molécula de ácido nucleico mediante un agente mutagénico, comprendiendo el análisis de la secuencia de la molécula de ácido nucleico para determinar el contexto de codón de mutaciones de un tipo de mutación en uno o más motivos, en los que se hace una determinación de

25 que se ha detectado mutagénesis somática dirigida o es probable que haya ocurrido cuando hay un porcentaje o número mayor de lo esperado de las mutaciones en una posición en codones en la molécula de ácido nucleico.

Generalmente, el porcentaje o el número esperado de mutaciones se calcula asumiendo que las mutaciones ocurren independientemente del contexto de codón. En algunas realizaciones, el porcentaje esperado de mutaciones es de

30 aproximadamente 11 % o 17 %, y/o el número esperado de mutaciones es aproximadamente 1 de cada 9 mutaciones o 1 de cada 6 mutaciones. En algunos ejemplos, se observa que el porcentaje de mutaciones es al menos del 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 80 %, 85 %, 90 %, 95 % o más.

Los métodos para determinar si ha ocurrido mutagénesis somática dirigida puede comprender además la

35 determinación de qué agente mutagénico es una causa probable de la mutagénesis somática dirigida. El agente mutagénico se puede seleccionar de, por ejemplo, aflatoxina, 4-aminobifenilo, ácidos aristolóquicos, compuestos arsénicos, asbestos, azatioprina, benceno, bencidina, berilio y compuestos de berilio, 1,3-butadieno, 1,4-butanodiol dimetilsulfonato, cadmio y compuestos de cadmio, clorambucilo, 1-(2-cloroetil)-3-(4-metilciclohexil)-1-nitrosourea (MeCCNU), bis(clorometil) éter y clorometil metil éter de grado técnico, compuestos cromo hexavalente, breas de

40 alquitrán, alquitranes, emisiones de horno de coque, ciclofosfamida, ciclosporina A, dietilestilbestrol (DES), erionita, óxido de etileno, formaldehído, melfalán, terapia de metoxsaleno con ultravioleta A (PUVA), gas mostaza, 2-naftilamina, neutrones, compuestos de níquel, radón, sílice cristalino (tamaño respirable), radiación solar, hollín, vahos de ácidos inorgánicos fuertes que contienen ácido sulfúrico, tamoxifeno, 2,3,7,8-tetraclorodibenzo-p-dioxin (TCDD), tiotepa, dióxido de torio, humo de tabaco, cloruro de vinilo, radiación ultravioleta, polvo de madera,

45 radiación X, radiación gamma, citidina desaminasa inducida por activación (AID), una citidina desaminasa similar a polipéptido catalítico de enzima editora de ARNm de apolipoproteína B (APOBEC), y ADN polimerasas propensas a errores. En algunos ejemplos, la citidina desaminasa APOBEC se selecciona de entre APOBEC1, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G y APOBEC3H.

50 En realizaciones particulares en las que el agente mutagénico se selecciona de entre AID, APOBEC1, APOBEC3G, APOBEC3H y aflatoxina, una determinación de que AID es una causa probable de mutagénesis somática dirigida se hace si el número o el porcentaje de mutaciones G>A observadas en motivos GYW en la segunda posición en los codones (sitios MC-2) en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; una determinación de que AID es una causa probable de mutagénesis somática dirigida se hace si el número o el

55 porcentaje de mutaciones C>T observadas en motivos WRC en la primera posición en los codones (sitios MC-1) en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; una determinación de que APOBEC3G es una causa probable de mutagénesis somática dirigida se hace si el número o el porcentaje de mutaciones G>A observadas en motivos CG en motivos en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; una determinación de que APOBEC3G es una causa probable de

60 mutagénesis somática dirigida se hace si hay un número o porcentaje mayor de lo esperado de mutaciones C>T observadas en motivos CG en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; una determinación de que APOBEC3G es una causa probable de mutagénesis somática dirigida se hace si hay un número o porcentaje mayor de lo esperado de mutaciones C>T observadas en motivos CC en sitios

65 MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; una determinación de que APOBEC3H es una causa probable de mutagénesis somática dirigida se hace si hay un número o porcentaje mayor de lo esperado de mutaciones G>A observadas en motivos GA en sitios MC-1 en la cadena no transcrita de la

molécula de ácido nucleico es mayor de lo esperado; una determinación de que APOBEC1 es una causa probable de mutagénesis somática dirigida se hace si el número o el porcentaje de mutaciones C>T observadas en motivos CA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; una determinación de que APOBEC1 es una causa probable de mutagénesis somática dirigida se hace si el número o el porcentaje de mutaciones G>A observadas en motivos TG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; y una determinación de que aflatoxina es una causa probable de mutagénesis somática dirigida se hace si el número o el porcentaje de mutaciones G>T observadas en motivos GG en sitios MC-3 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; en las que la molécula de ácido nucleico es de una muestra biológica de un sujeto.

Otras realizaciones de los métodos para determinar si ha ocurrido mutagénesis somática dirigida comprende además la determinación de si un proceso de mutación asociado a AID es una causa probable de la mutagénesis somática dirigida. Por ejemplo, se hace una determinación de que un proceso de mutación asociado a AID es una causa probable de mutagénesis somática dirigida si el número o el porcentaje de mutaciones A>G observadas en motivos WA en sitios MC-2, mutaciones G>A en motivos GYW en sitios MC-2, o mutaciones C>T en motivos WRC en sitios MC-1, en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.

En ejemplos particulares de los métodos de la presente invención, si se determina que AID es una causa probable de mutagénesis somática dirigida, los métodos comprenden además administrar un inhibidor de AID al sujeto; si se determina que APOBEC3G es una causa probable de mutagénesis somática dirigida, comprenden además administrar un inhibidor de APOBEC3G al sujeto; si se determina que APOBEC3H es una causa probable de mutagénesis somática dirigida, comprenden además administrar un inhibidor de APOBEC3G al sujeto; o si se determina que APOBEC1 es una causa probable de mutagénesis somática dirigida, comprenden además administrar un inhibidor de APOBEC1 al sujeto.

En realizaciones adicionales, los métodos también comprenden diagnosticar cáncer en el sujeto o determinar la probabilidad de que el sujeto desarrollará cáncer si se determina que ha ocurrido mutagénesis somática dirigida y/o un agente mutagénico es la causa probable de la mutagénesis somática dirigida.

En otros aspectos, la presente invención se dirige a métodos para determinar la probabilidad de que un sujeto tenga o desarrollará cáncer, comprendiendo analizar una molécula de ácido nucleico de una muestra biológica del sujeto para detectar si ha ocurrido mutagénesis somática dirigida por uno o más agentes mutagénicos, y determinar que el sujeto es probable que tenga o desarrolle cáncer cuando ha ocurrido mutagénesis somática dirigida.

En un ejemplo, la mutagénesis somática dirigida se detecta cuando: el número o el porcentaje de mutaciones G a A observadas en motivos GYW en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos WRC en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>A observadas en motivos CG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos CG en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos CA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>A observadas en motivos GA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>A observadas en motivos TG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>T observadas en motivos GG en sitios MC-3 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos CC en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; o el número o el porcentaje de mutaciones A>G observadas en motivos WA en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.

En ejemplos particulares, se determina que el agente mutagénico es AID si el número o el porcentaje de las mutaciones G>A observadas en motivos GYW en sitios MC-2 o mutaciones C>T en motivos WRC en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; APOBEC3G si el número o el porcentaje de las mutaciones G>A observadas en motivos CG en sitios MC-2, mutaciones C>T en motivos CG en sitios MC-1 o mutaciones C>T en motivos CC en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; APOBEC1 si el número o el porcentaje de las mutaciones C>T observadas en motivos CA en sitios MC-1 o mutaciones G>A en motivos TG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; APOBEC3H si el número o el porcentaje de las mutaciones G>A observadas en motivos GA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; o aflatoxina si el número o el porcentaje de mutaciones G>T observadas en motivos GG en sitios MC-3 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.

La muestra biológica puede comprender tejido o células de pecho, próstata, hígado, colon, estómago, pancreáticas, de piel, tiroideas, cervicales, linfoides, hematopoyéticas, de vejiga, pulmón, renales, rectales, de ovario, útero, y de cabeza o cuello, y, en algunos casos, el cáncer se selecciona de entre cáncer de mama, próstata, hígado, colon,

estómago, pancreático, de piel, tiroides, cervical, linfoides, hematopoyético, de vejiga, pulmón, renal, rectal, de ovario, útero y de cabeza y cuello. En ejemplos particulares, el carcinoma hepatocelular canceroso, melanoma o carcinoma adenoide quístico.

5 En algunas realizaciones de la presente invención, si la muestra comprende tejido o células de próstata, al sujeto se le diagnostica cáncer de próstata o se determina que es probable que tenga o desarrolle cáncer. En otras realizaciones, si la muestra comprende tejido o células de pecho, al sujeto se le diagnostica cáncer de mama o se determina que es probable que tenga o desarrolle cáncer de mama.

10 Los métodos de la presente invención pueden incluir además administrar terapia al sujeto, tal como, por ejemplo, radioterapia, cirugía, quimioterapia, terapia de ablación hormonal, terapia pro-apoptosis y/o inmunoterapia. En ejemplos particulares, los métodos incluyen administrar un inhibidor de AID, un inhibidor de APOBEC3G; un inhibidor de APOBEC1 y/o un inhibidor de APOBEC3H al sujeto.

15 En otro aspecto, la presente invención se dirige a métodos para identificar un motivo de ácido nucleico dirigido por un agente mutagénico, que comprenden analizar la secuencia de una molécula de ácido nucleico para identificar mutaciones somáticas de un tipo de mutación que se sabe que está asociado con el agente mutagénico; determinar el contexto de codón de las mutaciones para identificar la posición de nucleótido preferida en la que ocurren las mutaciones a una frecuencia mayor que la esperada; e identificar los nucleótidos que flanquean las mutaciones en la
20 posición de nucleótido preferida para identificar un motivo que es común a las mutaciones.

La invención también se dirige a métodos para identificar un motivo de ácido nucleico dirigido por un agente mutagénico, que comprende analizar la secuencia de una molécula de ácido nucleico para identificar mutaciones somáticas en la molécula de ácido nucleico; identificar un tipo de mutación que ocurre en una posición de nucleótido
25 preferida dentro de un codón a una frecuencia mayor que la esperada; e identificar los nucleótidos que flanquean el tipo de mutación en la posición de nucleótido preferida para identificar un motivo que es común al tipo de mutación.

El tipo de mutación se puede seleccionar de las mutaciones C>T, C>A, C>G, G>T, G>A, G>C, A>T, A>C, A>G, T>A, T>C y T>G, y la posición de nucleótido preferida se puede seleccionar de entre MC-1, MC-2 y MC-3.

30 En tales métodos, la frecuencia esperada se calcula asumiendo que las mutaciones ocurren independientemente del contexto de codón. Por ejemplo, la frecuencia esperada puede ser aproximadamente 1 de cada 9 mutaciones o 1 de cada 6 mutaciones. En algunas realizaciones, la mutación ocurre en la posición de nucleótido preferida al menos el 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 80 %, 85 %, 90 %, 95 % o más de las veces.

35 En algunas realizaciones de los métodos de la presente invención, se analiza la cadena no transcrita de la molécula de ácido nucleico.

40 El agente mutagénico puede ser endógeno o exógeno a las células de las que se obtiene el ácido nucleico. Por ejemplo el agente mutagénico se puede seleccionar de entre 4-aminobifenilo, ácidos aristolóquicos, compuestos arsénicos, asbestos, azatioprina, benceno, bencidina, berilio y compuestos de berilio, 1,3-butadieno, 1,4-butanodiol dimetilsulfonato, cadmio y compuestos de cadmio, clorambucilo, 1-(2-cloroetil)-3-(4-metilciclohexil)-1-nitrosourea (MeCCNU), bis(clorometil) éter y clorometil metil éter de grado técnico, compuestos de cromo hexavalente, brea de alquitrán, alquitranes, emisiones de horno de coque, ciclofosfamida, ciclosporina A, dietilestilbestrol (DES), erionita,
45 óxido de etileno, formaldehído, melfalán, terapia de metoxsaleno con ultravioleta A (PUVA), gas mostaza, 2-naftilamina, neutrones, compuestos de níquel, radón, sílice cristalino (tamaño respirable), radiación solar, hollín, vahos de ácido inorgánico fuerte que contienen ácido sulfúrico, tamoxifeno, 2,3,7,8-tetraclorodibenzo-p-dioxin (TCDD), tiotepa, dióxido de torio, humo de tabaco, cloruro de vinilo, radiación ultravioleta, polvo de madera, radiación X, radiación gamma, citidina desaminasa APOBEC y una ADN polimerasa propensa a error. En ejemplos
50 particulares de los métodos de la invención, la molécula de ácido nucleico o la célula de la que se obtiene la molécula de ácido nucleico se sabe que ha sido expuesto al agente mutagénico antes del análisis.

Las realizaciones de los métodos de la invención también pueden comprender primero aislamiento de la molécula de ácido nucleico y/o secuenciación de todo o parte de la molécula de ácido nucleico. La molécula de ácido nucleico
55 puede comprender todo o parte de un gen único o el ADNc de un gen único; o todo o parte de dos o más genes o el ADNc de dos o más genes. En algunos casos, el gen es un gen asociado con el cáncer. Por ejemplo, el gen se puede seleccionar de entre los genes de TP53, PIK3CA, ERBB2, DIRAS3, TET2 y óxido nítrico sintasa (NOS). En realizaciones adicionales, se analizaron la molécula de ácido nucleico que constituye el exoma completo de una célula o el genoma completo de una célula.

60 La invención también se dirige a un kit, que comprende un reactivo para su uso en un método descrito en el presente documento. El reactivo se puede seleccionar de, por ejemplo, entre un cebador, dNTP y polimerasa.

65 En realizaciones particulares de los métodos de la presente invención, todo o un parte del método se realiza mediante un sistema de procesamiento.

Breve descripción de los dibujos

- 5 La Figura 1 es un esquema que muestra la mutación somática dirigida en una región de interés en la cadena no transcrita de una molécula de ácido nucleico.
- La Figura 2 es un esquema que muestra un proceso como ejemplo del análisis de moléculas de ácido nucleico para determinar si ha ocurrido mutagénesis somática dirigida por AID o AOPBEC3G.
- 10 La Figura 3 es un esquema que muestra un análisis realizado para determinar si las mutaciones ocurren al azar o como resultado de la mutagénesis somática dirigida.
- La Figura 4 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1) y sitios WA en el gen de TP53 del ácido nucleico obtenido de los sujetos con cáncer cervical, y el análisis estadístico de la ocurrencia de las mutaciones.
- 15 La Figura 5 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1) y sitios WA en el gen de TP53 de ácido nucleico obtenido de sujetos con adenocarcinoma de colon, y el análisis estadístico de la ocurrencia de las mutaciones.
- 20 La Figura 6 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1), sitios WA, y sitios GG (aflatoxina) en el gen de TP53 del ácido nucleico obtenido de sujetos con carcinoma hepatocelular, y el análisis estadístico de la ocurrencia de las mutaciones.
- 25 La Figura 7 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1) y sitios WA en el gen de TP53 del ácido nucleico obtenido de sujetos con cáncer pancreático, y el análisis estadístico de la ocurrencia de las mutaciones.
- 30 La Figura 8 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1) y sitios WA en el gen de TP53 del ácido nucleico obtenido de sujetos con cáncer de próstata, y el análisis estadístico de la ocurrencia de las mutaciones.
- 35 La Figura 9 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA en el gen de TP53 del ácido nucleico obtenido de sujetos con melanoma maligno, y el análisis estadístico de la ocurrencia de las mutaciones.
- La Figura 10 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1) y sitios WA en el gen de TP53 del ácido nucleico obtenido de sujetos con adenocarcinoma cervical, y el análisis estadístico de la ocurrencia de las mutaciones.
- 40 La Figura 11 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA en el gen de NOS del ácido nucleico obtenido de sujetos con adenocarcinoma cervical, y el análisis estadístico de la ocurrencia de las mutaciones.
- 45 La Figura 12 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA en el gen de PIK3CA del ácido nucleico obtenido de sujetos con cáncer de mama, y el análisis estadístico de la ocurrencia de las mutaciones.
- 50 La Figura 13 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G), sitios TG/CA (APOBEC1) y sitios WA en el gen de TET2 del ácido nucleico obtenido de sujetos con cáncer hematopoyético y linfóide, y el análisis estadístico de la ocurrencia de las mutaciones.
- 55 La Figura 14 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA en el exoma completo de tejido obtenido de dos sujetos con carcinoma adenoide quístico, y el análisis estadístico de la ocurrencia de las mutaciones. (A) Sujeto PD3185a. (B) Sujeto PD3181a.
- 60 La Figura 15 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA en el exoma completo del tejido obtenido de cuatro sujetos con carcinoma de próstata, y el análisis estadístico de la ocurrencia de las mutaciones. (A) Sujeto WA7. (B) Sujeto WA26. (C) Sujeto PR-09-3421. (D) Sujeto PR-2762.

La Figura 16 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios GA (APOBEC3H) en el exoma completo del ácido nucleico obtenido de sujetos con cáncer de vejiga, y el análisis estadístico de la ocurrencia de las mutaciones.

5 La Figura 17 muestra la frecuencia y la localización dentro de los codones de mutaciones en los sitios CC (APOBEC3G) en el exoma completo del ácido nucleico obtenido de 8 sujetos con cáncer de vejiga (A), y un sujeto único con cáncer de vejiga (B), y el análisis estadístico de la ocurrencia de las mutaciones.

10 La Figura 18 es un esquema del proceso de detección de mutagénesis somática dirigida en una molécula de ácido nucleico usando un sistema de procesamiento.

TABLA A
SÍMBOLOS DE NUCLEÓTIDOS

SÍMBOLO	DESCRIPCIÓN
A	Adenosina
C	Citidina
G	Guanosina
T	Timidina
U	Uridina
M	Amino (adenosina, citosina)
K	ceto (guanosina, timidina)
R	Purina (adenosina, guanosina)
Y	Pirimidina (citosina, timidina)
w	Adenosina o citosina
N	Cualquier nucleótido

Descripción detallada de la invención

15 *1. Definiciones*

A menos que se defina lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen el mismo significado que el frecuentemente entendido por los expertos en la técnica a la que pertenece la invención. Aunque se pueden usar algunos métodos y materiales similares o equivalentes a los descritos en el presente documento en la práctica o el ensayo de la presente invención, se describen métodos y materiales preferidos. Para los fines de la presente invención, a continuación se definen los siguientes términos.

25 Los artículos “un” y “una” se usan en el presente documento para referirse a uno o a más de uno (es decir, a al menos uno) del objeto gramatical del artículo. A modo de ejemplo, “un elemento” significa un elemento o más de un elemento.

30 El término “muestra biológica” como se usa en el presente documento se refiere a una muestra que puede estar extraída, no tratada, tratada, diluida o concentrada de un sujeto o paciente. Adecuadamente, la muestra biológica se selecciona de cualquier parte del cuerpo de un paciente, incluyendo, pero no limitándose a, pelo, piel, uñas, tejidos o fluidos corporales tales como saliva y sangre.

