

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 602 072**

51 Int. Cl.:

G01N 33/574 (2006.01)

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.09.2012 PCT/US2012/058164**

87 Fecha y número de publicación internacional: **04.04.2013 WO13049750**

96 Fecha de presentación y número de la solicitud europea: **29.09.2012 E 12775575 (9)**

97 Fecha y número de publicación de la concesión europea: **26.10.2016 EP 2761299**

54 Título: **Mutaciones en SF3B1 y leucemia linfocítica crónica**

30 Prioridad:

29.09.2011 US 201161540618 P

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

17.02.2017

73 Titular/es:

**TROVAGENE, INC. (100.0%)
11055 Flintkote Ave., Suite B
San Diego, California 92121, US**

72 Inventor/es:

**ROSSI, DAVIDE;
GAIDANO, GIANLUCA y
FOA, ROBERT**

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 602 072 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

Mutaciones en SF3B1 y leucemia linfocítica crónica

5 **Campo de la divulgación**

La presente divulgación se refiere en general a los campos de la biología molecular, genética y cáncer. Específicamente, se utilizan las mutaciones del gen SF3B1 en métodos *in vitro* para diagnosticar, pronosticar, y determinar los regímenes de tratamiento óptimo para sujetos con leucemia linfocítica crónica.

10

Antecedentes de la divulgación

La leucemia linfocítica crónica (CLL) es un cáncer de unos leucocitos llamados linfocitos. Según progresa la CLL, aumenta el número de linfocitos B o células B presentes en la médula ósea. Estas células B cancerosas migran o se diseminan desde la médula ósea a la sangre. Por medio de la sangre estas células cancerosas tienen acceso a todos los órganos del cuerpo. Más comúnmente, las células B cancerosas afectan a los ganglios linfáticos, hígado y bazo. En último término la médula ósea no funciona apropiadamente, dando lugar a la muerte.

15

El curso clínico de la leucemia linfocítica crónica (CLL) varía desde un trastorno muy indolente con una esperanza de vida normal, hasta una enfermedad que progresa rápidamente que en último término se vuelve quimiorrefractaria y da lugar a la muerte. De manera ocasional, la CLL sufre una transformación histológica en el síndrome de Richter (RS).

20

El curso clínico variable de la CLL está promovido, al menos en parte, por la heterogeneidad molecular de la enfermedad. A pesar de los recientes avances, las lesiones genéticas identificadas hasta la fecha no recapitulan completamente la patogénesis molecular de CLL y no explican completamente el desarrollo de las graves complicaciones, tales como la quimiorrefractariedad y la transformación en RS, lo que sigue representando una necesidad clínica no satisfecha.

25

La identificación de las lesiones genéticas asociadas con la quimiorrefractariedad representa una etapa crítica para la identificación precoz de los pacientes con alto riesgo de CLL y para el desarrollo de fármacos molecularmente a medida.

30

Papaemmanuil et al (NEJM. 2011; 365(15); 1384-95) desvela un método para pronosticar un individuo con CLL detectando mutaciones en el gen SF3B1, que son indicativas de una supervivencia más larga libre de incidentes.

35

Sumario de la divulgación

Los métodos de la divulgación proporcionan una solución para la necesidad percibida durante mucho tiempo y sin resolver de un indicador biológico de la progresión de la enfermedad y la sensibilidad al tratamiento. La divulgación proporciona mutaciones sin sentido y de eliminación en el gen SF3B1 (factor de empalme 3b, subunidad 1, 155 kDa) que cambian la secuencia de aminoácidos de la proteína codificada. Estos cambios de la proteína tienen consecuencias funcionales. El SF3B1 codifica la subunidad 1 del complejo proteico de factor 3 b de corte y empalme. En condiciones normales o de tipo silvestre, el factor de corte y empalme 3b, junto con el factor de corte y empalme 3a y una unidad de ARN 12S, forman el complejo de ribonucleoproteínas nucleares pequeñas U2 (U2 RNPnp). El complejo de factores de corte y empalme 3b/3a se une al pre-ARNm. El factor de corte y empalme 3b también es un componente del empalmosoma menor tipo U12. Por lo tanto, la subunidad 1 del complejo proteico del factor 3b de corte y empalme tiene varios papeles críticos en el mecanismo de corte y empalme en la célula. Las mutaciones en SF3B1 afectan a la capacidad de una célula para convertir el pre-ARNm, que contiene la secuencia intrónica, en ARNm maduro. En el contexto de la CLL, estas mutaciones son predictivas de la disminución de la supervivencia en pacientes y del aumento de la resistencia al tratamiento con fludarabina.