35 Como se usa en el presente documento, el término “contexto de codón” con referencia a una mutación se refiere a la posición del nucleótido dentro de un codón en el que ocurre la mutación. Para los fines de la presente invención, las posiciones de nucleótido dentro de un codón mutado (MC; es decir, un codón que contiene la mutación) se anotan MC-1, MC-2 y MC-3, y se refieren a la primera, segunda y tercera posición de nucleótido, respectivamente, cuando la secuencia del codón se lee de 5' a 3'. Por consiguiente, la frase “determinar el contexto de codón de una mutación” o frase similar significa determinar en qué posición del nucleótido dentro del codón mutado ocurre la mutación, es decir, MC-1, MC-2 o MC-3.

40 Por toda esta memoria, a menos que el contexto requiera lo contrario, las palabras “comprenden”, “comprende” y que “comprende” se entenderá que implica la inclusión de una etapa o elemento indicado o grupo de etapas o elementos pero no la exclusión de ninguna otra etapa o elemento o grupo de etapas o elementos.

45 Por “gen” se quiere decir una unidad de herencia que ocupa un locus específico en un genoma y comprende secuencias reguladoras transcripcionales y/o de traducción y/o una región codificadora y/o secuencias no traducidas (es decir, intrones, secuencias no traducidas 5' y 3').

50 Como se usa en el presente documento, el término “probabilidad” se usa como una medida de si ha ocurrido mutagénesis somática dirigida, si un agente mutagénico particular es una causa de mutagénesis somática dirigida y de si sujetos con ácido nucleico que contiene mutaciones somáticas dirigidas tienen o desarrollarán cáncer basándose en un modelo matemático dado. Una probabilidad incrementada, por ejemplo, puede ser relativa o

absoluta y se puede expresar cualitativamente o cuantitativamente. Por ejemplo, una probabilidad incrementada o riesgo de que un sujeto desarrollará cáncer se puede expresar determinando simplemente el número de mutaciones somáticas dirigidas (como se enseña en el presente documento) y colocando al sujeto de ensayo en una categoría de "probabilidad o riesgo incrementado", basándose en estudios de población previos.

En algunas realizaciones, los métodos comprenden comparar el número o el porcentaje de mutaciones somáticas dirigidas con un número o porcentaje preseleccionado o umbral. Se pueden seleccionar umbrales que proporcionan una capacidad aceptable de predecir el diagnóstico, la probabilidad o el riesgo de pronóstico. En ejemplos ilustrativos, las curvas de la característica operativa del receptor (COR) se calculan trazando en una gráfica el valor de un variable frente a su frecuencia relativa en dos poblaciones en las que una primera población tiene un primer estado o riesgo y una segunda población tiene un segundo estado o riesgo (denominado arbitrariamente, por ejemplo, "estado sano" y "cáncer", o "riesgo bajo" y "riesgo alto").

Probablemente se superpondrá una distribución del número de mutaciones para sujetos con y sin una enfermedad. Bajo tales condiciones, un ensayo no distingue absolutamente un primer estado y un segundo estado con el 100 % de precisión, y el área de superposición indica dónde el ensayo no puede distinguir el primer estado y el segundo estado. Se selecciona un umbral, por encima del cual el ensayo se considera que es "positivo" y por debajo del cual el ensayo se considera que es "negativo". El área bajo la curva (ABC) de COR proporciona el estadístico C, el cual es una medida de la probabilidad de que la medida percibida permitirá correcta identificación de un estado (véase, por ejemplo, Hanley y col., *Radiology* 143:29-36 (1982)). El término "área bajo la curva" o "ABC" se refiere al área bajo la curva de una curva de característica operativa del receptor (COR), las cuales ambas son bien conocidas en la técnica. Las medidas de ABC son útiles para comparar la precisión de un clasificador a través del completo intervalo de datos. Los clasificadores con una ABC mayor tienen una mayor capacidad de clasificar correctamente incógnitas entre dos grupos de interés (por ejemplo, un estado de mutación de un estado sano y un estado de mutación de cáncer). Las curvas COR son útiles para trazar el resultado de una característica particular en la distinción o discriminación entre dos poblaciones (por ejemplo, casos que tienen un cáncer y controles sin el cáncer). Generalmente, los datos de la característica a través de la población entera (por ejemplo, los casos y los controles) se ordenan en orden ascendente basándose en el valor de una característica única. Entonces, para cada valor para esa característica, se calculan los índices verdadero positivo y falso positivo para los datos. La sensibilidad se determina contando el número de casos por encima del valor para esa característica y, a continuación, dividiendo por el número total de los casos. La especificidad se determina contando el número de controles por debajo del valor para esa característica y, a continuación, dividiendo por el número total de controles. Aunque esta definición se refiere a escenarios en los que una característica es elevada en los casos en comparación con los controles, esta definición también se aplica a escenarios en los que una característica es inferior en los casos en comparación con los controles (en dicho escenario, se contarían las muestras por debajo del valor para esa característica). Las curvas COR se pueden generar para una característica única así como para otros resultados únicos, por ejemplo, una combinación de dos o más características se pueden combinar matemáticamente (por ejemplo, añadir, sustraer, multiplicar, etc.) para producir un valor único, y este valor único se puede trazar en una curva COR. Además, cualquier combinación de las características múltiples (por ejemplo, uno u otros más marcadores epigenéticos), en la que la combinación deriva un valor de resultado único, se puede trazar en una curva COR. Estas combinaciones de las características pueden comprender un ensayo. La curva COR es la gráfica de la sensibilidad de un ensayo frente a la especificidad del ensayo, en la que la sensibilidad tradicionalmente se representa sobre el eje vertical y la especificidad tradicionalmente se representa en el eje horizontal. Por tanto, los "valores de ABC de COR" son iguales a la probabilidad de que un clasificador clasificará un caso positivo elegido al azar como mayor que uno negativo elegido al azar. Se puede pensar un valor de ABC de COR como equivalente al ensayo U de Mann-Whitney, el cual ensaya la diferencia media entre las puntuaciones obtenidas en los dos grupos considerados si los grupos son de datos continuos, o al ensayo de Wilcoxon de los rangos.

Alternativamente, o además, los umbrales se pueden establecer obteniendo un resultado de estado de mutación temprano a partir del mismo paciente, con el que se pueden comparar los resultados tardíos. En estas realizaciones, el efecto individual actúa como su propio "grupo control". En otra realización, los umbrales se pueden establecer analizando el número de mutaciones somáticas dirigidas en ácido nucleico del tejido no enfermo o sano de un paciente y comparándolo para analizar el número de mutaciones somáticas dirigidas en el ácido nucleico de tejido enfermo o canceroso.

El término "agente mutagénico" se refiere a un agente que puede causar mutagénesis de ADN. Los agentes mutagénicos incluyen agentes endógenos (es decir, agentes que son endógenos a, o se producen por, la célula en la cual está contenido el ADN) y agentes exógenos (es decir, agentes que son exógenos a, o no producidos por, la célula en la cual está contenido el ADN), e incluyen, por ejemplo, compuestos químicos, proteínas, enzimas, radiación y virus.

Como se usa en el presente documento, un "tipo de mutación" se refiere a la sustitución de nucleótido específica que comprende la mutación, y se selecciona de entre mutaciones C>T, C>A, C>G, G>T, G>A, G>C, A>T, A>C, A>G, T>A, T>C y T>G. Por tanto, por ejemplo, un tipo de mutación de C>T se refiere a una mutación en la que el nucleótido C dirigido o mutado se reemplaza con el nucleótido T sustituyente.

El “ácido nucleico” como se usa en el presente documento se designa ADN, ADNc, ARNm, ARN, ARNr o ARNc. El término generalmente se refiere a polinucleótidos mayores de 30 restos de nucleótido de longitud.

Los términos “paciente” y “sujeto” se usan intercambiamente y se refieren a pacientes y sujetos de ser humano u otro mamífero e incluye cualquier individuo que se desea examinar o tratar usando los métodos de la invención. Sin embargo, se entenderá que “paciente” no implica que los síntomas están presentes. Mamíferos adecuados que caen dentro del alcance de la invención incluyen, pero no se restringen a, seres humanos y otros primates, animales de ganadería (por ejemplo, ovejas, vacas, caballos, burros, cerdos), animales para ensayo de laboratorio (por ejemplo, conejos, ratones, ratas, cobayas, hámsteres), animales de compañía (por ejemplo, gatos, perros) y animales silvestres cautivos (por ejemplo, zorros, ciervos, dingos).

El término “mutación somática” se refiere a una mutación en el ADN de células somáticas (es decir, no células germinales), que ocurre después de la concepción. “Mutagénesis somática”, por lo tanto, se refiere al proceso por el cual ocurren las mutaciones somáticas.

Como se usa en el presente documento, “mutagénesis somática dirigida” se refiere al proceso de la mutagénesis somática resultante de uno o más agentes mutagénicos, en el que ocurre mutagénesis en un nucleótido dirigido dentro de un motivo, el nucleótido dirigido está presente en una posición particular dentro de un codón (por ejemplo, la primera, segunda o tercera posición del codón mutado que se lee de 5' a 3', anotado MC-1, MC-2 y MC-3, respectivamente), y el nucleótido dirigido se muta a un nucleótido sustituyente particular (es decir, la mutación es de un tipo de mutación particular, por ejemplo, C>T, no C>A o C>G). Por tanto, una determinación de que está ocurriendo mutagénesis somática dirigida requiere el análisis del tipo de mutación (por ejemplo, C>T), el motivo en el que ocurre la mutación (por ejemplo, WRC) y el contexto de codón de la mutación, es decir, la posición dentro del codón en la que ocurre la mutación (por ejemplo, MC-1, MC-2 o MC-3). Por lo tanto, “mutágeno somático dirigido” se refiere a la mutación resultante de la mutagénesis somática dirigida.

Como se usa en el presente documento, los términos “tratamiento”, “tratar” y similares, se refieren a la obtención de un efecto farmacológico y/o psicológico deseado. El efecto puede ser profiláctico en términos de prevenir completamente o parcialmente una afección (tal como cáncer) o síntoma de la misma y/o puede ser terapéutico en términos de una cura parcial o completa para una afección y/o efecto adverso atribuible a la afección. “Tratamiento”, como se usa en el presente documento, cubre cualquier tratamiento de la afección en un mamífero, particularmente en un ser humano, e incluye: (a) prevenir de que ocurra la afección en un sujeto que puede estar en riesgo de desarrollar la afección pero aún no se ha diagnosticado que la tiene; (b) inhibir la afección, es decir, detener su desarrollo, y (c) aliviar la afección, es decir, causar regresión de la afección.

Como se usa en el presente documento, “exoma completo” se refiere a todos los exones en el genoma. Por tanto, el análisis de la secuencia de un exoma completo de una célula se refiere al análisis de la secuencia de todos los exones en el genoma de la célula.

2. Agentes mutagénicos implicados en mutagénesis somática

Tanto factores exógenos como endógenos pueden actuar como agentes mutagénicos que causan o juegan un papel en la mutagénesis somática. Factores exógenos incluyen, pero no se limitan a, aflatoxinas, 4-aminobifenilo, ácidos aristolóquicos, compuestos arsénicos, asbestos, azatioprina, benceno, bencidina, berilio y compuestos de berilio, 1,3-butadieno, 1,4-butanodiol dimetilsulfonato (busulfán, Myleran®), cadmio y compuestos de cadmio, clorambucilo, 1-(2-cloroetil)-3-(4-metilciclohexil)-1-nitrosourea (MeCCNU), bis(clorometil) éter y clorometil metil éter de grado técnico, compuestos cromo de hexavalente, breas de alquitrán, alquitranes, emisiones de horno de coque, ciclofosfamida, ciclosporina A, dietilestilbestrol (DES), erionita, óxido de etileno, formaldehído, melfalán, terapia de metoxsaleno con ultravioleta A (PUVA), gas mostaza, 2-naftilamina, neutrones, compuestos de níquel, radón, sílice cristalino (tamaño respirable), radiación solar, hollín, vahos de ácido inorgánico fuerte que contienen ácido sulfúrico, tamoxifeno, 2,3,7,8-tetraclorodibenzo-p-dioxin (TCDD), tiotepa, dióxido de torio, humo de tabaco, cloruro de vinilo, radiación ultravioleta, polvo de madera, radiación X, radiación gamma. Los factores endógenos incluyen, pero no se limitan a, citidina desaminasa inducida por activación (AID), citidina desaminasas similares a polipéptido catalítico de enzima editora de ARNm de apolipoproteína B (APOBEC), y ADN polimerasas propensas a errores, tales como ADN polimerasa-eta.

2.1 AID

La citidina desaminasa inducida por activación (AID) es una enzima importante en la inmunidad adaptativa, implicada en hipermutación somática (HMS) y recombinación de cambio de clase de genes de inmunoglobulina en linfocitos B. AID desencadena HMS por desaminación de citidinas a uracilos (C>U) para diversificar los genes de la región variable de la inmunoglobulina (VDJ) y crear nuevos sitios de unión de antígeno.

Ha sido bien reconocido que los patrones de mutación resultantes de estos procesos de HMS que dan lugar nuevos sitios de unión de antígeno no son aleatorios. Se han identificado agrupaciones y puntos calientes de la actividad mutacional influenciados por las secuencias de bases vecinas, y las propiedades catalíticas y los espectros de

mutación específica resultantes de AID y su implicación en la desaminación de ADN mediada por AID de genes de la región variable de inmunoglobulina recolocados están bien documentados.

El proceso de HMS actualmente se considera que ocurre en dos fases. En la fase 1, el gen que codifica la proteína AID se regula al alza en linfocitos B de centro germinal (Muramatsu y col. (2000) *Cell* 102:553-563). A continuación, AID dirige mutaciones a los pares de base G:C en los puntos calientes complementos inversos GYW/WRCC (en los que Y=C/T, W=A/T, R=A/G; y los nucleótidos subrayados constituyen el par de bases modificado) mediante la desaminación directa de citidina a uracilo (C>U) en las regiones monocatenarias (mc) transcritas del ADN expuesto durante la transcripción (Di Noia y Neuberger (2007) *Annu Rev. Biochem.* 76:1-22; Teng y Papavasiliou (2007) *Annu Rev. Genet.* 41:107-120) AID ocupa la citidina diana antes de la desaminación (Bhutani y col. (2011) *Cell* 146:866-872). Los uracilos en ADN son muy mutagénicos si se dejan no reparados y activan un proceso de reparación por escisión de base (REB) de ADN que implica la uracil ADN glicosilasa (UNG) que causa sitios apririmidínicos (AP) apurínicos, o "abásicos", conduciendo a cortes (*nicks*) de ADNmc (por una actividad de endonucleasa apurínica/apirimidínica, APE) y atrayendo además actividad de reparación de parche de ADN (Peled y col. (2008) *Ann. Rev. Immunol.* 26:481-511). Una UNG desencadena las rutas de BER para eliminar los uracilos, el sitio abásico creado puede, en la replicación y la reparación, ser reemplazado por cualquiera de las bases A, G, C o T.

El patrón de mutación de sesgo de cadena principal asociado con la fase 1 se caracteriza por transiciones C>T y G>A dominantes, y con el número total de mutaciones de G que superan el número de C (Steele (2009) *Mol. Immunol.* 46:305-320). Se ha deducido que el patrón de mutación sesgada de cadena resultante es coherente con el distintivo de la mala incorporación conocida de la ARN polimerasa II de mamífero que copia la cadena de ADN molde que lleva lesiones de AID, uracilos y sitios AP (véase, por ejemplo, Steele (2009) *Mol. Immunol.* 46:305-320).

En la fase 2, las mutaciones se dirigen a los pares de base A:T predominantemente en motivos de punto caliente WA y son claramente sesgados en cadena con mutaciones de A que superan las mutaciones de T en 2 a 3 veces (véase, por ejemplo, Steele (2009) *Mol. Immunol.* 46:305-320). En la fase 2, los mal emparejamientos de G:U reclutan la unión del complejo MSH2-MSH6 del heterodímero de reparación de ADN de errores de emparejamiento, lo cual por turnos recluta la proteína de translesión de la familia Y propensa a error ADN polimerasa-eta que dirige mutaciones en un proceso de reparación de ADN propenso a error de parche corto a ambos sitios WA y algunos otros trechos de secuencia en la región de secuencia diana VDJ.

Diversos estudios sugieren la posibilidad de que los procesos de HMS iniciados por AID aberrantes podrían dar como resultado la conversión de C>U en ADN fuera del ambiente del centro germinal y, por tanto, contribuyen a oncogénesis en otros genes (Beale y col. (2004) *J. Mol. Biol.* 337:585-594; Marusawa H. (2008) *Int. J. Biochem Cell Biol.* 40:1.399-1.402). Se ha encontrado que ocurre actividad similar a HMS en un intervalo de genes tales como *BCL-6* en linfocitos B tonsilares humanos (Yavuz y col. (2002) *Mol. Immunol* 39:485-493), los genes *CD54*, *PIM 1* y *CMYC* en linfomas T (Kotani y col. (2005) *PNAS* 102:4.506-4.511), y *BCL-6* y *C-MYC* en linfomas B (Nielsen y col. (2005) *Oncogene* 24:3.063-3.066). También se ha investigado la actividad de HMS iniciada por AID como una fuente potencial de mutaciones de *TP53* en un número de estudios. En dicho estudio, se encontró que la mutación dirigida en *TP53* en leucemia linfocítica crónica de linfocito B (LLC-B) presenta los rasgos característicos del proceso de HMS (Malcikova y col. (2008) *Molecular Immunology* 45:1.525-9). Aunque el número de mutaciones era bajo para los dos pacientes observados, los datos revelan un sesgo significativo para mutaciones puntuales en pares CG, y una preferencia significativa para los motivos RGYW/WRCY (28 % y 44 % en el primer y segundo paciente, respectivamente). En el segundo paciente, se encontró que 6/8 mutaciones puntuales que afectaban a los pares A:T estaban localizadas en los motivos WA/TW, los cuales son una característica principal del espectro de mutación puntual sencilla de HMS. En el primer paciente se encontró una alta expresión del transcrito de AID, pero no en el segundo que estaba ya mutado en IgVH. Como se muestra en el presente documento y se describe en Lindley y Steele (*ISRN Genomics* (2013) 921418) y Lindley (*Cancer Genet.* (2013) 206(6):222-6), los procesos de mutación similares a HMS sesgado en cadena parecen muy asociados con el cáncer.

Hay también ejemplos de agentes infecciosos que inducen activamente la expresión de AID y dan como resultado un patrón de mutación de *TP53* que es coherente con las características conocidas de la actividad de HMS en los genes de Ig. Ejemplos incluyen virus de la hepatitis C (Machida y col. (2004) *Proc. Natl. Acad. Sci. U.S.A.* 101:4.262-4.267), virus Epstein Barr (Epeldegui y col. (2007) *Mol. Immunol* 44:934-942) y *Helicobacter pylori* (Matsumoto y col. (2007) *Nat. Med.* 13:470-476). AID se ha unido a tumorigénesis de linfocito B y otros cánceres (Honjo y col. (2012) *Adv. Cancer Res.* 2012;113:1-44), y la expresión transgénica de AID causa formación de tumor en ratones (Okazaki y col. (2003) *J. Exp. Med.* 197:1.173-1.181).

2.2 Citidina desaminasas APOBEC

Además de AID, el genoma humano codifica varias citidina desaminasas APOBEC homólogas que se sabe que están implicadas en la inmunidad innata y la edición de ARN (Smith y col. (2012) *Semin. Cell. Dev. Biol.* 23:258-268). En seres humanos, al menos APOBEC1, APOBEC3A, APOBEC3B, APOBEC3C, APOBEC3D, APOBEC3F, APOBEC3G y APOBEC3H están implicados en proporcionar inmunidad innata y/o edición de ARNm celular.

Por ejemplo, APOBEC 1 es responsable de la edición de pre-ARNm de *ApoB*, en el que causa desaminación de la citidina 6666 para cambiar un codón de glutamina en un codón de parada, generando así una forma más corta de ApoB (ApoB48). APOBEC1 también puede desaminar la citidina en ADN (Harris y col. (2002) *Mol. Cell.* 10:1.247-1.253; Petersen-Mahrt y Neuberger (2003) *J. Biol. Chem.* 278:19.583-19.586). Las enzimas APOBEC3 desaminan elementos genéticos móviles (es decir, retroelementos endógenos y virus exógenos), mutando el ADN como una forma de inmunidad innata. Por ejemplo, APOBEC3G actúa sobre VIH y otros retrovirus (por ejemplo, virus de inmunodeficiencia de simio (VIS), virus de anemia infecciosa equina (VAIE), virus de leucemia murina (VLM), y virus foamy (FV)) para mutar el ADN de cadena tipo negativo durante la transcripción inversa. También se ha mostrado que otras enzimas APOBEC3 actúan sobre el VIH y otros retrovirus, así como virus de la hepatitis B, parvovirus y AAV-2 (publicado en Smith y col. (2012) *Sem. Cell Dev. Biol.* 23:258-268).

Como AID, las citidina desaminasas APOBEC han estado implicadas en oncogénesis. Por ejemplo, la expresión transgénica de APOBEC1 causa formación de tumor en ratones (Yamanaka y col. (1995) *PNAS* 92:8.483-8.487); alta expresión de APOBEC3B conduce a mutación somática en genes asociados a tumor (Shinohara y col. (2012) *Scientific Reports* 2:806); APOBEC3B se regula al alza en al menos cáncer de mama, vejiga, cérvix (adenocarcinoma y carcinoma de célula escamosa) y de cabeza y cuello, con un incremento asociado en mutaciones en motivos de APOBEC3B (Burns y col (2013) *Nature* 494:366-370; Burns y col. (2013) *Nature Genetics* 45:977-983); y se ha mostrado que los distintivos de mutación de la enzima APOBEC se extiende en una diversidad de cánceres.

Un estudio que compara preferencias de modificaciones dirigidas por AID, APOBEC 1 y APOBEC3G usando una prueba de mutación bacteriana demostró la importancia crítica de los nucleótidos inmediatamente 5' y 3' de la C dirigida para las especificidades de las citidina desaminasas (Beale y col. (2004) *J. Mol. Biol.* 337:585-594). Mientras que APOBEC3G solamente puede desaminar citidinas sobre ADNmc, APOBEC1 puede editar citidinas sobre ADN o ARN bicatenarios. Se observó que el 79 % de las transiciones en presencia de APOBEC1 estaban asociadas con una T 5', implicando así un motivo de TG/CA para APOBEC1. El motivo de APOBEC3G se sugiere que es CG/CG y/o CC (Beale y col. (2004) *J. Mol. Biol.* 337:585-594; y Rathmore y col (2013). *J. Mol. Biology* 425(22):4.442-54). Otros estudios indican que otras enzimas APOBEC, tales como APOBEC3A, APOBEC3B y APOBEC3F tienen un motivo TC, o un motivo TCW más riguroso (en el que W corresponde a o bien A o T) (Bishop y col. (2004) *Curr. Biol.* 14:1.392-1.396; Thielen y col. (2010) *J. Biol. Chem.* 285:27.753-27.766; Henry y col. (2009) *PLoS One.* 4:e4277; Shinohara y col. (2012) *Scientific Reports* 2:806; Burns y col. (2013) *Nature Genetics* 45:977-983). APOBEC3H se ha sugerido que se dirige a un motivo GA/TC.

3. Métodos para detectar mutagénesis somática dirigida

Como se demuestra en el presente documento, algunos agentes mutagénicos no solamente causan mutagénesis de un nucleótido en uno o más motivos particulares, sin que el motivo y el nucleótido mutado se reconocen dentro del contexto de codón, es decir, el nucleótido mutado está en una posición particular dentro de la estructura del codón, tal como el primer, segundo o tercer nucleótido en el codón mutado (léase de 5' a 3'). También hay una clara preferencia para el reemplazamiento o sustitución de nucleótido. Esta combinación de modificación dirigida específica a motivo y específica a contexto de codón por agentes mutagénicos se califica en el presente documento mutagénesis somática dirigida. A modo de ejemplo no limitante, y como se muestra en la Figura 1, la mutación de A en un motivo WA en la cadena no transcrita de una molécula de ácido nucleico preferiblemente puede ocurrir en la primera posición del codón mutado (MC-1) y es una mutación a C (es decir, A>C). Por tanto, la probabilidad de si ha ocurrido o no mutación somática dirigida de una molécula de ácido nucleico se puede determinar analizando la secuencia de una molécula de ácido nucleico para determinar el contexto de codón de mutaciones de un tipo de mutación (por ejemplo, A>C) en uno o más motivos particulares (por ejemplo, un motivo WA). Si no hay una preferencia codónica en la localización de las mutaciones del tipo de mutación en el motivo (es decir, las mutaciones básicamente están regularmente distribuidas a través de cada posición en los codones), entonces lo más probable es que las mutaciones surjan por cambio y no como resultado de mutaciones somáticas dirigidas por un agente mutagénico. Sin embargo, si hay un porcentaje o número de mutaciones mayor de lo esperado del tipo de mutación en una posición particular en los codones (por ejemplo, sitios MC-1, MC-2 o MC-3) en la molécula de ácido nucleico, entonces esto indica que ha ocurrido mutagénesis somática dirigida o es probable que haya ocurrido.

El "número o porcentaje esperado" de las mutaciones anteriormente descritas es el número o porcentaje de mutaciones esperadas si las mutaciones son independientes de otras mutaciones y el contexto de codón, es decir, la distribución de las mutaciones en cada nucleótido dirigido en cada posición en el codón es básicamente uniforme. Por tanto, por ejemplo, cuando se evalúan las mutaciones que surgen a través de las posiciones o sitios MC-1, MC-2 y MC-3, se esperaría que la mutación de un nucleótido (por ejemplo, A) a uno cualquiera de los otros tres nucleótidos (por ejemplo, G, C o T) en uno cualquiera de los tres sitios (por ejemplo MC-1, MC-2 o MC-3) ocurriera como 1 de cada 9 mutaciones (es decir, 1 de 3 cambios de A a uno cualquiera de G, C o T, y 1 de 3 cambios en cualquier sitio, que es igual a 1 en 9 cambios total) o aproximadamente el 11 % de las veces. Cuando se evalúan las mutaciones que surgen a través de justo dos de las posiciones de nucleótido en el codón mutado, tales como los sitios MC-1 y MC-2, se esperaría que la mutación de un nucleótido (por ejemplo, A) a uno cualquiera de los otros nucleótidos (por ejemplo, G, C o T) en cualquiera de los dos sitios (por ejemplo, MC-1 o MC-2), ocurriera como 1 de cada 6 mutaciones, o aproximadamente el 17 % de las veces (es decir, 1 de 3 cambios de A a uno cualquiera de G, C o T, y

1 de 2 cambios en cualquier sitio, que es igual a 1 de 6 cambios total). Igualmente, cuando se evalúan las mutaciones que surgen a través de solo uno de los sitios (por ejemplo, MC-1), se esperaría que la mutación de un nucleótido (por ejemplo, A) a uno cualquier de los otros nucleótidos (por ejemplo, G, C o T) ocurriera como 1 de cada 3 mutaciones, o aproximadamente el 33 % de las veces.

Esto se ilustra en la Figura 2, en la que la prevalencia de las mutaciones C>T en los sitios MC-1 (es decir, en la primera posición de nucleótido dentro del codón mutado) se evalúa para determinar si ha ocurrido mutación somática dirigida o si las mutaciones observadas surgen al azar. Si la mutación de citosinas a través de los sitios MC-1 y MC-2 en un motivo WRC es aleatoria, entonces se esperaría que el tipo y la posición de las mutaciones se distribuyan regularmente, y que ocurra una mutación C>T en MC-2 una vez cada seis veces (o aproximadamente el 17 %), siendo las otras 5 mutaciones C>A en MC-1, C>A en MC-2, C>G en MC-1, C>G en MC-2, y C>T en MC-1. En el ejemplo particular mostrado en la Figura 2, hay un total de 82 mutaciones de una citosina en los sitios MC-1 o MC-2 en un motivo WRC. Si la mutagénesis era aleatoria, se esperaría que un sexto (o el 17%) de estas fueran mutaciones C>T en los sitios MC-2, equivalente a aproximadamente 14 ocurrencias. Sin embargo, en este ejemplo, hay 72 mutaciones C>T observadas en los sitios MC-2, indicando que ha ocurrido mutagénesis somática dirigida del ácido nucleico.

Generalmente, cuando ocurre mutagénesis somática dirigida como resultado de la actividad de uno o más agentes mutagénicos y se hace una evaluación a través de los tres sitios del codón (por ejemplo, MC-1, MC-2 y MC-3), las mutaciones particulares que están asociadas con el agente mutagénico se observan en al menos o aproximadamente el 20 %, 25 %, 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o más de las veces. Cuando se hace una evaluación a través de dos sitios (por ejemplo, MC-1 y MC-2; MC-1 y MC-3; o MC-2 y MC-3), las mutaciones particulares que están asociadas con el agente mutagénico generalmente se observan en al menos o aproximadamente el 30 %, 35 %, 40 %, 45 %, 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o más de las veces. Cuando se hace una evaluación a través un sitio (por ejemplo, MC-1; MC-2 o MC-3), las mutaciones particulares que están asociadas con el agente mutagénico generalmente se observan en al menos o aproximadamente el 50 %, 55 %, 60 %, 65 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o más de las veces.

Al evaluar el tipo de mutación en un motivo particular (por ejemplo, mutaciones C>T en un motivo WRC) así como el contexto de codón de la mutación (por ejemplo, si la mutación está en un sitio MC-2), se puede hacer una evaluación más precisa de la actividad del agente mutagénico en comparación con cuando se evalúa solamente la mutación en el motivo y el contexto de codón no se factoriza. Por consiguiente, usando los métodos descritos en el presente documento, la probabilidad de que un agente mutagénico particular o un proceso mutagénico, tal como un proceso mutagénico asociado con AID, sea una causa de mutagénesis somática dirigida de una molécula de ácido nucleico se puede evaluar analizando la secuencia de la molécula de ácido nucleico para determinar el contexto de codón de las mutaciones en motivo(s) dirigido(s) por el agente mutagénico o el proceso mutagénico.

3.1 Mutagénesis somática dirigida por AID, APOBEC1, APOBEC3G, APOBEC3H y aflatoxina

Como se describió anteriormente, se sabe que AID se dirige al motivo GYW/WRC, en el que el nucleótido subrayado está mutado. Como se demuestra en el presente documento, hay una preferencia significativa para que la modificación dirigida de la G ocurra en los sitios MC-2, dando como resultado mutaciones G>A. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones G>A en motivos GYW en los sitios MC-2 en la cadena no transcrita de una molécula de ácido nucleico indica que AID es una causa probable de mutagénesis somática dirigida del ácido nucleico, y que AID es activa en las células y/o tejido del cual se obtuvo el ácido nucleico. Como también se demuestra en el presente documento, hay una preferencia significativa para que la modificación dirigida de la C ocurra en los sitios MC-1, dando como resultado mutaciones C>T. Por consiguiente, un número o porcentaje de lo esperado de mutaciones C>T en motivos WRC en los sitios MC-1 en la cadena no transcrita de una molécula de ácido nucleico indica que AID es una causa probable de mutagénesis somática dirigida del ácido nucleico, y que AID es activa en las células y/o tejido del cual se obtiene el ácido nucleico.

Se sabe que APOBEC3G se dirige a los motivos CG/CG. Los estudios descritos en el presente documento demuestran que hay una preferencia significativa para que la modificación dirigida de la G ocurra en los sitios MC-2, dando como resultado mutaciones G>A. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones G>A en motivos CG en los sitios MC-2 en la cadena no transcrita de una molécula de ácido nucleico indica que APOBEC3G es causa probable de mutagénesis somática dirigida del ácido nucleico, y que APOBEC3G es activa en las células y/o tejido del cual se obtiene el ácido nucleico. También hay una preferencia significativa para que la modificación dirigida de la C ocurra en sitios MC-1, dando como resultado mutaciones C>T. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones C>T en motivos CG en los sitios MC-1 en la cadena no transcrita de una molécula de ácido nucleico indica que APOBEC3G es una causa probable de mutagénesis somática dirigida del ácido nucleico, y que APOBEC3G es activa en las células y/o tejido del cual se obtiene el ácido nucleico.

También se sabe que APOBEC3G se dirige a los motivos CC. Los estudios descritos en el presente documento demuestran que hay una preferencia significativa para que la modificación dirigida de la C ocurra en los sitios MC-1, dando como resultado mutaciones C>T. Por consiguiente, un número o porcentaje mayor de lo esperado de

mutaciones C>T en motivos CC en los sitios MC-1 en la cadena no transcrita de una molécula de ácido nucleico indica que APOBEC3G es una causa probable de mutagénesis somática dirigida del ácido nucleico, y que APOBEC3G es activa en las células y/o tejido del cual se obtiene el ácido nucleico.

5 APOBEC1 preferentemente se dirige a los motivos TG/CA en moléculas de ácido nucleico. Además, hay una preferencia significativa para que la modificación dirigida de la C en el motivo CA ocurra en los sitios MC-1, dando como resultado mutaciones C>T. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones C>T en motivos CA en los sitios MC-1 en la cadena no transcrita de una molécula de ácido nucleico indica que APOBEC1 es una causa probable de mutagénesis somática dirigida del ácido nucleico, y que APOBEC1 es activa en las células y/o tejido del cual se obtiene el ácido nucleico. También hay una preferencia para que la modificación dirigida de la G en el motivo TG ocurra en los sitios MC-2, dando como resultado mutaciones G>A. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones G>A en motivos TG en los sitios MC-2 en la cadena no transcrita de una molécula de ácido nucleico indica que APOBEC1 es una causa probable de mutagénesis somática dirigida del ácido nucleico, y que APOBEC1 es activa en las células y/o tejido del cual se obtiene el ácido nucleico.

Se sabe que las mutaciones somáticas en un motivo WA ocurren en la fase 2 del proceso de HMS asociado a AID en linfocitos B de centro germinal, y por tanto son indicativas de procesos de mutaciones asociadas a AID y, por extensión, pueden ser indicativas de la actividad de AID. Como se demuestra en el presente documento, hay una preferencia para que la modificación dirigida de la A ocurra en los sitios MC-2, dando como resultado mutaciones A>T. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones A>T en motivos WA en los sitios MC-2 en la cadena no transcrita de una molécula de ácido nucleico indica que los procesos de mutación somática asociada a AID son activos en las células y/o tejido del cual se obtiene el ácido nucleico, y que AID también puede ser activa en las células y/o tejido. También puede hacer una determinación de que un proceso de mutación asociado a AID es una causa probable de mutagénesis somática dirigida si también se pueden hacer el número o el porcentaje de mutaciones G>A observadas en motivos GYW en los sitios MC-2, o mutaciones C>T en motivos WRC en los sitios MC-1, las cuales son representativas de actividad de AID.

La aflatoxina está asociada a las transversiones de G>T en la tercera posición del codón 249 en *TP53*. Se ha determinado en el presente documento que hay una preferencia de mutación dirigida de la G dentro de un motivo GG, en el que el nucleótido dirigido está en un sitio MC-3. Por consiguiente, un número o porcentaje mayor de lo esperado de mutaciones G>T en motivos GG en los sitios MC-3 en la cadena no transcrita de una molécula de ácido nucleico indica que la aflatoxina es una causa probable de mutagénesis somática dirigida del ácido nucleico. En ejemplos particulares, la aflatoxina es aflatoxina B1. En otros ejemplos, la aflatoxina es aflatoxina B2, G1, G2, M1 o M2.

3.2 Identificación de motivos para otros agentes mutagénicos

Como claramente se demuestra en el presente documento, los agentes mutagénicos se pueden dirigir a un nucleótido en un motivo dentro de un contexto de codón particular. Por tanto, la mutación somática dirigida mediante tales agentes generalmente da como resultado un tipo de mutación (por ejemplo, C>T, y no C>G o C>A), en una posición dentro de la estructura del codón (por ejemplo, MC-1 y no MC-2 o MC-3) y en un motivo (por ejemplo, CG). Al analizar las secuencias de ácido nucleico para el tipo de mutación particular en el motivo y dentro de un contexto de codón particular, como se describió anteriormente, se puede obtener un indicio más preciso de la actividad de un agente mutagénico que si solo la incidencia de las mutaciones en el motivo se examinaba.

Este sesgo para el contexto de codón se puede usar para identificar motivos para otros agentes mutagénicos. Al analizar una molécula de ácido nucleico para la incidencia de mutaciones somáticas de un tipo de mutación conocido por estar asociado con un agente mutagénico (por ejemplo, G>T), y también al evaluar el contexto de codón de las mutaciones y los nucleótidos que flanquean la mutación, se puede identificar el motivo para el agente mutagénico. Cuando ocurre una mutación particular (por ejemplo, G>T) en una posición particular dentro de un codón (por ejemplo, MC-3) más frecuentemente que lo que se daría al azar, es decir, hay una posición de nucleótido preferida en la que ocurre la mutación, entonces es probable que las mutaciones en esta posición ocurran como resultado de mutación somática dirigida por el agente mutagénico. Al analizar los nucleótidos que flanquean la mutación en la posición de nucleótido preferida (por ejemplo, MC-3), se puede identificar cualquier motivo común a las mutaciones y, por tanto, modificado por el agente mutagénico.

Esto se demuestra en el Ejemplo 7 de a continuación. Previamente se ha unido una transversión G>T en la tercera posición del codón 249 en *TP53* a aflatoxina. Cuando el ácido nucleico de una muestra de exoma completo de un sujeto con carcinoma hepatocelular se analizó para mutaciones G>T, se observó que había 9 mutaciones G>T en sitios MC-3, y que cada mutación era coincidente con otra G inmediatamente 5' de la G mutada, sugiriendo que la aflatoxina se dirige a motivos GG, en los que la G dirigida (subrayada) está en un sitio MC-3, para causar mutaciones G>T.

Por tanto, la presente invención también proporciona métodos para identificar un motivo dirigido por un agente mutagénico. Los métodos implican analizar la secuencia de una molécula de ácido nucleico para determinar si un

tipo de mutación asociado con el agente mutagénico ocurre predominantemente en una posición o sitio de un codón (por ejemplo MC-1, MC-2 o MC-3). Si hay una coincidencia de tipo y sitio de mutación, entonces se identifican los nucleótidos que flanquean el nucleótido mutado para identificar un motivo común que incluye el nucleótido mutado. Más específicamente, los métodos implican analizar la secuencia de una molécula de ácido nucleico para identificar mutaciones somáticas de un tipo de mutación conocido por estar asociado con el agente mutagénico, determinando el contexto de codón de las mutaciones para identificar una posición de nucleótido preferida en la que las mutaciones ocurren a una frecuencia mayor que la esperada, e identificar los nucleótidos que flanquean las mutaciones en la posición de nucleótido preferida para identificar un motivo que es común a las mutaciones.