40

45

50

La divulgación proporciona un método *in vitro* para pronosticar un sujeto con leucemia linfocítica crónica (CLL) que comprende: (a) determinar la secuencia de una parte del gen *SF3B1* en una muestra biológica obtenida del sujeto, donde la parte del gen *SF3B1* comprende una secuencia que codifica un dominio HEAT3, HEAT4 o HEAT5; y (b) analizar la secuencia de gen *SF3B1* en cuanto a una mutación, donde la presencia de la mutación en la secuencia del gen *SF3B1* predice una disminución de la supervivencia del sujeto, pronosticando de esta manera el sujeto con leucemia linfocítica crónica (CLL). En muchas realizaciones, la mutación está presente en el dominio HEAT3, HEAT4 o HEAT5. En realizaciones adicionales, el método comprende la preparación de una molécula de ácido nucleico de un sujeto después del análisis como se desvela en el presente documento para detectar alteraciones de ácido nucleico que predicen o pronostican el curso y/o resultado probable de la CLL.

55

60

De manera alternativa, o adicionalmente, la divulgación proporciona un método *in vitro* para pronosticar un sujeto con leucemia linfocítica crónica (CLL), que comprende: (a) determinar la secuencia de una parte del polipéptido SF3B1 en una muestra biológica obtenida del sujeto, donde la parte del polipéptido SF3B1 comprende la secuencia de un dominio HEAT3, HEAT4 o HEAT5; y (b) analizar la secuencia del polipéptido SF3B1 en cuanto a una

65

mutación, donde la presencia de la mutación en la secuencia del polipéptido SF3B1 predice una disminución de la supervivencia del sujeto, pronosticando de esta manera el sujeto con una leucemia linfocítica crónica (CLL). En muchas realizaciones, la mutación está presente con el dominio HEAT3, HEAT4 o HEAT5. En realizaciones adicionales, el método comprende la preparación de una molécula de polipéptido a partir de un sujeto seguido por el análisis como se desvela en el presente documento para detectar las alteraciones de aminoácidos que predicen o pronostican el curso y/o resultado probable de CLL.

Con respecto a los métodos para pronosticar un sujeto, la expresión disminución de la supervivencia incluye la supervivencia libre de tratamiento o la supervivencia total. Algunas realizaciones de la invención incluyen la predicción o pronóstico del curso y/o resultado probables de CLL en un sujeto con la alteración de ácido nucleico, o aminoácido, en ausencia de tratamiento para CLL. En otras realizaciones, el curso y/o resultado probables son para un sujeto con la alteración de ácido nucleico, o aminoácido, si se trata con un tratamiento desvelado para CLL.

La divulgación también proporciona un método *in vitro* para determinar la respuesta de un sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina, que comprende: (a) determinar la secuencia de una parte del gen SF3B1 en una muestra biológica obtenida del sujeto, donde la parte del gen SF3B1 comprende una secuencia que codifica un dominio HEAT3, HEAT4 o HEAT5; y (b) analizar la secuencia del gen SF3B1 en cuanto a una mutación donde la presencia de la mutación en la secuencia del gen SF3B1 indica que el sujeto es resistente o refractario a la fludarabina, determinando de esta manera la respuesta del sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina. En muchas realizaciones, la mutación está presente en el dominio HEAT3, HEAT4 o HEAT5. En realizaciones adicionales, el método comprende la preparación de una molécula de ácido nucleico de un sujeto seguido por el análisis como se desvela en el presente documento para detectar alteraciones de ácido nucleico que predicen o pronostican la probable falta de respuesta de la CLL en el sujeto al tratamiento con fludarabina.