También se puede aplicar un proceso similar cuando aún no se conoce el tipo de mutación asociada con un agente mutagénico. En tales casos, la secuencia de una molécula de ácido nucleico primero se analiza para identificar mutaciones somáticas, y también se identifica cualquier tipo de mutación (por ejemplo, G>T) que ocurre en una posición dentro de un codón (por ejemplo, MC-3) a una frecuencia que es mayor que la esperada si la mutación ocurre al azar (es decir, a una posición de nucleótido preferida). La secuencia que flanquea la mutación en la posición de nucleótido preferida, a continuación, se evalúa para determinar si hay un motivo que es común a la mutación. Si lo hay, este motivo probablemente es la diana del agente mutagénico.

En otros ejemplos, los motivos conocidos de los agentes mutagénicos además se pueden analizar para determinar la preferencia codónica y la preferencia para un tipo de mutación. Las secuencias de ácido nucleico se pueden evaluar como se describe en el presente documento, tal como en el Ejemplo 1, para determinar el contexto de codón y el tipo de mutación asociado con las mutaciones en el motivo para evaluar si hay una preferencia para un tipo de mutación en una posición de nucleótido en el codón. Por ejemplo, se cree que APOBEC3A, APOBEC3B, APOBEC3F y APOBEC3H se dirigen a un motivo TC, o un motivo TCW más riguroso. La secuencia de una o más moléculas de ácido nucleico se pueden analizar para determinar el contexto de codón en el que ocurren las mutaciones en el motivo, es decir, si la C está en MC-1, MC-2 o MC-3 y qué tipo de mutación ocurre, (por ejemplo, C>A, C>T, o C>G). Una vez que el tipo de mutación coincidente, el motivo y el contexto de codón se identifican, este conjunto de criterios, o norma de diagnóstico, se puede usar para determinar más precisamente si APOBEC3A, APOBEC3B, APOBEC3F o APOBEC3H (u otro agente mutagénico) es la causa probable de mutagénesis somática dirigida en una molécula de ácido nucleico y, por tanto, es activa en las células de las cuales se obtiene el ácido nucleico.

Para identificar motivos y/o normas diagnósticas usando los métodos anteriormente descritos, el ácido nucleico que se analiza es generalmente ácido nucleico que se sabe o sospecha que ha estado en contacto con el agente mutagénico o es ácido nucleico que se ha obtenido de las células que se sabe o sospecha que han estado en contacto con el agente mutagénico. Por ejemplo, las células que comprenden el ácido nucleico se pueden exponer *in vitro* al agente mutagénico antes de que se analice el ácido nucleico. En otros ejemplos, el ácido nucleico se puede obtener de tejido o células de sujetos que se sabe que han estado expuestos al agente mutagénico. Se pueden realizar estudios múltiples que usan ejemplos múltiples para validar los resultados.

3.3 Evaluación de la molécula de ácido nucleico

En los métodos de la presente invención se puede usar cualquier método conocido en la técnica para obtener y evaluar la secuencia de una molécula de ácido nucleico. La molécula de ácido nucleico analizada usando los métodos de la presente invención puede ser cualquier molécula de ácido nucleico, aunque generalmente es ADN (incluyendo ADNc). Generalmente, el ácido nucleico es ácido nucleico de mamífero, tal como ácido nucleico humano. El ácido nucleico se puede obtener de cualquier muestra biológica. Por ejemplo, la muestra biológica puede comprender sangre, tejido o células. En algunos ejemplos, la muestra biológica es una biopsia. Además, la muestra puede ser de cualquier parte del cuerpo y puede comprender cualquier tipo de células o tejido, tal como, por ejemplo, tejido o células de pecho, próstata, células de hígado, colon, estómago, pancreáticas, de piel, tiroides, cervicales, linfoides, hematopoyéticas, de vejiga, pulmón, renales, rectales, de ovario, uterinas, y de cabeza o cuello, o células de fluido cerebroespinal. En algunos ejemplos, el ácido nucleico se obtiene de una muestra de célula o tejido de un sujeto sospechoso o en riesgo de tener cáncer, o se obtiene de una muestra de célula o tejido de un sujeto que tiene cáncer.

La molécula de ácido nucleico puede contener una parte o todo un gen, o una parte o todo de dos o más genes, y es la secuencia de este gen o genes la que se analiza según los métodos de la invención. Por ejemplo, las moléculas de ácido nucleico pueden comprender todo o parte de los genes de TP53, PIK3CA, ERBB2, DIRAS3, TET2 u óxido nítrico sintasa (NOS). En algunos casos, la molécula de ácido nucleico comprende el genoma completo o el exoma completo, y es la secuencia del genoma completo o exoma completo la que se analiza en los métodos de la invención.

Cuando se usan los métodos de la presente invención, se ha podido predeterminar la secuencia de la molécula de ácido nucleico. Por ejemplo, la secuencia se puede almacenar en una base de datos u otro medio de almacenamiento, y es esta secuencia la que se analiza según los métodos de la invención. En otros casos, la secuencia de la molécula de ácido nucleico debe ser determinada primero antes del empleo de los métodos de la

invención. En ejemplos particulares, la molécula de ácido nucleico también primero se debe aislar de la muestra biológica.

Los métodos para obtener el ácido nucleico y/o la secuenciación del ácido nucleico son bien conocidos en la técnica, y cualquier método así se puede utilizar para los métodos descritos en el presente documento. En algunos casos, los métodos incluyen amplificación del ácido nucleico aislado antes de secuenciación, y las técnicas de amplificación de ácido nucleico adecuadas son bien conocidas por los expertos en la técnica. Las técnicas de secuenciación de ácido nucleico son bien conocidas en la técnica y se pueden aplicar a genes sencillos o múltiples, o exomas o genomas completos. Estas técnicas incluyen, por ejemplo, métodos de secuenciación capilar que dependen de "secuenciación de Sanger" (Sanger y col. (1977) *Proc. Natl. Acad. Sci. USA* 74:5.463-5.467) (es decir, los métodos que implican la secuenciación de la terminación de cadena), así como técnicas de "secuenciación de nueva generación" que facilitan la secuenciación de miles a millones de moléculas de una vez. Tales métodos incluyen, pero no se limitan a, pirosecuenciación, que hace el uso de luciferasa para leer señales cuando se añaden nucleótidos individuales a moldes de ADN; tecnología de "secuenciación por síntesis" (Illumina), que usa técnicas terminadoras de tinte reversible que añaden un único nucleótido al molde de ADN en cada ciclo; y secuenciación SOLiD™ (Secuenciación por Ligación de Oligonucleótido y Detección; Life Technologies), que secuencia por ligación preferencial de oligonucleótidos de longitud fijada. Estas técnicas de secuenciación de nueva generación son particularmente útiles para secuenciar exomas y genomas completos.

Una vez que se obtiene la secuencia de la molécula de ácido nucleico, a continuación, se identifican mutaciones somáticas puntuales sencillas. Se pueden identificar mutaciones puntuales sencillas comparando la secuencia con una secuencia control. La secuencia control puede ser la secuencia de una molécula de ácido nucleico obtenido de una muestra de un individuo control, tal como un individuo sano que está libre de enfermedad; la secuencia de una molécula de ácido nucleico obtenido de una muestra control, tal como una muestra de tejido no enfermo sano; o puede ser una secuencia consenso que se entiende que no contiene mutaciones somáticas. Además, de identificar las mutaciones puntuales sencillas, se identifica el codón que contiene la mutación y la posición de la mutación dentro del codón (MC-1, MC-2 o MC-3). También se identifican nucleótidos en los codones 5' y 3' flanqueantes para identificar los motivos. Generalmente, para los métodos de la presente invención, se analiza la secuencia de la cadena no transcrita (equivalente a la secuencia de ADNc) de las moléculas de ácido nucleico. En algunos casos, se analiza la secuencia de la cadena transcrita.

La Figura 2 muestra un ejemplo del análisis que se puede realizar sobre el ácido nucleico de una muestra biológica como se describió anteriormente para determinar si APOBEC3G y/o AID son un caso probable de mutagénesis somática. En este ejemplo, la localización de mutaciones puntuales sencillas en la secuencia de ADNc (en la que el codón de inicio "ATG" comprende los 1º, 2º y 3º nucleótidos de la molécula) para la muestra PD3185a se ha identificado y su contexto de codón determinado para evaluar cuánto y qué tipo de mutaciones ocurren en los motivos GYW/WRC, CG/CG y WA en cada posición. A continuación, los datos se tabulan y se aplica análisis estadístico para determinar si las mutaciones surgieron por cambio o como resultado de mutagénesis somática dirigida causada por AID y/o APOBEC3G. En este ejemplo mostrado en la Figura 2, debido a que hay más mutaciones G>A en el motivo GYW en los sitios MC-2 y más mutaciones C>T en el motivo WRC en los sitios MC-1 sobre la cadena no transcrita de lo esperado, es probable que AID sea una causa de mutagénesis somática dirigida en esta molécula de ácido nucleico. Además, debido a que hay más mutaciones G>A en el motivo CG en los sitios MC-2 sobre la cadena no transcrita de lo esperado, es probable que APOBEC3G sea también una causa de mutagénesis somática dirigida en esta molécula de ácido nucleico.

Como se muestra en el presente documento, usando los métodos de la presente invención, solamente se analiza un número pequeño de mutaciones en motivos necesarios de ser analizados para determinar con significancia estadística si ha ocurrido mutagénesis somática dirigida como resultado de la actividad del agente mutagénico particular. En algunos casos, el número de mutaciones en un motivo particular analizado usando los métodos de la presente invención puede ser tan pequeño como 2 mutaciones. Por ejemplo, si se encuentra que un paciente aparentemente sano tiene solamente 2 mutaciones somáticas en el ácido nucleico analizado, y ambas son mutaciones G>A en un motivo GYW en un sitio MC-2, entonces, la probabilidad de que este patrón surgiera por cambio es de 0,04238 ($p < 95\%$, usando una prueba Chi cuadrado, $9-1=9$ df). Alternativamente, la probabilidad de cada una de las mutaciones que ocurren por cambio se puede decir que es 1/9 (es decir, 1/3 de cambio de una mutación G>A, y un 1/3 de cambio de la mutación que está en un sitio MC-2, como se discutió anteriormente), y la probabilidad de que 2 de 2 mutaciones ocurran en este patrón es, por lo tanto, 1/81 (o 0,012346). Sin embargo, como los expertos en la técnica entenderían, la significancia estadística se puede mejorar cuando se analiza más mutaciones en un motivo particular. Por tanto, en algunos casos, el número de mutaciones en un motivo particular analizado usando los métodos de la presente invención puede ser al menos 20. Muchos ejemplos de ácido nucleico de sujetos antes o después del tratamiento tienen 40 o más mutaciones, con alguno que porta hasta 400 o más mutaciones. Por consiguiente, el número de mutaciones en un motivo particular analizado usando los métodos de la presente invención puede ser al menos o aproximadamente 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300 o más.

Todos los materiales imprescindibles y reactivos requeridos para detectar la mutagénesis somática dirigida en un sujeto y además identificar la probabilidad de que un agente mutagénico es la causa de la mutagénesis somática

dirigida, y métodos relacionados como se describe en el presente documento, se pueden reunir juntos en un kit. Por ejemplo, cuando los métodos de la presente invención incluyen primero aislar y/o secuenciar el ácido nucleico a analizar, se prevén kits que comprendan reactivos para facilitar que el aislamiento y/o la secuenciación. Tales reactivos pueden incluir, por ejemplo, cebadores para la amplificación de ADN, polimerasa, dNTP (incluyendo dNTP marcados), controles positivos y negativos, y tampones y soluciones. Tales kits también generalmente comprenderán, de manera adecuada, recipientes distintos para cada reactivo individual. El kit también puede presentar diversos dispositivos, y/o instrucciones impresas para usar el kit.

En algunas realizaciones, los métodos generalmente descritos en el presente documento se realizan, al menos en parte, mediante un sistema de procesamiento, tal como un sistema informático adecuadamente programado. Se puede usar un ordenador autónomo, con el microprocesador que ejecuta un programa de aplicaciones que permiten que se realicen los métodos anteriormente descritos. Alternativamente, los métodos se pueden realizar, al menos en parte, mediante uno o más sistemas de procesamiento que funcionan como parte de una arquitectura distribuida. Por ejemplo, se puede usar un sistema de procesamiento para identificar tipos de mutación, el contexto de codón de una mutación y/o motivos dentro de una o más secuencias de ácido nucleico. En algunos ejemplos, los comandos entrados en el sistema de procesamiento por un usuario ayudan al sistema de procesamiento en la realización de estas determinaciones.

En un ejemplo, un sistema de procesamiento incluye al menos un microprocesador, una memoria, un dispositivo de entrada/salida, tal como un teclado y/o pantalla, y una interfaz externa, interconectados por un bus. La interfaz externa se puede utilizar para conectar el sistema de procesamiento con dispositivos periféricos, tales como una red de comunicaciones, bases de datos o dispositivos de almacenamiento. El microprocesador puede ejecutar instrucciones en la forma de programa de aplicación almacenadas en la memoria para permitir que se realicen los métodos de la presente invención, así como realizar algunos otros procesos requeridos, tales como comunicación con los sistemas del ordenador. El programa de aplicación puede incluir uno o más módulos informáticos, y se pueden ejecutar en un entorno de ejecución adecuado, tal como un entorno del sistema operativo, o similares.

En otro ejemplo, el sistema de procesamiento se puede usar para cargar información de secuencia y otros datos relevantes de bases de datos u otras fuentes. Algoritmos concebidos para ser apropiados para los métodos descritos en el presente documento se pueden aplicar a los datos, tal como se muestra en la Figura 18. En este ejemplo, los datos de entrada [1] y los parámetros de ensayo (tales como motivos a usar) [2] se cargan o entran en el sistema. A continuación, se genera una tabla de sustitución de bases para mutaciones dentro de la región genómica de interés con datos alineados y unidos a mutaciones con datos de contexto de codón y otra información unida a detalles de muestra y secuencia de nucleótidos [3, 4, 5]. La próxima etapa implica la identificación de ocurrencias coincidentes de cada tipo de mutación en cada motivo en cada posición de nucleótido dentro de los codones [6]. Los datos se tabulan para registrar ocurrencias coincidentes de cada tipo de mutación de cada motivo con el contexto de codón [7], incluyendo los grados de probabilidad relativos con niveles de confianza para cada diagnóstico [8]. Los resultados se unen para identificar los agentes mutagénicos (o estructuras moleculares) y los procesos bioquímicos probables de estar implicados en la producción de las mutaciones y la información clínica relevante [8]. Un informe de resultados se genera según la información de solicitud de servicio usada como entrada [9] y se genera un resultado legible [10].

4. Aplicaciones diagnósticas y terapéuticas

Los métodos descritos en el presente documento para detectar si ha ocurrido mutación somática dirigida y determinar la probabilidad de que un agente mutagénico sea una causa de mutagénesis somática de una molécula de ácido nucleico tienen muchas aplicaciones diagnósticas y terapéuticas útiles. Se sabe que la mutagénesis somática está asociada con el desarrollo y la progresión de muchos cánceres. Igualmente, se sabe que algunos agentes mutagénicos están asociados con el desarrollo y la progresión de muchos cánceres. Usando los métodos descritos en el presente documento, se puede determinar la presencia y/o la extensión de la mutagénesis somática dirigida resultante de uno o más agentes mutagénicos, y la identidad del agente mutagénico que es la causa probable de mutagénesis somática. Esto puede facilitar diagnóstico prematuro de cáncer, una determinación de la probabilidad de que un sujeto tenga o desarrollará cáncer, y/o el desarrollo de protocolos terapéuticos o preventivos apropiados. Además, la evaluación en curso de mutaciones somáticas dirigidas atribuibles a uno o más agentes mutagénicos se pueden usar para evaluar si un cáncer está progresando o revirtiendo y/o el éxito o fallo de un régimen de tratamiento. Por ejemplo, un incremento en el número de mutaciones somáticas dirigidas detectadas en el ácido nucleico de una muestra, tal como una biopsia, con el paso del tiempo en el mismo sujeto puede indicar un empeoramiento del cáncer o un fallo de un régimen de tratamiento, mientras una estabilización o reducción en el número de mutaciones puede indicar la remisión de la afección o el éxito de un régimen de tratamiento.

En casos particulares, los métodos de la presente invención se pueden extender al diagnóstico de cáncer en un sujeto o una determinación de la probabilidad de que un sujeto tenga o desarrollará cáncer. Por ejemplo, la probabilidad de que un sujeto tenga o desarrollará cáncer se puede evaluar analizando una molécula de ácido nucleico a partir de una muestra biológica del sujeto para determinar si ha ocurrido mutagénesis somática dirigida por uno o más agentes mutagénicos. Si ha ocurrido mutagénesis somática dirigida, se puede hacer una determinación de que el sujeto es probable que tenga o desarrollare cáncer.

En algunos ejemplos, las normas diagnósticas descritas anteriormente se utilizan para determinar la probabilidad de que se ha dado mutagénesis somática dirigida. Por ejemplo, la mutagénesis somática dirigida se puede detectar cuando el número o el porcentaje de mutaciones G a A observadas en motivos GYW en los sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos WRC en los sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>A observadas en motivos CG en los sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos CG en los sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos CA en los sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>A observadas en motivos GA en los sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>A observadas en motivos TG en los sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones G>T observadas en motivos GG en los sitios MC-3 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; el número o el porcentaje de mutaciones C>T observadas en motivos CC en los sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; o el número o el porcentaje de mutaciones A>G observadas en motivos WA en los sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado, como se describió anteriormente para AID, APOBEC3G, APOBEC3H, APOBEC1 y aflatoxina. En otros ejemplos, se usan las normas diagnósticas determinadas para otros agentes mutagénicos usando los métodos descritos en el presente documento para detectar la ocurrencia de mutación somática dirigida.

En algunos casos, cuando se detectan mutaciones somáticas dirigidas en una muestra que contiene células o tejido de una región o localización particular en un sujeto, tal como tejido o células de pecho, próstata, hígado, colon, estómago, pancreáticas, de piel, tiroides, cervicales, linfoides, hematopoyéticas, de vejiga, pulmón, renales, rectales, de ovario, útero, y de cabeza o cuello, entonces se hace una determinación de que el sujeto tiene o es probable que desarrolle cáncer que implica ese tejido o esas células. Por tanto, por ejemplo, se puede hacer una determinación de que el sujeto tiene o es probable que desarrolle cáncer de mama, próstata, hígado, colon, estómago, pancreático, de piel, tiroides, cervical, linfóide, hematopoyético, de vejiga, pulmón, renal, rectal, de ovario, útero, o de cabeza y cuello.

En ejemplos particulares, si se observa que un agente mutagénico, tal como AID o APOBEC3G, es la causa probable de mutagénesis somática dirigida de ácido nucleico en tejido o células de próstata, entonces al sujeto se le puede diagnosticar cáncer de próstata o se puede determinar que es probable que tenga o desarrolle cáncer de próstata. Igualmente, si se observa que un agente mutagénico es la causa probable de mutagénesis somática dirigida del ácido nucleico en tejido o células de pecho, entonces al sujeto se le puede diagnosticar cáncer de mama o se puede determinar que es probable que tenga o desarrolle cáncer de mama.

La extensión de la mutagénesis somática dirigida mediante el agente mutagénico (es decir, el número de mutaciones somáticas dirigidas atribuibles al agente mutagénico en el ácido nucleico) también se puede usar para ayudar en la determinación de la probabilidad de que el sujeto tenga o desarrollará cáncer, el cáncer está progresando o revirtiendo, y/o el tratamiento está funcionando o no. Generalmente, a mayor número de mutagénesis somática dirigida, mayor probabilidad de que el sujeto tenga o desarrolle cáncer. Además, si hay un incremento en el número de mutaciones somáticas dirigidas con el tiempo en un sujeto, mayor probabilidad de que el cáncer esté progresando y/o el tratamiento haya fallado. Por el contrario, si hay un descenso en el número de mutaciones somáticas dirigidas con el tiempo en un sujeto, mayor probabilidad de que el cáncer esté revirtiendo y/o el tratamiento haya sido exitoso.

Los métodos de la presente invención también se extienden a protocolos terapéuticos o preventivos. En casos en los que un sujeto se determina que es probable que desarrolle cáncer, se pueden diseñar y aplicar protocolos diseñados para reducir esa probabilidad. Por ejemplo, si un sujeto se determina que está en riesgo de desarrollar un cáncer asociado con un agente mutagénico particular, se puede advertir al sujeto que reduzca la exposición a ese agente mutagénico. Por ejemplo, si un sujeto se determina que está en riesgo de desarrollar melanoma, se puede advertir al sujeto que reduzca la exposición a radiación UV. En casos en los que a un sujeto se le ha diagnosticado cáncer o se ha determinado que tiene una alta probabilidad de desarrollar cáncer usando los métodos anteriormente descritos, se puede diseñar un protocolo terapéutico apropiado para el sujeto y administrar. Esto puede incluir, por ejemplo, radioterapia, cirugía, quimioterapia, terapia de ablación de hormona, terapia de proapoptosis y/o inmunoterapia. En algunos ejemplos, se pueden realizar pruebas diagnósticas adicionales para confirmar el diagnóstico antes de la terapia.

Únicamente para ilustrar, las radioterapias incluyen radiación y ondas que inducen daño de ADN, por ejemplo, irradiación γ , rayos X, irradiación UV, microondas, emisiones electrónicas, radioisótopos, y similares. La terapia se puede conseguir irradiando el sitio de tumor localizado con las anteriores formas descritas de radiaciones. Lo más probable es que todos estos factores efectúen un amplio rango de daño de ADN, sobre los precursores de ADN, la replicación y la reparación de ADN, y el ensamblaje y mantenimiento de los cromosomas.

Los intervalos de dosis para rayos X oscilan de dosis diarias de 50 a 200 roentgens durante periodos prolongados de tiempo (3 a 4 semanas), a dosis únicas de 2.000 a 6.000 roentgens. Los intervalos de dosis para radioisótopos varían mucho, y dependen de la semivida del isótopo, la fuerza y el tipo de radiación emitida, y el consumo mediante las células neoplásicas.

5 Ejemplos no limitantes de radioterapias incluyen radioterapia por rayo externo conforme (50 a 100 Grey dados como fracciones durante 4 a 8 semanas), o bien disparo único o fraccionado, braquiterapia de tasa de dosis alta, braquiterapia intersticial permanente, radioisótopos sistémicos (por ejemplo, Estroncio 89). En algunas realizaciones, se puede administrar la radioterapia en combinación con un agente radiosensibilizante. Ejemplos ilustrativos de
10 agentes radiosensibilizantes incluyen pero no se limitan a efaproxiral, etanidazol, fluosol, misonidazol, nimorazol, temoporfin y tirapazamina.

Los agentes quimioterapéuticos se pueden seleccionar de una cualquiera o más de las siguientes categorías:

15 (i) fármacos antiproliferativos/antienoplásicos y sus combinaciones, como se usa en la oncología médica, tal como agentes alquilantes (por ejemplo, cisplatino, carboplatino, ciclofosfamida, mostaza de nitrógeno, melfalán, clorambucilo, busulfán y nitrosoureas); antimetabolitos (por ejemplo, antifolatos tales como fluoropiridinas como 5-fluorouracilo y tegafur, raltitrexed, metotrexato, arabinósido de citosina e hidroxiurea; antibióticos antitumorales (por ejemplo, antraciclinas como adriamicina, bleomicina, doxorubicina, daunomicina, epirubicina, idarubicina, mitomicina-C, dactinomicina y mitramicina); agentes antimitóticos (por ejemplo, alcaloides de vinca como vincristina, vinblastina, vindesina y vinorelbina y taxoides como paclitaxel y docetaxel; e inhibidores de topoisomerasa (por ejemplo, epipodofilotoxinas como etopósido y tenipósido, amsacrina, topotecán y camptotecina);

20 (ii) agentes citostáticos tales como antioestrógenos (por ejemplo, tamoxifeno, toremifeno, raloxifeno, droloxifeno y yodoxifeno), reguladores a la baja del receptor de estrógeno (por ejemplo, fulvestrant), antiandrógenos (por ejemplo, bicalutamida, flutamida, nilutamida y acetato de ciproterona), antagonistas de UH o agonistas de LHRH (por ejemplo, goserelina, leuprorelina y buserelina), progestógenos (por ejemplo, acetato de megestrol), inhibidores de aromatasas (por ejemplo como anastrozol, letrozol, vorazol y exemestano) e inhibidores de 5 α -reductasa tal como finasterida;

25 (iii) agentes que inhiben la invasión de célula cancerosa (por ejemplo, inhibidores de metaloproteinasas como marimastat e inhibidores de la función del receptor del activador del plasminógeno tipo uroquinasa);

30 (iv) inhibidores de la función del factor de crecimiento, por ejemplo, tales inhibidores incluyen anticuerpos del factor de crecimiento, anticuerpos del receptor del factor de crecimiento (por ejemplo, el anticuerpo anti-erbb2 trastuzumab [Herceptin™] y el anticuerpo anti-erbb1 cetuximab [C225]), inhibidores de la farnesil transferasa, inhibidores de MEK, inhibidores de la tirosina quinasa e inhibidores de la serina/treonina quinasa, por ejemplo, otros inhibidores de la familia del factor de crecimiento epidérmico (por ejemplo, otros inhibidores de la tirosina quinasa de la familia de EGFR tales como N-(3-cloro-4-fluorofenil)-7-metoxi-6-(3-morfolinopropoxi)quinazolin-4- -amina (gefitinib, AZD1839), N-(3-etinilfenil)-6,7-bis(2-metoxietoxi)quinazolin-4-amina (erlotinib, OSI-774) y 6-acrilamido-N-(3-cloro-4-fluorofenil)-7-(3-morfolinopropoxi)quinazolin-4-amina (CI 1033)), por ejemplo, inhibidores de la familia del factor de crecimiento derivado de plaqueta y, por ejemplo, inhibidores de la familia del factor de crecimiento de hepatocito;

35 (v) agentes antiangiogénicos tales como los que inhiben los efectos del factor de crecimiento endotelial vascular (por ejemplo, el anticuerpo anti-factor de crecimiento de célula endotelial vascular bevacizumab [Avastin™], compuestos tales como los descritos en las Solicitudes de Patente Internacional WO 97/22596, WO 97/30035, WO 97/32856 y WO 98/13354) y compuestos que funcionan mediante otros mecanismos (por ejemplo, linomida, inhibidores de la función de la integrina $\alpha\text{v}\beta\text{3}$ y angiostatina);

40 (vi) agentes de daño vascular tales como Combretastatina A4 y compuestos descritos en las Solicitudes de Patente internacional WO 99/02166, WO 00/40529, WO 00/41669, WO 01/92224, WO 02/04434 y WO 02/08213; (vii) terapias antisentido, por ejemplo, aquellas que están dirigidas a las dianas anteriormente enumeradas, tales como ISIS 2503, un antisentido anti-ras; y

45 (viii) propuestas de terapia génica, que incluyen, por ejemplo, planteamientos para reemplazar genes aberrantes tales como propuestas de p53 aberrante o GDEPT aberrante (terapia con profármaco enzimático dirigida a gen) tales como los que usan citosina desaminasa, timidina quinasa o una enzima nitroreductasa bacteriana y propuestas para incrementar la tolerancia del paciente a quimioterapia o radioterapia tal como terapia génica de resistencia a fármaco múltiple.

50 Propuestas de inmunoterapia, incluyen, por ejemplo, propuestas *ex vivo* e *in vivo* para incrementar la inmunogenicidad de las células tumorales del paciente, tal como transfección con citoquinas tales como interleuquina 2, interleuquina 4 o factor estimulante de colonia de granulocito-macrófago, propuestas para disminuir la anergia de linfocito T, propuestas que usan células inmunes transfectadas tales como células dendríticas

transfectadas por citoquina, propuestas que usan líneas de célula tumoral transfectada por citoquina y propuestas que usan anticuerpos antiidiotípicos. Estas propuestas generalmente dependen del uso de células efectoras inmunes y moléculas para seleccionar y destruir células cancerosas. El efector inmune puede ser, por ejemplo, un anticuerpo específico para algún marcador sobre la superficie de una célula maligna. El anticuerpo solo puede servir como un efector de terapia o puede reclutar otras células para realmente facilitar la muerte celular. El anticuerpo también se puede conjugar a un fármaco o toxina (compuestos quimioterapéuticos, radionúclido, cadena A de ricina, toxina del cólera, toxina pertussis, etc.) y sirven simplemente como agente dirigido. Alternativamente, el efector puede ser un linfocito que lleva una molécula de superficie que interactúa, o bien directamente o indirectamente, con una diana de célula maligna. Diversas células efectoras incluyen linfocitos T citotóxicos y células NK.

Ejemplos de otras terapias de cáncer incluyen fototerapia, crioterapia, terapia con toxina o terapia proapoptosis. Un experto en la técnica sabrá que esta lista no es exhaustiva de los tipos de modalidades de tratamiento disponibles para cáncer y otras lesiones hiperplásicas.

En algunos casos, en los que se determina la identidad probable del agente mutagénico que causa las mutaciones somáticas dirigidas, la terapia o las medidas preventivas pueden incluir la administración al sujeto de un inhibidor de ese agente mutagénico. Los inhibidores pueden incluir, por ejemplo, ARNip, miARN, antagonistas proteicos (por ejemplo, mutantes negativos dominantes del agente mutagénico), pequeños inhibidores de molécula, anticuerpos y fragmentos de los mismos. Por ejemplo, los ARNpi comercialmente disponibles y los anticuerpos específicos para citidina desaminasas APOBEC y AID están muy disponibles y son conocidos por los expertos en la técnica. Otros ejemplos de inhibidores de APOBEC3G incluyen las pequeñas moléculas descritas por Li y col. (*ACS Chem. Biol.* (2012) 7(3):506-517), muchas de las cuales contienen restos de catecol, los cuales se sabe que son sulfhidrilo reactivos después de la oxidación a la ortoquinona. Los inhibidores de APOBEC1 también incluyen, pero no se limitan a, polipéptidos de APOBEC1 mutantes dominantes negativos, tales como el mutante mu1 (H61K/C93S/C96S) (Oka y col. (1997) *J. Biol. Chem.* 272, 1.456-1.460).

Generalmente, los agentes terapéuticos se administrarán en las composiciones farmacéuticas junto con un vehículo farmacéuticamente aceptable y en una cantidad eficaz para conseguir su propósito previsto. La dosis de compuestos activos administrados a un sujeto debería ser suficiente para conseguir una respuesta beneficiosa en el sujeto con el tiempo tal como una reducción en, o alivio de, los síntomas de cáncer y/o la reducción, regresión o eliminación de tumores o células cancerosas. La cantidad del(de los) compuesto(s) farmacéuticamente activo(s) a administrar puede depender del sujeto a tratar incluso de la edad, sexo, peso y estado de salud general del mismo. A este respecto, las cantidades precisas del(de los) compuesto(s) activo(s) para la administración dependerá de la opinión del médico, y los expertos en la técnica pueden determinar fácilmente dosis adecuadas de los agentes terapéuticos y regímenes de tratamiento adecuados sin experimentación indebida.

Para que la invención se pueda entender fácilmente y poner en efecto práctico, a continuación, se describirán realizaciones particulares preferidas por medio de los siguientes ejemplos no limitantes.

Ejemplos

Ejemplo 1

Análisis de mutaciones somáticas de TP53 en cáncer de mama

Se evaluó la frecuencia y el contexto de las mutaciones somáticas en el gen de TP53 en los cánceres de mama accediendo a la base de datos de TP53 IARC y extrayendo los datos específicos para cáncer de mama. El número de mutaciones puntuales en este conjunto de datos era grande (N=2.514). La mayoría de las mutaciones eran mutaciones puntuales sencillas, predominantemente centradas en la región de unión a ADN (codones aproximadamente 130 a 300) de TP53. Solamente una fracción menor de las muestras llevaban una mutación exónica en TP53. Se asumió que hay solamente ligeras variaciones debido a la composición de bases de TP53, y no se hicieron correcciones. La selección de diversos criterios facilitó la construcción y el análisis de todos los tipos de mutaciones con contexto de secuencia flanqueante 5' y 3' en relación con la secuencia de exón de TP53 no mutada (y en algunos casos secuencia intrónica). Esto facilitó el desarrollo de las distribuciones de frecuencia de diversos tipos de mutación (por ejemplo, A a G) frente a posición de nucleótido y codón a través de las regiones de interés.

Se analizaron las secuencias de los transcritos de ADNc (es decir, el mismo contexto de secuencia que las cadenas no transcritas). Se usaron transcritos de ADNc ya que estos están públicamente disponibles en las bases de datos COSMIC y Ensembl con fines de extracción y análisis. Usando estos transcritos, el contexto de secuencia alrededor de cada mutación se analizó para las mutaciones en el motivo de AID (GYW/WRC), motivo de APOBEC1 (TG/CA) y motivo de APOBEC3G (CG/CG), así como el motivo WA, que es representativo de sitios potenciales para las mutaciones en los pares de bases A:T en la fase II del proceso de HMS (y, por tanto, asociado con la actividad de AID). Las mutaciones se evaluaron en relación con sus posiciones en un codón mutado.

La Figura 1 muestra un ejemplo de una secuencia mutada en la "región de interés" definida para este análisis. La región de interés incluye 9 nucleótidos que abarcan el codón mutado, el codón 5-prima (5') flanqueante y el codón 3-

prima (3'). Las respectivas posiciones de los nucleótidos en la secuencia de codón mutado (MC) se anotan como MC-1, MC-2 y MC-3 (léase de 5' a 3'). Las respectivas posiciones de los nucleótidos (N) en el codón 5' flanqueante se anotan como 5'N1, 5'N2 y 5'N3 respectivamente (también léase de 5' a 3'). Igualmente, las posiciones de los nucleótidos en el codón 3' flanqueante se anotan como 3'N1, 3'N2 y 3'N3 respectivamente. En el ejemplo mostrado para una mutación puntual A a C (A>C), una A en un sitio MC-1 en la cadena no transcrita (NTS) está mutada a una C en la cadena no transcrita replicada (NTS'). La mutación de A en el codón mutado está asociada con una G en la posición 5'N3. Esta se anota como "S ·· A" (en la que S es una G o C). Esta anotación se usa sin reparar en la localización de una mutación dentro del codón mutado.

El contexto de codón y la frecuencia de las 2.514 mutaciones somáticas en el gen *TP53* del conjunto de datos de cáncer de mama reunidos se muestra en la Tabla 1. Como se indicó anteriormente, MC-1, MC-2 y MC-3 se refieren a la posición de las mutaciones dentro del codón mutado (MC). Estas se leen de 5' a 3' a partir de la cadena no transcrita. Para determinar si el contexto de codón era importante para cada tipo de mutación, se usó una prueba Chi cuadrado para ensayar la significancia estadística frente a un límite al nivel $P < 0,01$ (2 DF).

Tabla 1:

Mutación	Localización de todas las mutaciones dentro del codón mutado						
	MC-1 (p, 2df)		MC-2 (p, 2df)		MC-3 (p, 2df)		Total
A>T	30		29		0	($p < 0,001$)	
A>C	11		30	($p < 0,001$)	2	($p < 0,01$)	43
A>G	64		194	($p < 0,001$)	11	($p < 0,001$)	269
Total de A	105		253		13		371
T>A	30		29		8		67
T>C	48		64	($p < 0,01$)	18	($p < 0,001$)	130
T>G	23		44		19		86
Total de T	101		137		45		283
C>A	23		22		23		68
C>T	397	($p < 0,001$)	118	($p < 0,001$)	78	($p < 0,001$)	593
C>G	42		25		22		89
Total de C	462		165		123		750
G>A	203	($p < 0,001$)	505	($p < 0,001$)	87	($p < 0,001$)	795
G>T	69		87		37	($p < 0,01$)	193
G>C	35		67	($p < 0,001$)	20	($p < 0,01$)	122
Total de G	307		659		144		1.110
Total	975		1.214		325		2.514

Se observó que había muchas más transiciones (es decir, A < > G o C < > T) que transversiones (es decir, A o G < > C o T). Como resultado, el patrón de mutación muestra sesgos de cadena significativos en los que las mutaciones de A superan las mutaciones de T (371/283=1,3), y las mutaciones de G superan las mutaciones de C (1.110/750=1,5). Esto concuerda con el trabajo previo que muestra patrones de sesgo de cadena similares para los procesos de HMS en regiones VDJ de genes de Ig, así como datos de mutación del gen de la proteína quinasa a través del genoma completo para un rango de cánceres no linfoides que incluyen el cáncer de mama (Steele y Lindley (2010) *DNA Repair* 9:600-603). El patrón de sesgo de cadena también concuerda con los datos de mutación tomados de pacientes con leucemia linfocítica crónica de linfocito B (Malcikova y col. (2008) *Molecular Immunology* 45:1.525-9).

El conjunto de datos reunidos mostrados en la Tabla 1 también reveló patrones de preferencia codónica de mutación significativos no previamente publicados. Los sesgos de contexto de codón más significativos eran para las transiciones C>T ($P < 0,001$, 2DF), G>A ($P < 0,001$, 2DF) y A>G ($P < 0,001$, 2DF), las cuales se sabe que dan como resultado los patrones de sesgo de cadena principales asociados con procesos de HMS.

Se encontró que 397/593 (66,9 %) de todas las transiciones C>T ocurrieron en un sitio MC-1, y 397/750 (52,9 %) de todas las mutaciones de C (es decir, C>A/G/T) eran transiciones C>T en un sitio MC-1. Por el contrario, 505/795 (63,5 %) de todas las transiciones G>A ocurrieron en un sitio MC-2, y 505/1110 (45,5 %) de todas las mutaciones de G (es decir, G>A/C/T) eran transiciones G>A en un sitio MC-2. Si las mutaciones ocurren al azar e independientemente de la estructura de codón, se espera que solamente 1 de 9 (o alrededor del 11,1 %) mutaciones ocurriría en un sitio particular (es decir, MC-1, MC-2 o MC-3) para cada uno de los 3 tipos diferentes de mutación de un nucleótido particular.

Para las transiciones A>G, 194/269 (72,1 %) de todas las transiciones A>G ocurrieron en un sitio MC-2, y 194/371 (52,3 %) de todas las mutaciones de A (es decir, A>C/G/T) eran transiciones A>G en un sitio MC-2.