De manera alternativa, o adicionalmente, la divulgación proporciona un método *in vitro* para determinar la respuesta de un sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina, que comprende: (a) determinar la secuencia de una parte del polipéptido SF3B1 en una muestra biológica obtenida del sujeto, donde la parte del polipéptido SF3B1 comprende una secuencia de un dominio HEAT3, HEAT4 o HEAT5; y (b) analizar la secuencia del polipéptido SF3B1 en cuanto a una mutación, donde la presencia de la mutación en la secuencia del polipéptido SF3B1 indica que el sujeto es resistente o refractario a fludarabina, determinando de esta manera la respuesta del sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina. En muchas realizaciones, la mutación está presente en el dominio HEAT3, HEAT4 o HEAT5. En realizaciones adicionales, el método comprende la preparación de una molécula de polipéptido a partir de un sujeto seguido por el análisis como se desvela en el presente documento para detectar alteraciones de aminoácidos que predicen o pronostican la probable falta de respuesta de la CLL en el sujeto al tratamiento con fludarabina.

En ciertos aspectos del presente método, la parte del polipéptido SF3B1 que comprende la secuencia de un dominio HEAT3, HEAT4 o HEAT5 se selecciona de entre el grupo que consiste en las SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, y 19. En algunas realizaciones, la parte del polipéptido que comprende la secuencia de un dominio HEAT3, HEAT4 o HEAT5 es SEQ ID NO: 1 o 19.

Con respecto a los métodos para determinar la respuesta de un sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina, en sujetos que albergan una o más mutaciones en el gen SF3B1 o el polipéptido SF3B1, se suspende el tratamiento con fludarabina o se sustituye por el tratamiento con clorambucilo, ciclofosfamida, rituximab, alemtuzumab, bendamustina, o una combinación de los mismos. En muchas realizaciones, el tratamiento con fludarabina se sustituye con el tratamiento con alemtuzumab. En otras realizaciones, el tratamiento con fludarabina no se inicia, y a su vez, el sujeto se trata con clorambucilo, ciclofosfamida, rituximab, alemtuzumab, bendamustina, o una combinación de los mismos. Por lo tanto, la divulgación incluye adicionalmente un método para tratar un sujeto con CLL que comprende la determinación de la sensibilidad al tratamiento con fludarabina como se desvela en el presente documento y suspender o alterar el tratamiento como se ha descrito anteriormente. En realizaciones adicionales, la divulgación incluye un método para tratar un sujeto con CLL que comprende la determinación de la sensibilidad al tratamiento con fludarabina como se desvela en el presente documento e iniciar el tratamiento para la CLL con una terapia distinta a la fludarabina como se ha descrito anteriormente.

Con respecto a cualquier método de la divulgación, la etapa de análisis puede incluir la reacción en cadena de la polimerasa (PCR), la secuenciación de Sanger, la secuenciación de próxima generación o una combinación de las mismas como conoce el experto. En algunas realizaciones, el análisis puede utilizar moléculas de ADN preparadas o aisladas que se utilizan como matrices o para la hibridación. En algunos casos, se amplifica una molécula de ADN matriz, tal como por PCR o PCR cuantitativa, con marcado detectable opcional de las moléculas amplificadas para ayudar a su detección. En otros casos, las moléculas amplificadas pueden detectarse basándose en la hibridación con una sonda de polinucleótido. En casos alternativos, la molécula de ADN preparada o aislada no se copia ni amplifica, sino que a su vez se secuencia directamente, con inmobilización opcional directa o indirecta en un soporte sólido o medio de fase sólida, antes de la secuenciación. Ejemplos no limitantes de un soporte sólido o un medio de fase sólida incluyen una perla, una microperla, u otro material insoluble. En casos adicionales, una molécula de ADN

se puede preparar o aislar por incorporación en, o como parte de, una emulsión o compartimento, tal como una gota y microgota, u otra suspensión en solución.