Los datos en la Tabla 1 también soportan la expectativa de la selección para mutaciones de aminoácido en el gen *TP53* ya que el número de mutaciones en el MC-3 era significativamente menor que en las posiciones MC-1 o MC-2 para cada una de las transiciones C>T, G>A y A>G. Para ARN, la ruta de la degradación de ARNm mediada por mutaciones terminadoras (NMD) es un sistema de vigilancia celular conocido que depende de la información de

contexto de codón para permitir que la célula identifique y disponga de productos génicos defectuosos que contienen mutaciones “terminadoras (*nonsense*)” o señales de PARADA (UAG, UGA y UAA) que podrían parar prematuramente la traducción. El resultado es selección para mutaciones de aminoácido (*missense*) en *TP53*. Los datos también son coherentes con otro estudio previo que informa de una tendencia para mutabilidad mayor que la esperada para las posiciones de codón MC-1 y MC-2 en regiones determinadoras complementarias de genes de la región variable (V) de Ig (Shapiro y col. (2002) *J. Immunology* 168:2.302-2.306).

El análisis también reveló una preferencia estadística altamente significativa para transiciones C>T para que ocurran en sitios MC-1 ($P<0,001$, 2DF), y para transiciones G>A para que ocurran en sitios MC-2 ($P<0,001$, 2DF). Como citidinas sobre la TS o la NTS de ADNbc en una “burbuja de transcripción” abierta son ambas capaces de someterse a desaminación, los datos soportan la conclusión de que los mecanismos moleculares implicados son capaces de leer dentro del marco (*in-frame*), y distinguir entre las citidinas sobre la TS y la NTS.

La Tabla 2 muestra el contexto de codón de las 2.514 mutaciones somáticas para el conjunto de datos de cáncer de mama de *TP53* que ocurren en los motivos de AID, APOBEC1 y APOBEC3G, así como los motivos WA. Se usó una prueba Chi cuadrado para determinar la significancia estadística frente a un límite al nivel $P<0,01$ (2 DF).

Si las mutaciones ocurren independientemente de la estructura del codón 5', y no se hace corrección para la composición de bases, entonces se espera que alrededor de un tercio de cada tipo de mutación estará localizada en un sitio MC-1, MC-2 o MC-3. Igualmente, se espera que solamente una novena parte (11,1 %) de todas las mutaciones de un nucleótido único estará localizada en un sitio MC-1, MC-2 o MC-3. Se encontró que el sesgo de contexto de codón para las transiciones en los motivos clave asociados con la actividad de AID, APOBEC1 y APOBEC3G era incluso más estadísticamente significativo que lo que se encontró en el conjunto de datos reunidos mostrados en la Tabla 1

Tabla 2

Mutación	MC-1 (p, 2df)	MC-2 (p, 2df)	MC-3 (p, 2df)	Total
Sitios <u>GYW/WRC</u> (AID)				
G>A	9 ($p<0,0001$)	185 ($p<0,0001$)	6 ($p<0,0001$)	200
G>T	3	32 ($p<0,0001$)	2	37
G>C	0	10	1	11
C>A	7	9	0	16
C>T	106 ($p<0,0001$)	13 ($p<0,0001$)	13 ($p<0,0001$)	132
C>G	3	2	15 ($p<0,01$)	20
Sitios <u>CG/CG</u> (APOBEC3G)				
G>A	46 ($p<0,0001$)	358 ($p<0,0001$)	3 ($p<0,0001$)	407
G>T	19	24	2 ($p<0,01$)	45
G>C	10	43 ($p<0,0001$)	0 ($p<0,0001$)	53
C>A	6	0	7	13
C>T	240 ($p<0,0001$)	6 ($p<0,0001$)	2 ($p<0,0001$)	248
C>G	20 ($p<0,01$)	1	6	27
Sitios <u>TG/CA</u> (APOBEC1)				
G>A	46	62	47	155
G>T	17	38 ($p<0,01$)	9	64
G>C	7	4	6	17
C>A	6	2	6	14
C>T	93 ($p<0,0001$)	16 ($p<0,0001$)	51	160
C>G	5	5	4	14
Sitios <u>WA</u>				
A>T	5	5	0	10
A>C	0	15 ($p<0,0001$)	1 ($p<0,01$)	16
A>G	7 ($p<0,0001$)	128 ($p<0,0001$)	6 ($p<0,0001$)	141

Para los motivos GYW ligados a la actividad de AID, 185/200 (92,5 %) de todas las transiciones G>A ocurrieron en un sitio MC-2, y 185/248 (74,6 %) de todas las mutaciones en los sitios GYW (es decir, G>A/C/T) eran transiciones G>A en un sitio MC-2. Por el contrario, en los sitios WRC, 106/132 (80,3 %) de todas las transiciones C>T ocurrieron en un sitio MC-1, y 106/168 (63,1 %) de todas las mutaciones de C (es decir, C>A/G/T) eran transiciones C>T en un sitio MC-1.

En los motivos CG ligados a la actividad de APOBEC3G, 358/407 (87,7 %) de todas las transiciones G>A ocurrieron en un sitio MC-2, y 358/505 (70,9 %) de todas las mutaciones en los sitios CG (es decir, G>A/C/T) eran transiciones G a A en un sitio MC-2. Por el contrario, en los sitios CG, 240/248 (96,8 %) de todas las transiciones C>T ocurrieron en un sitio MC-1, y 240/288 (83,3 %) de todas las mutaciones de C (es decir, C>A/G/T) eran transiciones C>T en un sitio MC-1.

Para los motivos TG/CA ligados a la actividad de APOBEC1, el sesgo del contexto de codón no era estadísticamente significativo. En los sitios CA, 93/160 (58,1 %) de las transiciones C>T ocurrieron en un sitio MC-1, y 93/188 (49,5 %) de todas las mutaciones de C (es decir, C>A/G/T) eran transiciones C>T en un sitio MC-1. Solamente 62/155 (40,0 %) de todas las transiciones G>A en un sitio TG ocurrieron en un sitio MC-2, y 62/136 (45,6 %) de todas las mutaciones de G (es decir, G>A/C/T) en los sitios TG eran transiciones G>A en un sitio MC-2.

Otra característica de los patrones de preferencia codónica observados en los motivos clave mostrados en la Tabla 2 es que la mayoría de todas las mutaciones de G para cada uno de los motivos para AID, APOBEC1 y APOBEC3G preferentemente ocurren en un sitio MC-2. En comparación, la mayoría de las mutaciones de C para cada uno de los motivos ocurrieron en un sitio diana de MC-1. Esto implica que un mecanismo de detección dentro del marco está implicado en el nivel de ADN durante la iniciación de transcripción, y esto es capaz de distinguir entre citidinas en la NTS y aquellas en la TS en el contexto de una “burbuja de transcripción”.

Para las transiciones A>G en los sitios WA, 128/141 (90,8 %) ocurrieron en un sitio MC-2, y 128/167 (76,6 %) de todas las mutaciones de A en los sitios WA (es decir, A>C/G/T) eran transiciones G>A en un sitio MC-2. Ya que se reconoce un nivel elevado de mutaciones A>G en sitios WA como un rasgo característico de la actividad de la HMS y diagnóstico de la intervención de un intermedio molde de ARN, este descubrimiento soporta una predicción de que los procesos de mutación iniciados por AID endógenos son activos en al menos muchas de las muestras en el conjunto de datos.

La Tabla 3 muestra el contexto de codón de mutaciones que ocurren en motivos clave asociados con AID, APOBEC1 y APOBEC3G y localizados conjuntamente con un nucleótido fuerte (S=G/C) en la posición 5'N3. La anotación “S ·· M” (en la que M es el nucleótido mutado A, G, C o T) se usa para indicar la presencia de un nucleótido “S” en la posición 5'N3 que flanquea el codón mutado, y con la diana nucleótido mutada en una cualquiera de las posiciones MC-1, MC-2 o MC-3. Si las mutaciones ocurren independientemente de la estructura del codón 5', y no se hace corrección para la composición de bases, entonces se espera que solamente la mitad de las mutaciones en cada uno de los motivos estarán localizadas conjuntamente con un S en la posición 5'N3.

Tabla 2.

Mutación	MC-1	MC-2	MC-3
Mutaciones en sitios <u>GYW/WRC</u> (AID) y S ·· <u>G/C</u>			
G>A	2/9	184/185 (99,5 %)	4/6
G>T	2/3	31/32	2/2
G>C	0/0	10/10	1/1
C>A	5/7	0/9*	0/0
C>T	102/106 (96,2 %)	0/13*	9/13
C>G	3/3	0/2*	8/15
Mutaciones en sitios <u>CG/CG</u> (APOBEC3G) y S ·· <u>G/C</u>			
G>A	46/46*	352/358 (98,3 %)	2/3
G>T	19/19*	24/24	2/2
G>C	10/10*	41/43	0/0
C>A	6/6	0/0	7/7
C>T	239/240 (99,6 %)	6/6	2/2
C>G	20/20	1/1	4/6
Mutaciones en sitios <u>TG/CA</u> (APOBEC 1) y S ·· <u>G/C</u>			
G>A	0/46*	36/62 (58,1 %)	34/47
G>T	0/17*	27/38	9/9
G>C	0/7*	3/4	5/6
C>A	4/6	2/2	4/6
C>T	52/93 (55,9 %)	10/16	45/51
C>G	5/5	5/5	2/4
Mutaciones en sitios <u>WA</u> y S ·· <u>A</u>			
A>T	0/5*	4/5	0/0
A>C	0/0*	11/15	0/1
A>G	0/8*	121/127 (95,3 %)	6/6
* Es imposible para un nucleótido en la posición 5'N3 ser ambas “S” y <u>WA</u> o <u>TG</u> para mutaciones en la posición MC-1. Igualmente, los nucleótidos en la posición 5'N3 no pueden ser “S” y <u>WRC</u> para mutaciones en la posición MC-2, y todas las mutaciones en un sitio <u>CG</u> en la posición MC-1 tienen una “S” en la posición 5'N3.			

El análisis reveló una unión inesperadamente alta entre los sitios S ·· M y las transiciones en los motivos asociados con la actividad de AID, APOBEC3G y en sitios WA, pero no sitios de APOBEC1. Para los motivos GYW/WRC asociados con la actividad de AID, 184/185 (99,5 %) de todas las transiciones G>A en la posición MC-2 tenían una S presente en la posición 5'N3, y 102/106 (96,2 %) de todas las transiciones C>T en la posición MC-1 tenían una S

presente en la posición 5'N3. Para los motivos CG/CG asociados con la actividad de APOBEC3G, 352/358 (98,3 %) de todas las transiciones G>A en la posición MC-2 tenían una S presente en la posición 5'N3, y 239/240 (99,6 %) de todas las transiciones C a T en la posición MC-1 tenían una S presente en la posición 5'N3. Para los motivos TG/CA asociados con la actividad de APOBEC1, los resultados no eran estadísticamente significativos. Solamente 36/62 (58,1 %) de las transiciones G>A en un sitio MC-2 tenían una S presente en la posición 5'N3, y 52/93 (55,9 %) de las transiciones C a T en la posición MC-1 tenían una S presente en la posición 5'N3. Para los sitios WA, 121/127 (95,3 %) de las transiciones A a G en un sitio MC-2 tenían una S presente en la posición 5'N3.

Los datos en la Tabla 3 también revelan una proporción inesperadamente alta de alguna de las transversiones en los motivos seleccionados que están localizados conjuntamente con una S en la posición 5'N3. En particular, hay una unión mayor que la esperada entre las mutaciones G a T/C en sitios diana GYW o CG y un sitio S ·· G. Para todas las transiciones y transversiones de G que ocurren en una diana MC-2 para los motivos seleccionados de AID, APOBEC3G y WA, un 778/799 (97,4 %) altamente significativo están localizados conjuntamente con un S ·· G. Igualmente, 375/382 (98,2 %) de todas las transiciones y transversiones de C que ocurren en un sitio diana MC-1 para los motivos seleccionados de AID y APOBEC3G, están localizados conjuntamente con un S ·· C.

Por lo tanto, la locación conjunta de un S ·· M (M=A/G/C/T) parece ser una parte integral de la unión de contacto directo y mecanismos sensores del marco de lectura de codón asociados con la actividad desaminasa de AID y APOBEC3G, así como el(los) mecanismo(s) mutágeno(s) que actúa(n) en sitios los WA.

Ejemplo 2

Desarrollo de las normas diagnósticas para predecir la actividad de AID, APOBEC1 o APOBEC3G

Los patrones de preferencia codónica observados para la mutación en los motivos de AID, APOBEC1, APOBEC3G y WA (anteriormente descritos) se usaron para generar las siguientes "normas" o criterios diagnósticos para su uso en la predicción de si la mutación somática dirigida (MSD) de una molécula de ácido nucleico está ocurriendo como resultado de la actividad de AID, APOBEC1 y/o APOBEC3G:

Un número mayor de lo esperado de mutaciones G>A del motivo GYW (AID) en sitios MC-2 está asociado con la actividad desaminasa de AID sobre la cadena transcrita.

Un número mayor de lo esperado de mutaciones C>T del motivo WRC (AID) en sitios MC-1 está asociado con la actividad desaminasa de AID sobre la cadena transcrita.

Un número mayor de lo esperado de mutaciones G>A del motivo CG (APOBEC3G) en sitios MC-2 está asociado con la actividad de APOBEC3G.

Un número mayor de lo esperado de mutaciones C>T del motivo CG (APOBEC3G) en sitios MC-1 está asociado con la actividad de APOBEC3G

Un número mayor de lo esperado de mutaciones A>G del motivo WA en MC-2 es un indicio de procesos de mutación ligados a AID y, por tanto, a la actividad de AID.

Cuando se aplican estas normas, se asume que el conjunto de mutaciones de cada nucleótido son independientes unas de otras, y que si los agentes mutagénicos no están presentes, la distribución de las mutaciones de cada nucleótido en cada uno de los sitios de codón MC-1 y MC-2 será distribuida al azar para las mutaciones de A, G, C o T.

La Figura 2 muestra un ejemplo de cómo se pueden usar los criterios diagnósticos anteriores para determinar la probabilidad de que la distribución de la mutación de preferencia codónica surgiera por cambio o por mutación somática dirigida por AID o APOBEC3G. Para cada una de las categorías diagnósticas anteriormente seleccionadas, el número de las mutaciones observadas (O) y las esperadas (E) se tabulan en forma de Tabla. Para cada una de las categorías diagnósticas, el número de mutaciones esperadas (E) se calcula usando el número total de mutaciones probables de surgir a través de los sitios MC-1 y MC-2 para cada uno de los 3 tipos posibles de la mutación de un nucleótido particular si las mutaciones son aleatorias. (Cuando se analizan mutaciones del gen de TP53, como se muestra en la Figura 2, las mutaciones que ocurren en los sitios MC-3 se exclúan como un comparador ya que las variantes mutadas del gen de TP53 se han seleccionado para la función de unión. La ruta de degradación de ARN mensajero mediada por mutaciones terminadoras (NMCD) implicada es un sistema de vigilancia celular conocido que depende de la información del contexto de codón para permitir que la célula identifique y disponga de los productos génicos defectuosos que contienen mutaciones terminadoras o señales de PARADA que podrían parar prematuramente la traducción). Por ejemplo, respecto al motivo WRC que está asociado con la actividad de AID que da como resultado mutaciones de citosina (C) de la cadena no transcrita, si el número de mutaciones en C se distribuyeran al azar, las mutaciones estarían regularmente distribuidas a través de los sitios MC-1/MC-2 y a través de C>A, C>G y C>T (C>A/G/T). Por tanto, en este ejemplo el número esperado (E) de mutaciones C>T en un sitio MC-1 es el número total de mutaciones C>A/T/G observadas en los sitios MC-1 y MC-2

(es decir, $1+1+72+6+1+1$), dividido por el número de tipos/posiciones posibles de mutaciones (es decir, 6), lo cual es igual a 13,67. A continuación, se aplica una simple prueba de Chi cuadrado para determinar la probabilidad de que la distribución observada surgiera por cambio. En el ejemplo mostrado en la Figura 3, la probabilidad de que la distribución de preferencia codónica MC-1/MC-2 surgiera por cambio para el conjunto seleccionado de los criterios diagnósticos aplicados al conjunto de mutación para el gen TET2 es $7,42 \times 10^{-128}$. Este resultado implica un nivel muy alto de significancia ($P < 10^{-127}$).

Refiriéndose de nuevo a la Figura 3, el número mayor de lo esperado de mutaciones G>A del motivo GYW en sitios MC-2 y el número mayor de lo esperado de mutaciones C>T del motivo WRC en sitios MC-1 indica actividad desaminasa de AID, mientras que el número mayor de lo esperado de mutaciones G>A del motivo CG en sitios MC-2 y el número mayor de lo esperado de mutaciones C>T del motivo CG en sitios MC-1 indica la actividad de APOBEC3G.

Ejemplo 3

Análisis de mutaciones somáticas de TP53 en otros cánceres

Para determinar si la preferencia codónica para mutaciones en los motivos de AID, APOBEC3G y WA observadas en TP53 en muestras de cáncer de mama también ocurre en TP53 en otros cánceres, los datos se extrajeron de la base de datos de TP53 IARC para cáncer cervical (todos los tipos), adenocarcinoma cervical, adenocarcinoma de colon, carcinoma hepatocelular, cáncer pancreático, cáncer de próstata y melanoma maligno y se analizaron como se describió anteriormente.

Las Figuras 4 a 11 muestran la frecuencia y la localización dentro de los codones de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA. Como se muestra en estas figuras, se observaron los patrones de preferencia codónica para las mutaciones en los motivos de AID, APOBEC3G y/o WA en TP53 en cada uno de estos cánceres, indicando que hay una probabilidad estadísticamente muy alta de que un amplio rango de cánceres con mutaciones en TP53 estén asociados con la actividad desaminasa de AID/APOBEC.

Ejemplo 4

Análisis de mutaciones somáticas atribuibles a AID o APOBEC3G en PIK3CA y TET2

La frecuencia y el contexto de codón de las mutaciones somáticas en los motivos de AID, APOBEC3G y WA observadas en PIK3CA de muestras de tejido de cáncer de mama y TET2 de muestras de tejido hematopoyético y linfóide se analizaron usando datos de muestra agregados para diferentes cohortes de pacientes obtenidos de la base de datos COSMIC. Como se muestra en las Figuras 12 y 13, la frecuencia y el contexto de codón de las mutaciones somáticas en los motivos de AID, APOBEC3G y WA indicaron que AID y APOBEC3G eran activas en estos tejidos, y la causa probable de un número significativo de mutaciones somáticas observadas.

Ejemplo 5

Análisis de exomas completos de muestras de sujetos con carcinoma adenoide quístico

Los criterios diagnósticos anteriormente descritos se usaron para evaluar la probabilidad de que AID y/o APOBEC3G estén implicadas en mutagénesis somática dirigida en células de tejido de carcinoma adenoide quístico (CAQ) de pacientes. Los datos de secuencia se obtuvieron de un estudio en el que la secuenciación de exoma completo se realizó en 23 especímenes con CAQ primario con pretratamiento y 1 con metástasis de nudo linfóide local-regional, así como las correspondientes muestras parenquimales de glándula salivar normal de emparejamiento (Stephens y col. (2013) *J. Clin. Invest.* 123(7):2.965-2.968). La secuenciación de exoma identificó 312 mutaciones, con una media de 13 mutaciones por exoma, lo cual es relativamente poco en comparación con otros tumores sólidos. Las mutaciones somáticas se analizaron como se describió anteriormente para determinar la frecuencia y la localización de codón de mutaciones en los sitios GYW/WRC (AID), sitios CG/CG (APOBEC3G) y sitios WA.

La Figura 14 muestra análisis representativo de las mutaciones encontradas en dos muestras de paciente: PD3185a y PD3181a. Aplicando los criterios diagnósticos, se observó que la mutación somática dirigida ocurría en el ácido nucleico de la muestra PD3185a, y que tanto AID como APOBEC3G eran probables de ser activas en las células de esta muestra y la causa de la mutagénesis somática dirigida. Por el contrario, en la muestra que tenía el mayor número de mutaciones somáticas (PD3181a), no se observó evidencia de mutación somática dirigida, sin indicio de que AID o APOBEC3G fueran responsables de las mutaciones somáticas presentes en el ácido nucleico de esta muestra.

En conjunto, se encontró que solamente 9 de 24 de las muestras de CAQ examinadas eran positivas para mutagénesis somática dirigida resultante de la actividad de AID y/o APOBEC3G (Tabla 4). No había correlación entre el número de mutaciones y la mutagénesis somática dirigida, o entre la puntuación de activación de MYB. Esta

puntuación de activación de MYB se derivó para indicar si una muestra particular tiene o no fusiones de los genes MYB-NFIB (Stephens y col. (2013) *J. Clin. Invest.* 123(7):2.965-2.968).

Tabla 4.

ID muestra	Histología	Activación MYB	Análisis MSD		
			Mutaciones (n)	Actividad de AID y/o APOBEC3G	Estadísticamente Significativo (p valor)
PD3178a	Cribriforme	Sí	7	Negativo	ND
PD3179a	Cribriforme	Sí	10	Negativo	ND
PD3180a	Sólido	Sí	13	Negativo	ND
PD3181a	Sólido	Sí	23	Negativo	ND
PD3182a	Cribriforme	Sí	7	Negativo	ND
PD3184a	Sólido	Sí	5	Negativo	ND
PD3185a	Sólido	Sí	11	Positivo	0,0008
PD3186a	Cribriforme con sólido	Sí	13	Positivo	0,0293
PD3188a	Sólido	Sí	14	Negativo	ND
PD3189a	Sólido	No	8	Negativo	ND
PD3190a	Sólido	No	17	Positivo	0,0039
PD3191a	Sólido	Sí	11	Negativo	ND
PD3192a	Sólido	Sí	6	Negativo	ND
PD3193a	Cribriforme	No	16	Negativo	ND
PD3194a	Cribriforme	No	8	Negativo	ND
PD3195a	Sólido	Sí	10	Positivo	0,0132
PD3196a	Cribriforme	Sí	12	Positivo	0,0032
PD3197a	Cribriforme	Sí	7	Positivo	0,0039
PD3198a	Cribriforme	Sí	2	Negativo	ND
PD3199a	Cribriforme	Sí	8	Positivo	0,0109
PD3200a	Sólido	No	12	Positivo	0,0019
PD3208a	Sólido	Sí	7	Positivo	0,000045
PD3216a	Sólido	Sí	3	Negativo	ND
PD3226a	Sólido	Sí	7	Negativo	ND

5

Ejemplo 6

Análisis de exomas completos de muestras de sujetos con carcinoma de próstata

10 Los datos de mutaciones extensas de exoma de cuatro muestras de carcinoma de próstata se obtuvieron de la base de datos COSMIC ("Wellcome Trust Sanger Institute"; <http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/>) y se analizaron como se describió anteriormente para determinar si el ácido nucleico en las muestras contenían mutaciones somáticas dirigidas resultantes de la actividad de AID y/o APOBEC3G. Dos de las muestras eran de pacientes sometidos a autopsia con cáncer de próstata resistente a castración metastático (CPRC), y las otras dos

15 muestras eran de pacientes con cáncer de próstata en estadio pT2c y pT3a, respectivamente.

Como se resume en la Tabla 5, se encontró que tres de las muestras eran positivas para mutación somática dirigida resultante de la actividad de AID y/o APOBEC3G. Interesantemente, la mutación somática dirigida se observó en sujetos con muestras con bajo APE, indicando que este tipo de análisis se podría usar para la detección prematura de carcinoma de próstata antes de que los niveles de APE comiencen a aumentar.

20

La Figura 15 muestra el análisis individual de las mutaciones encontradas en las cuatro muestras de paciente. Además de las indicaciones de la actividad de AID y/o APOBEC3G, altos números de mutaciones G>T en sitios MC-1 y mutaciones C>T en sitios MC-3 en la muestra PR-09-3421, y altos números de mutaciones G>A y mutaciones C>T en la muestra PR-2762 sugieren que otras desaminasas APOBEC pueden estar activas en estos pacientes.

25

Tabla 5.

ID muestra	Estadio de cáncer	Análisis MSD		
		Mutaciones (n)	Actividad de AID y/o APOBEC3G	Estadísticamente Significativo (p valor)
WA7	Autopsia, CPRC	41	positivo	0,000127
WA26	Autopsia, CPRC	115	positivo	0,010397
PR-09-3421	estadio de APE en suero de pT3a (ng/ml)-4,8	49	positivo	5,1x10 ⁻⁰⁵
PR-2762	estadio de APE en suero de pT3a (ng/ml)-5,5	42	negativo	ND

Ejemplo 7Identificación de un motivo de aflatoxina

5 Una transversión G>T en la tercera posición del codón 249 en *TP53* está ligada a aflatoxina, un agente mutagénico exógeno de *Aspergillus* sp., y se ha usado como marcador diagnóstico. Como se muestra en la Figura 5, hay un número muy alto de mutaciones G>T en los sitios MC-3 en combinación con los motivos GG en los genes de *TP53* de muestras de carcinoma hepatocelular (CHC). Para investigar esto más, se analizó una muestra de exoma completo (HCC53T) de la base de datos COSMIC para las mutaciones G>T. Se observó que había 9 mutaciones
 10 G>T en un sitio MC-3 en el exoma completo, cada una coincidente con un motivo GG. Esto sugiere que la aflatoxina causa mutaciones G>T en un sitio MC-3 de un motivo GG.

Ejemplo 8Desarrollo de normas diagnósticas para predecir la actividad de APOBEC3H

Se cree que APOBEC3H se dirige a un motivo GA. Para analizar más el contexto de codón de mutaciones en este motivo, se analizó el exoma completo de tejido de un sujeto con carcinoma de vejiga (secuencia obtenida de la base de datos COSMIC). Como se muestra en la Figura 16, había una predominancia de mutaciones G>A en los sitios
 20 MC-1, indicando que APOBEC3H dirige preferentemente mutaciones a la G en motivos GA cuando la G está en un sitio MC-3, dando como resultado mutaciones G>A.

Ejemplo 9Desarrollo de normas diagnósticas para predecir la actividad de APOBEC3H

Se ha sugerido que APOBEC3G se dirige a un motivo CC además del motivo CG/CG. Para analizar más el contexto de codón de mutaciones en el motivo CC, se analizaron los exomas completos de tejido de 8 sujetos con carcinoma de vejiga antes del tratamiento (secuencias obtenidas de la base de datos COSMIC). Las secuencias de los exomas completos de 8 sujetos (B2, B5, B8-10, B13, B15 y B20) se analizaron como datos reunidos (Figura 17A) y la
 30 secuencia del exoma completo de un sujeto (B13) se analizó independientemente (Figura 17A). Como se muestra en las Figuras 17A y B, había una predominancia estadísticamente significativa de mutaciones C>T en los sitios MC-1, indicando que APOBEC3G preferentemente dirige mutaciones a la C en motivos CC cuando la C dirigida está en un sitio MC-1, dando como resultado mutaciones C>T.

35 La cita de cualquier referencia en el presente documento no se debería interpretar como una admisión de que tal referencia es adecuada como "Técnica anterior" a la presente solicitud.

40 Por toda la memoria la reivindicación ha sido para describir las realizaciones preferidas de la invención sin limitar la invención a una realización cualquiera o colección específica de características. Por lo tanto, los expertos en la técnica apreciarán que, teniendo en cuenta la presente descripción, se pueden hacer diversas modificaciones y cambios en las realizaciones particulares ilustradas sin apartarse del alcance de la presente invención. Todas dichas modificaciones y cambios pretenden estar incluidos dentro del alcance de las reivindicaciones adjuntas.

REIVINDICACIONES

1. Un método para determinar la probabilidad de que un sujeto tenga o desarrollará cáncer, que comprende analizar la secuencia de una molécula de ácido nucleico de una muestra biológica del sujeto para determinar una pluralidad de mutaciones de un tipo de mutación en uno o más motivos reconocidos o dirigidos por el agente mutagénico el contexto de codón de esas mutaciones para identificar de ese modo la localización de una mutación y el tipo de mutación para cada uno de una pluralidad de codones mutados en la molécula de ácido nucleico, en el que el contexto de codón de una mutación individual se determina determinando en cuál de las tres posiciones de un correspondiente codón mutado ocurre la mutación individual, y determinando que la mutagénesis somática dirigida es probable que haya ocurrido cuando hay un porcentaje o número mayor de lo esperado de mutaciones de un tipo de mutación en una de las tres posiciones en la pluralidad de codones mutados, y determinando que el sujeto es probable que tenga o desarrolle cáncer cuando ha ocurrido mutagénesis somática dirigida; en el que el agente mutagénico se selecciona de entre aflatoxina, AID y una citidina desaminasa APOBEC.
2. El método de la reivindicación 1, en el que la mutagénesis somática dirigida se detecta cuando:
- el número o el porcentaje de mutaciones G a A observadas en motivos GYW en la segunda posición de codones mutados (sitios MC-2) en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones C>T observadas en motivos WRC en la primera posición de codones mutados (sitios MC-1) en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones G>A observadas en motivos CG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones C>T observadas en motivos CG en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones C>T observadas en motivos CA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones G>A observadas en motivos GA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones G>A observadas en motivos TG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones G>T observadas en motivos GG en la tercera posición de codones mutados (sitios MC-3) en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
 - el número o el porcentaje de mutaciones C>T observadas en motivos CC en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado; o
 - el número o el porcentaje de mutaciones A>G observadas en motivos WA en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado;
3. El método de la reivindicación 2, en el que el agente mutagénico es AID si el número o el porcentaje de mutaciones G>A observadas en motivos GYW en sitios MC-2 o mutaciones C>T en motivos WRC en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.
4. El método de la reivindicación 2, en el que el agente mutagénico es APOBEC3G si el número o el porcentaje de mutaciones G>A observadas en motivos CG en sitios MC-2, mutaciones C>T en motivos CG en sitios MC-1 o mutaciones C>T en motivos CG en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.
5. El método de la reivindicación 2, en el que el agente mutagénico es APOBEC1 si el número o el porcentaje de mutaciones C>T en motivos CA en sitios MC-1 o mutaciones G>A en motivos TG en sitios MC-2 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.
6. El método de la reivindicación 2, en el que el agente mutagénico es APOBEC3H si el número o el porcentaje de mutaciones G>A observadas en motivos GA en sitios MC-1 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.
7. El método de la reivindicación 2, en el que el agente mutagénico es aflatoxina si el número o el porcentaje de mutaciones G>T observadas en motivos GG en sitios MC-3 en la cadena no transcrita de la molécula de ácido nucleico es mayor de lo esperado.
8. El método de una cualquiera de las reivindicaciones 1 a 7, en el que la muestra biológica comprende tejido o células de pecho, próstata, hígado, colon, estómago, pancreáticas, de piel, tiroides, cervicales, linfoides, hematopoyéticas, de vejiga, pulmón, renales, rectales, de ovario, uterinas, y de cabeza y cuello.
9. El método de una cualquiera de las reivindicaciones 1 a 8, en el que el cáncer se selecciona de entre cáncer de mama, próstata, hígado, colon, estómago, pancreático, de piel, tiroides, cervical, linfoide, hematopoyético, de vejiga, pulmón, renal, rectal, de ovario, de útero, y de cabeza y cuello.

10. El método de una cualquiera de las reivindicaciones 1 a 9, en el que, si la muestra comprende tejido o células de próstata, al sujeto se le diagnostica cáncer de próstata o se determina que es probable que tenga o desarrolle cáncer.
- 5 11. El método de una cualquiera de las reivindicaciones 1 a 9, en el que, si la muestra comprende tejido o células de pecho, al sujeto se le diagnostica cáncer de mama o se determina que es probable que tenga o desarrolle cáncer de mama.
- 10 12. Un agente quimioterapéutico, isótopo radioactivo o efector inmune para su uso en el tratamiento o prevención de un cáncer en un sujeto, en el que el sujeto se expone al agente quimioterapéutico, isótopo radioactivo o efector inmune basándose en que se determina que el sujeto tiene el cáncer o está en riesgo de desarrollar el cáncer mediante el método según una cualquiera de las reivindicaciones 1 a 11; en el que:
- 15 el compuesto quimioterapéutico se selecciona de entre un agente alquilante, antimetabolito, antibiótico antitumoral, agentes antimitóticos, inhibidor de la topoisomerasa, antiestrógeno, regulador a la baja del receptor de estrógeno, antiandrógeno, antagonista de UH, agonista de LHRH, inhibidor de aromatasa, inhibidor de 5 α -reductasa, inhibidor de metaloproteinasa, inhibidor de la función del receptor del activador de plasminógeno tipo uroquinasa, inhibidor de la función del factor de crecimiento, agente antiangiogénico, un inhibidor de AID, un inhibidor de APOBEC3G, un inhibidor de APOBEC1 y un inhibidor de APOBEC3H; y
- 20 el efector inmune se selecciona de entre un anticuerpo, un linfocito y una célula NK.

Figura 1

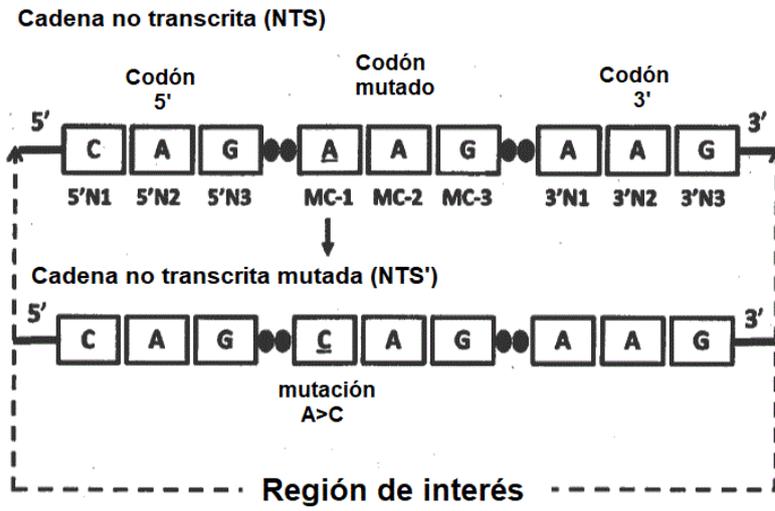


Figura 2

Identificación de mutación somática dirigida en una molécula de ácido nucleico

1. Analizar secuencia de cadena no transcrita de ácido nucleico para mutaciones puntuales sencillas e indicar el tipo de mutación, contexto de codón de la mutación y nucleótidos que flanquean la mutación, e identificar esos en los motivos de APOBEC3G y AID

muestra	ADNc				APO3G			AID			
	ADNc	Codón 5'	Codón mutado	Codón 3'/IC1-	WT	Mut	WA	CG	CG	WRC	GYW
PD3185a	c.1474C>T	CTA	GGG	GGA	1	C	T			WRC	
PD3185a	c.289C>T	CAA	GGC	TCC	1	C	T			WRC	
PD3185a	c.31G>T	AAG	GCA	CCT	1	G	T				GYW
PD3185a	c.2248G>T	CTG	GGC	TGG	1	G	T				
PD3185a	c.3869C>G	GAA	ACT	GAA	2	C	G			WRC	
PD3185a	c.527G>A	TTT	GGC	AAT	2	G	A				GYW
PD3185a	c.4256G>A	ATT	CGT	GGC	2	G	A		CG		
PD3185a	c.8G>A	TTG	GGG	TTG	2	G	A		CG		
PD3185a	c.495C>T	GTA	AAC	TCG	3	C	T			WRC	
PD3185a	c.495C>T	GTA	AAC	TCG	3	C	T			WRC	
PD3185a	c.1473C>T	AGG	GGC	TAC	3	C	T				



2. Tabular los espectros de mutación y calcular la probabilidad de que las mutaciones surgieran por cambio. Para evitar el doble conteo de las mutaciones que son coincidentes con 2 motivos conocidos, los duplicados (anteriormente destacados) se separan según una jerarquía deseada, por ejemplo, AID sobre APOBEC3G

Ensayo MSD para actividad de AID/APO					Muestra: PD3185a	
Mutación Tipo	Localización de mutac.				Mutaciones (n):	11
	MC-1	MC-2	MC-3	Total	Activación MSD:	AID/APO3G POSITIVO
sitios GYW/WRC (AID)						
G>A	0	1	0	1	Significancia (p valor):	0,00083897
G>T	1	0	0	1		
G>C	0	0	0	0		
C>A	0	0	0	0		
C>T	2	0	2	4		
C>G	0	1	0	1		
sitios CG/CG (APOBEC3G)						
G>A	0	2	0	2		
G>T	0	0	0	0		
G>C	0	0	0	0		
C>A	0	0	0	0		
C>T	0	0	0	0		
C>G	0	0	0	0		
sitios WA						
A>T	0	0	0	0		
A>C	0	0	0	0		
A>G	0	0	0	0		
					Categorías diagnósticas	
					Observadas	Esperadas
1. G>A mutaciones del motivo GYW en sitios MC-2					1,00	0,333
2. C>T mutaciones del motivo WRC en sitios MC-1					2,00	0,500
3. G>A mutaciones del motivo CG en sitios MC-2					2,00	0,333
4. C>T mutaciones del motivo CG en sitios MC-1					0,00	0,000
5. A>G mutaciones del motivo WA en sitios MC-2					0,00	0,000
Comentarios:						

Figura 3

Tejido Hematopoyético y linfoide - TET2
(n=615)

Mutación Tipo	Localización de mutaciones			Total
	MC-1	MC-2	MC-3	
sitios <u>GYW</u>WRC (AID)				
G>A	4	11	2	17
G>T	10	3	0	13
G>C	0	1	1	2
C>A	1	1	5	7
C>T	72	6	0	78
C>G	1	1	7	9
sitios <u>CG</u>/CG (APOBEC3G)				
G>A	2	7	1	10
G>T	7	0	0	7
G>C	0	1	0	1
C>A	1	0	0	1
C>T	84	1	3	88
C>G	0	1	3	4
sitios WA				
A>T	4	0	0	4
A>C	1	2	0	3
A>G	13	2	1	16

mutaciones C>T del motivo WRC en sitios MC-1

Observadas (O)=72

Esperadas (E)=(72+10)/6=13,67

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	11	4,83
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	72	13,67
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	7	2,83
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	84	14,50
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	2	3,67