5 En la práctica de la divulgación, el sujeto puede estar diagnosticado de leucemia linfocítica crónica (CLL) o puede no estar diagnosticado de leucemia linfocítica crónica (CLL). Estos métodos se pueden aplicar en cualquier punto del diagnóstico o tratamiento de un sujeto. Los sujetos con CLL pueden presentar uno o más de los síntomas de la siguiente lista no limitante: agrandamiento de ganglios linfáticos, hígado o bazo; sudoración excesiva o sudores nocturnos; fatiga; fiebre; infecciones recurrentes; y pérdida de peso espontánea. Los sujetos con CLL pueden presentar un recuento de leucocitos mayor de lo normal, anemia y/o trombocitopenia. Los sujetos se pueden tratar con uno o más de los siguientes tratamientos ejemplares no limitantes: fludarabina (Fludara), clorambucilo, ciclofosfamida (Cytoxan), rituximab (Rituxan), alemtuzumab (Campath), bendamustina, o una combinación de los mismos. Los sujetos que son resistentes o refractarios al tratamiento con fludarabina debido a que albergan una o más de las mutaciones descritas en el presente documento se pueden tratar con Alemtuzumab (Campath). Los sujetos que albergan una o más de las mutaciones descritas en el presente documento también se pueden tratar con Bendamustina, particularmente cuando la CLL recurre tras un tratamiento inicial o cuando el sujeto o paciente sufre una recaída.

20 En la práctica de la divulgación, la muestra biológica incluye una molécula de ADN genómico, ADNc, o ARN aislada o purificada. De manera alternativa, la muestra biológica incluye una molécula de polipéptido aislada y purificada. La muestra biológica se puede obtener a partir de uno o más tejidos o fluidos corporales. Para la detección de la mutación, los tejidos ejemplares incluyen, pero no se limitan a médula ósea, células de la sangre, células de sangre periférica, ganglios linfáticos, bazo, tejido muscular (incluyendo músculo liso, visceral, estriado, esquelético, o cardíaco compuesto por células o fibras musculares), tejidos del sistema nervioso (incluyendo, pero sin limitarse a, las neuronas y la glía o el sistema nervioso central y periférico), o tejidos epiteliales (incluyendo, pero sin limitarse a, células epiteliales que comprenden la piel, tracto respiratorio, tracto digestivo, y tracto digestivo). Para confirmar el origen somático de las mutaciones, los especímenes incluyen saliva, células epidérmicas obtenidas por raspado no invasivo de la piel o una torunda de la parte interna del carrillo. En algunas realizaciones, la muestra de tejido comprende glóbulos rojos o leucocitos aislados de sangre completa. Los fluidos corporales ejemplares incluyen pero no se limitan a, humor acuoso, humor vítreo, bilis, sangre completa, suero sanguíneo, leche natural, líquido cefalorraquídeo (LCR), endolinfa, perilinf, jugos gástricos, mucus (incluyendo drenaje nasal y flemas), fluido peritoneal, líquido pleural, saliva, sebo (grasa cutánea), sudor, lágrimas, y orina. En algunas realizaciones, el fluido corporal es sangre entera, suero sanguíneo, endolinfa, perilinf, saliva, u orina.

35 En la práctica de la divulgación, la mutación en SF3B1 puede ser una mutación sin sentido o una eliminación en fase en la secuencia de polinucleótido o polipéptido de un gen *SF3B1* o el polipéptido resultante codificado por un gen *SF3B1*, respectivamente. En muchas realizaciones, una mutación sin sentido o una eliminación en fase está en el dominio HEAT3, HEAT4 o HEAT5.

40 En algunos métodos desvelado para pronosticar un sujeto, la mutación puede ser una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2044 o 2146; una sustitución de una Timina (T) por una Guanina (G) en la posición de base de nucleótido 2046; o una sustitución de una Alanina (A) por una Guanina (G) en la posición de base de nucleótido 2267, de SEQ ID NO: 17 (Nº de registro GenBank N° NM_012433.2).

45 En otros métodos desvelados para pronosticar un sujeto, la mutación puede ser una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Lisina (Lys o K) en el codón 666 o 700; dar como resultado una sustitución de un resto de Asparagina (Asn o N) por un resto de Lisina (Lys o K) en el codón 666; o una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Glicina (Gly o G) en el codón 740, de SEQ ID NO: 19 (Nº de registro GenBank NP_036565.2).