CHI CUADRADO 7,42E-128

Figura 4

Cérvix - "Todo cérvix"				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	1	1	2
G>T	0	3	0	3
G>C	0	1	0	1
C>A	0	0	0	0
C>T	9	0	1	10
C>G	0	0	1	1
sitios CG/CG (APOBEC3G)				
G>A	3	16	0	19
G>T	0	1	0	1
G>C	0	2	0	2
C>A	1	0	0	1
C>T	12	0	0	12
C>G	0	0	0	0
TG/CA (APOBEC1)				
G>A	3	2	3	8
G>T	1	1	0	2
G>C	0	0	0	0
C>A	0	0	0	0
C>T	2	0	3	5
C>G	0	0	0	0
sitios WA				
A>T	0	0	0	0
A>C	0	0	0	0
A>G	0	3	0	3

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	1	0,83
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	9	1,50
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	16	3,67
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	12	2,17
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	3	0,50

CHI, CUADRADO 1,88581E-28

Figura 5

Adenocarcinoma de colon				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	9	3	12
G>T	1	5	1	7
G>C	1	0	0	1
C>A	0	0	5	5
C>T	36	0	0	36
C>G	0	0	0	0
sitios CG/CG (APOBEC3G)				
G>A	19	106	0	125
G>T	2	4	0	6
G>C	0	3	0	3
C>A	0	0	0	0
C>T	95	2	0	97
C>G	2	0	0	2
TG/CA (APOBEC1)				
G>A	2	10	3	15
G>T	2	5	0	7
G>C	1	0	0	1
C>A	0	0	0	0
C>T	5	0	1	6
C>G	0	0	0	0
sitios WA				
A>T	0	0	0	0
A>C	0	3	0	3
A>G	1	11	0	12

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	9	2,67
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	36	6,00
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	106	22,33
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	95	16,50
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	11	2,50

CHI, CUADRADO 2,349E-189

Figura 6

Hígado - Carcinoma Hepatocelular (CHC)				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	1	13	3	17
G>T	2	11	2	15
G>C	0	7	0	7
C>A	3	0	5	8
C>T	28	0	1	29
C>G	0	0	7	7
sitios CG/CG (APOBEC3G)				
G>A	5	31	0	36
G>T	28	23	1	52
G>C	9	8	0	17
C>A	5	2	2	9
C>T	49	1	0	50
C>G	1	0	0	1
TG/CA (APOBEC1)				
G>A	4	19	4	27
G>T	6	11	2	19
G>C	1	3	1	5
C>A	0	0	0	0
C>T	11	0	1	12
C>G	0	0	0	0
sitios WA				
A>T	4	3	1	8
A>C	0	4	0	4
A>G	4	36	1	41
sitios GG				
G>A	16	9	2	27
G>T	14	9	184	207
G>C	4	6	7	17

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	13	5,67
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	28	5,17
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	31	17,33
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	49	9,67
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	36	8,50

CHI, CUADRADO 7,6521E-79

Figura 7

Páncreas - "Todo páncreas"				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	15	2	17
G>T	0	5	2	7
G>C	0	0	0	0
C>A	2	0	2	4
C>T	31	1	1	33
C>G	2	0	2	4
sitios CG/CG (APOBEC3G)				
G>A	6	63	0	69
G>T	3	5	0	8
G>C	0	5	0	5
C>A	1	1	0	2
C>T	59	0	0	59
C>G	5	0	0	5
TG/CA (APOBEC1)				
G>A	8	12	4	24
G>T	4	3	2	9
G>C	2	0	0	2
C>A	0	0	0	0
C>T	6	0	0	6
C>G	0	0	0	0
sitios WA				
A>T	1	1	0	2
A>C	1	0	0	1
A>G	0	16	1	17

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	15	3,33
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	31	6,00
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	63	13,67
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	59	11,00
5. A>G mutaciones del motivo <u>wA</u> en sitios MC-2	16	3,17

CHI, CUADRADO 3,4278E-125

Figura 8

Próstata - "Toda próstata"				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	8	5	13
G>T	2	3	1	6
G>C	0	2	0	2
C>A	0	1	4	5
C>T	22	1	0	23
C>G	0	0	2	2
sitios CG/CG (APOBEC3G)				
G>A	5	34	3	42
G>T	0	3	0	3
G>C	0	2	0	2
C>A	1	0	0	1
C>T	30	2	0	32
C>G	3	0	1	4
TG/CA (APOBEC1)				
G>A	2	5	10	17
G>T	4	5	1	10
G>C	0	0	0	0
C>A	0	0	0	0
C>T	3	2	0	5
C>G	0	0	0	0
sitios WA				
A>T	2	1	1	4
A>C	1	2	0	3
A>G	1	14	0	15

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	8	2,50
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	22	4,00
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	34	7,33
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	30	6,00
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	14	3,50

CHI, CUADRADO 1,75449E-67

Figura 9

Piel - "Melanoma maligno"				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	1	2	0	3
G>T	2	0	0	2
G>C	0	0	0	0
C>A	0	0	0	0
C>T	3	0	2	5
C>G	0	0	0	0
sitios CG/CG (APOBEC3G)				
G>A	2	5	0	7
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	0	0	0
C>T	6	1	1	8
C>G	0	0	0	0
sitios WA				
A>T	0	1	0	1
A>C	0	1	0	1
A>G	0	4	0	4

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	2	0,83
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	3	0,50
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	5	1,17
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	6	1,17
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	4	1,00

CHI, CUADRADO 2,25977E-11

Figura 10

Cérvix - Todo adenocarcinoma				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	0	0	0
G>T	0	1	0	1
G>C	0	0	0	0
C>A	0	0	0	0
C>T	1	0	0	1
C>G	0	0	0	0
sitios CG/CG (APOBEC3G)				
G>A	0	3	0	3
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	0	0	0
C>T	5	0	0	5
C>G	0	0	0	0
TG/CA (APOBEC1)				
G>A	0	0	0	0
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	0	0	0
C>T	0	0	0	0
C>G	0	0	0	0
sitios WA				
A>T	0	1	0	1
A>C	0	0	0	0
A>G	0	0	0	0

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	0	0,17
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	1	0,17
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	3	0,50
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	5	0,83
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	0	0,17

CHI, CUADRADO 1,21287E-07

Figura 11

Cérvix - Adenocarcinoma, NOS				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	2	0	2
G>T	0	1	0	1
G>C	0	0	0	0
C>A	0	0	0	0
C>T	1	0	0	1
C>G	0	0	0	0
sitios CG/CG (APOBEC3G)				
G>A	0	1	0	1
G>T	0	0	1	1
G>C	0	0	0	0
C>A	0	0	0	0
C>T	1	0	0	1
C>G	0	0	0	0
sitios WA				
A>T	0	0	0	0
A>C	0	0	0	0
A>G	1	0	0	1

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	2	0,50
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	1	0,17
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	1	0,17
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	1	0,17
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	0	0,17

CHI, CUADRADO 0,001793999

Figura 12

Pecho - PIK3CA (n=105)

Mutación	Localización de mutaciones			
	MC-1	MC-2	MC-3	Total
sitios <u>GYW</u> / <u>WRC</u> (AID)				
G>A	0	12	0	12
G>T	0	1	2	3
G>C	0	0	1	1
C>A	8	0	0	8
C>T	2	0	0	2
C>G	0	0	0	0
sitios <u>CG</u> / <u>CG</u> (APOBEC3G)				
G>A	3	22	0	25
G>T	0	3	0	3
G>C	0	1	0	1
C>A	1	0	0	1
C>T	3	0	2	5
C>G	0	0	0	0
sitios <u>WA</u>				
A>T	1	1	0	2
A>C	0	0	0	0
A>G	7	3	0	10

Pecho - PIK3CA (n=105)

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	12	2,17
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	2	1,67
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	22	4,83
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	3	0,67
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	3	2,00

CHI, CUADRADO 8,662E-24

Figura 13

Tejido hematopoyético y linfoide - TET2 (n=615)				
Mutación	Localización de mutaciones			
Tipo	MC-1	MC-2	MC-3	Total
sitios <u>GYW</u>/WRC (AID)				
G>A	4	11	2	17
G>T	10	3	0	13
G>C	0	1	1	2
C>A	1	1	5	7
C>T	72	6	0	78
C>G	1	1	7	9
sitios <u>CG</u>/CG (APOBEC3G)				
G>A	2	7	1	10
G>T	7	0	0	7
G>C	0	1	0	1
C>A	1	0	0	1
C>T	84	1	3	88
C>G	0	1	3	4
TG/CA (APOBEC1)				
G>A	5	16	1	22
G>T	16	4	0	20
G>C	3	3	0	6
C>A	1	13	2	16
C>T	147	7	1	155
C>G	3	19	3	25
sitios WA				
A>T	4	0	0	4
A>C	1	2	0	3
A>G	13	2	1	16

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	11	4,83
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	72	13,67
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	7	2,83
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	84	14,50
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	2	3,67

CHI, CUADRADO 7,42E-128

Figura 14

A.

Ensayo MSD para actividad de AID/APO

Mutación Tipo	Localización de mutac.			
	MC-1	MC-2	MC-3	Total
sitios <u>GYW/WRC</u> (AID)				
G>A	0	1	0	1
G>T	1	0	0	1
G>C	0	0	0	0
C>A	0	0	0	0
C>T	2	0	2	4
C>G	0	1	0	1
sitios <u>CG/CG</u> (APOBEC3G)				
G>A	0	2	0	2
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	0	0	0
C>T	0	0	0	0
C>G	0	0	0	0
sitios <u>WA</u>				
A>T	0	0	0	0
A>C	0	0	0	0
A>G	0	0	0	0

Muestra: PD3185a
 Mutaciones (n): 11
 Activación MSD: AID/APO3G POSITIVO
 Significancia (p valor): 0,00083897

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	1,00	0,333
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	2,00	0,500
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	2,00	0,333
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	0,00	0,000
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	0,00	0,000

B.

Ensayo MSD para actividad de AID/APO

Mutación Tipo	Localización de mutac.			
	MC-1	MC-2	MC-3	Total
sitios <u>GYW/WRC</u> (AID)				
G>A	0	0	0	0
G>T	0	0	0	0
G>C	0	0	1	1
C>A	0	0	0	0
C>T	0	0	1	1
C>G	0	0	0	0
sitios <u>CG/CG</u> (APOBEC3G)				
G>A	0	0	0	0
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	0	0	0
C>T	0	1	1	2
C>G	0	0	0	0
sitios <u>WA</u>				
A>T	0	1	0	1
A>C	0	0	1	1
A>G	0	0	0	0
A>C	0	0	0	0
A>G	0	0	0	0

Muestra: PD3181a
 Mutaciones (n): 23
 Activación MSD: NEGATIVO
 Significancia (p valor):

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	0,00	0,000
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	0,00	0,000
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	0,00	0,000
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	0,00	0,167
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	0,00	0,167

Figura 15A

A.

Ensayo MSD para actividad de AID/APO

Mutación Tipo	Localización de mutaciones			
	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	0	1	0	1
G>T	0	1	0	1
G>C	1	0	0	1
C>A	0	0	0	0
C>T	2	0	1	3
C>G	0	0	0	0
sitios CG/CG (APOBEC3G)				
G>A	0	1	1	2
G>T	0	1	0	1
G>C	0	0	0	0
C>A	0	0	0	0
C>T	2	0	0	2
C>G	0	0	0	0
sitios WA				
A>T	1	0	0	1
A>C	0	0	1	1
A>G	0	2	0	2

Muestra: WA7
 Mutaciones (n): 41
 Activación MSD: POSITIVO
 Significancia (p valor): 0,000126626

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	1,00	0,500
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	2,00	0,333
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	1,00	0,333
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	2,00	0,333
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	2,00	0,500

Comentarios:

B.

Ensayo MSD para actividad de AID/APO

Mutación Tipo	Localización de mutaciones			
	MC-1	MC-2	MC-3	Total
sitios GYW/WRC (AID)				
G>A	1	1	3	5
G>T	2	0	1	3
G>C	0	0	0	0
C>A	0	0	0	0
C>T	3	0	4	7
C>G	1	0	0	1
sitios CG/CG (APOBEC3G)				
G>A	7	5	4	16
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	0	0	0
C>T	3	9	2	14
C>G	0	1	0	1
sitios WA				
A>T	2	1	0	3
A>C	0	1	0	1
A>G	0	1	0	1

Muestra: WA26
 Mutaciones (n): 115
 Activación MSD: POSITIVO
 Significancia (p valor): 0,010396508

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo <u>GYW</u> en sitios MC-2	1,00	0,667
2. C>T mutaciones del motivo <u>WRC</u> en sitios MC-1	3,00	0,667
3. G>A mutaciones del motivo <u>CG</u> en sitios MC-2	5,00	2,000
4. C>T mutaciones del motivo <u>CG</u> en sitios MC-1	3,00	2,167
5. A>G mutaciones del motivo <u>WA</u> en sitios MC-2	1,00	0,833

Comentarios:

Figura 15B

C.

Ensayo MSD para actividad de AID/APO

Mutación Tipo	Localización de mutaciones			Total
	MC-1	MC-2	MC-3	
sitios GYW/WRC (AID)				
G>A	0	1	1	2
G>T	1	0	0	1
G>C	0	0	0	0
C>A	0	0	0	0
C>T	1	1	1	3
C>G	0	0	0	0
sitios CG/CG (APOBEC3G)				
G>A	1	5	1	7
G>T	5	0	0	5
G>C	0	0	0	0
C>A	0	0	0	0
C>T	6	2	6	14
C>G	0	0	0	0
sitios WA				
A>T	1	0	0	1
A>C	0	1	0	1
A>G	0	1	0	1

Muestra:	PR-09-3421
Mutaciones (n):	49
Activación MSD:	POSITIVO
Significancia (p valor):	5,10206E-05

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo GYW en sitios MC-2	1,00	0,333
2. C>T (mutaciones del motivo WRC en sitios MC-1	1,00	0,333
3. G>A mutaciones del motivo CG en sitios MC-2	5,00	1,833
4. C>T (mutaciones del motivo CG en sitios MC-1	6,00	1,333
5. A>G mutaciones del motivo WA en sitios MC-2	1,00	0,500

Comentarios:

También indica números significativos de G>T en MC-1 y C>T en MC-3

Sugiere que otra desaminasa de la familia APO con una preferencia codónica diferente podría ser activa en este paciente.

D.

Ensayo MSD para actividad de AID/APO

Mutación Tipo	Localización de mutaciones			Total
	MC-1	MC-2	MC-3	
sitios GYW/WRC (AID)				
G>A	2	1	1	4
G>T	0	0	0	0
G>C	0	0	0	0
C>A	1	0	0	1
C>T	0	1	0	1
C>G	0	0	1	1
sitios CG/CG (APOBEC3G)				
G>A	3	2	1	6
G>T	0	0	0	0
G>C	0	0	0	0
C>A	0	1	0	1
C>T	0	3	4	7
C>G	0	0	0	0
sitios WA				
A>T	0	0	0	0
A>C	0	0	0	0
A>G	1	1	0	2

Muestra:	PR-2762
Mutaciones (n):	42
Activación MSD:	NEGATIVO
Significancia (p valor):	0,346518269

Categorías diagnósticas	Observadas	Esperadas
1. G>A mutaciones del motivo GYW en sitios MC-2	1,00	0,500
2. C>T (mutaciones del motivo WRC en sitios MC-1	0,00	0,333
3. G>A mutaciones del motivo CG en sitios MC-2	2,00	0,833
4. C>T (mutaciones del motivo CG en sitios MC-1	0,00	0,667
5. A>G mutaciones del motivo WA en sitios MC-2	1,00	0,333

Comentarios:

Actividad desaminasa (alto G>A/C>T) puede ser debida a uno u otros más miembros de la familia APOBEC de desaminasas

Figura 16

Fuente: Nombre muestra B10, base de datos COSMIC

Mutación Tipo	Localización de mutaciones			
	MC-1	MC-2	MC-3	Total
sitios GA (APOBEC3H)				
G>A	19	2	3	24
G>T	4	0	0	4
G>C	5	4	1	10
Total	28	6	4	38

Carcinoma de vejiga, grado 2
Macho, 52
Observación - Descubrimiento
(pretratamiento)

Muestra:	0
Mutaciones (n):	88(38/88 ocurre en sitios GA)
Activación MSD para APOBEC3H:	POSITIVO
Significancia (p valor):	6,589E-11

Ensayo diagnóstico para APOBEC3H	Observadas	Esperadas
1. G>A mutaciones del motivo GA en sitios MC-1	19	4,222
2. G>A mutaciones del motivo GA en sitios MC-2	2	4,222
3. G>A mutaciones del motivo GA en sitios MC-3	3	4,222
4. G>T mutaciones del motivo GA en sitios MC-1	4	4,222
5. G>T mutaciones del motivo GA en sitios MC-2	0	4,222
6. G>T mutaciones del motivo GA en sitios MC-3	0	4,222
7. G>C mutaciones del motivo GA en sitios MC-1	5	4,222
8. G>C mutaciones del motivo GA en sitios MC-2	4	4,222
7. G>C mutaciones del motivo GA en sitios MC-3	1	4,222

Figura 17

A.

Ensayo MSD para actividad de APOBEC3G (usando dinucleótido CC)

Muestras B2, B5, B8-10, B13, B15, B20 COSMIC (v.66)

Mutación Tipo	Localización de mutaciones			
	MC-1	MC-2	MC-3	Total
sitios CC (APOBEC3G)				
C>A	2	0	4	6
C>G	2	2	3	7
C>T	9	1	2	12
Total	13	3	9	25

Carcinoma de vejiga

Observación - todo descubrimiento (pre-tratamiento)

Muestra: 8 muestras (pretratamiento)

Mutaciones (n): 473

Activación MSD para APOBEC3H: POSITIVO (p < 0,05)
Significancia (p valor): 0,01343184

Ensayo diagnóstico para APOBEC3H

	Observadas	Esperadas
1. C>A mutaciones del motivo CC en sitios MC-1	2	2,778
2. C>A mutaciones del motivo CC en sitios MC-2	0	2,778
3. C>A mutaciones del motivo CC en sitios MC-3	4	2,778
4. C>G mutaciones del motivo CC en sitios MC-1	2	2,778
5. C>G mutaciones del motivo CC en sitios MC-2	2	2,778
6. C>G mutaciones del motivo CA en sitios MC-3	3	2,778
7. C>T mutaciones del motivo CC en sitios MC-1	9	2,778
8. C>T mutaciones del motivo CC en sitios MC-2	1	2,778
7. C>T mutaciones del motivo CC en sitios MC-3	2	2,778

B.

Ensayo MSD para actividad de APOBEC3G (usando motivo de dinucleótido de CC)

Fuente: Nombre muestra B13, base de datos COSMIC (v.66)

Mutación Tipo	Localización de mutaciones			
	MC-1	MC-2	MC-3	Total
sitios CC (APOBEC3G)				
C>A	0	0	0	0
C>G	0	0	0	0
C>T	2	0	0	2
Total	2	0	0	2

Carcinoma vejiga, grado 2

Macho, 64

Observación - Descubrimiento (pre-tratamiento)

Muestra: 0

Mutaciones (n): 39

Activación MSD para APOBEC3H: POSITIVO (p < 0,05)
Probabilidad de que las mutaciones sean al azar: $=(1/9)*(1/9) = 0,0123$

Figura 18