50 En algunos métodos desvelados para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación puede ser una sustitución de una Timina (T) por una Adenina (A) en la posición de base de nucleótido 1938; una sustitución de una Adenina (A) por una Citosina (C) en la posición de base de nucleótido 2034; una sustitución de una Guanina (G) por una Citosina (C) en la posición de base de nucleótido 2032; una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2044; una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2146; una eliminación de la secuencia de nucleótido CAGAAA correspondiente a las posiciones de bases 2143 a 2148; o una sustitución de una Guanina (G) por una Citosina (C) en la posición de base de nucleótido 2056, de SEQ ID NO: 17.

60 En algunos métodos desvelados para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación puede ser una sustitución de un resto de Serina (Ser o S) por un resto de Arginina (Arg o R) en el codón 630; una sustitución de un resto de Glutamina (Gln o Q) por un resto de Histidina (His o H) en el codón 662; una sustitución de un resto de Ácido Aspártico (Asp o D) por un resto de Histidina (His o H) en el codón 662; una sustitución de Ácido Glutámico (Glu o E) por un resto de Lisina (Lys o K) en el codón 666 o 700; una eliminación de un resto de Glutamina (Gln o Q) en el codón 699 y un resto de Lisina (Lys o K) en el codón 700; una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Glutamina (Gln o Q) en el codón 670, de SEQ ID NO: 19.

En muchos métodos desvelados para pronosticar un sujeto, la mutación da como resultado una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Lisina (Lys o K) en el codón 666 o 700. Cuando están presentes la sustitución K666E o K700E, la mutación es una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2044 o 2146, respectivamente. Con respecto a algunos métodos desvelados para pronosticar un sujeto, la mutación da como resultado una sustitución de un resto de Asparagina (Asn o N) por un resto de Lisina (Lys o K) en el codón 666. Cuando está presente la sustitución K666N, la mutación es una sustitución de una Timina (T) por una Guanina (G) en la posición de nucleótido 2046. Con respecto a otros métodos desvelados para pronosticar un sujeto, la mutación da como resultado una sustitución de un Ácido Glutámico (Glu o E) por un resto de Glicina (Gly o G) en el codón 740. Cuando está presente la sustitución G740E, la mutación es una sustitución de una Adenina (A) por una Guanina (G) en la posición de nucleótido 2267.

En otros métodos desvelados para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación da como resultado una sustitución de un resto de Serina (Ser o S) por un resto de Arginina (Arg o R) en el codón 630. Cuando está presente la sustitución R630S, la mutación es una sustitución de una Timina (T) por una Adenina (A) en la posición de base de nucleótido 1938. En algunos métodos desvelados para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación da como resultado una sustitución de un resto de Glutamina (Gln o Q) por un resto de Histidina (His o H) en el codón 662. Cuando está presente una sustitución H662Q, la mutación es una sustitución de una Adenina (A) por una Citosina (C) en la posición de base de nucleótido 2034. En métodos desvelados adicionales para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación da como resultado una sustitución de un resto de Ácido Aspártico (Asp o D) por un resto de Histidina (His o H) en el codón 662. Cuando está presente una sustitución H662D, la mutación es una sustitución de una Guanina (G) por una Citosina (C) en la posición de base de nucleótido 2032. En métodos desvelados adicionales para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación da como resultado una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Lisina (Lys o K) en el codón 666 o 700. Cuando está presente la sustitución K666E, la mutación es una sustitución de una Guanina (G) por una Adenina (A) en la posición de nucleótido 2044. Cuando está presente la sustitución K700E, la mutación es una sustitución de una Guanina (G) por Adenina (A) en la posición de base de nucleótido 2146. En casos de métodos desvelados para la determinación de la respuesta de un sujeto al tratamiento con fludarabina, la mutación da como resultado una eliminación de un resto de Glutamina (Gln o Q) en el codón 699 y un resto de Lisina (Lys o K) en el codón 700. Cuando está presente la eliminación Q699-K700, la mutación es la eliminación de la secuencia CAGAAA correspondiente con las posiciones de base de nucleótido 2143 a 2148. En casos adicionales de métodos desvelados para determinar la respuesta de un sujeto al tratamiento con fludarabina, la mutación da como resultado una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Glutamina (Gln o Q) en el codón 670. Cuando está presente la sustitución Q670E, la mutación es una sustitución de una Guanina (G) por una Citosina (C) en la posición de base de nucleótido 2056.

El factor de empalme 3b de *Homo sapiens*, subunidad 1, 155 kDa, *SF3B1*, transcrito variante 1, está codificado por la siguiente secuencia de ARNm (NM_012433.2, SEQ ID NO: 17) (la parte que codifica los dominios HEAT3, HEAT4 y HEAT5 está subrayada):

1 ggaagtttett gggagcgcca gttccgtctg tgtgttcgag tggacaaaat ggcgaagatc
 61 gccaagactc acgaagatat tgaagcacag attcgagaaa ttcaaggcaa gaaggcagct
 121 cttgatgaag ctcaaggagt gggcctcgat tctacaggtt attatgacca ggaaatttat
 181 ggtggaagtg acagcagatt tgctggatac gtgacatcaa ttgctgcaac tgaacttgaa
 241 gatgatgacg atgactatcc atcatctacg agtttgcttg gtcagaagaa gccaggatat
 301 catgcccctg tggcattgct taatgatata ccacagtcaa cagaacagta tgatccattt
 361 gctgagcaca gacctccaaa gattgcagac cgggaagatg aatacaaaaa gcataggcgg
 421 accatgataa tttccccaga gcgtcttgat ccttttgag atggagggaa aaccctgat
 481 cctaaaatga atgctaggac ttacatggat gtaatgagag aacaacactt gactaaagaa
 541 gaacgagaaa ttaggcaaca gctagcagaa aaagctaaag ctggagaact aaaagtcgtc
 601 aatggagcag cagcgtccca gcctccatca aaacgaaaac ggcgttggga tcaaacagct
 661 gatcagactc ctggtgccac tccccaaaaa ctatcaagtt gggatcaggc agagaccctt
 721 gggcatactc ctcccttaag atgggatgag acaccaggtc gtgcaaaggg aagcgagact
 781 cctggagcaa ccccaggctc aaaaatatgg gatcctacac ctageccacac accagcggga
 841 gctgctactc ctggacgagg tgatacacca ggccatgcca caccaggcca tggaggcgca
 901 acttccagtg ctcgtaaaaa cagatgggat gaaaccccca aacagagag agatactcct
 961 gggcatggaa gtggatggc tgagactcct cgaacagatc gaggtggaga ttctattggt
 1021 gaaacaccga ctctggagc cagtaaaaga aaatcacggt gggatgaaac accagctagt
 1081 cagatgggtg gaagcactcc agttctgacc cctggaaaga caccaattgg cacaccagcc
 1141 atgaacatgg ctacccctac tccaggctac ataatgagta tgactcctga acagcttcag
 1201 gcttggcggg gggaaagaga aattgatgag agaaatcgcc cactttctga tgaggaatta
 1261 gatgctatgt tcccagaagg atataaggta ctctctctc cagctggtta tgttctatt
 1321 cgaactccag ctcgaaagct gacagctact ccaacacctt tgggtggtat gactggtttc
 1381 cacatgcaaa ctgaagatcg aactatgaaa agtggttaag accagccatc tggaaatcct
 1441 ccatttttaa aacctgatga tattcaatac tttgataaac tattggttga tgttgatgaa
 1501 tcaacactta gtccagaaga gcaaaaagag agaaaaataa tgaagttgct tttaaaaatt
 1561 aagaatggaa caccaccaat gagaaaggct gcattgcgtc agattactga taaagctcgt
 1621 gaatttgag ctggtccttt gtttaatcag attcttctc tgctgatgtc tctacactt
 1681 gaggatcaag agcgtcattt acttgtgaaa gttattgata ggatactgta caaacttgat
 1741 gacttagttc gtccatagt gcataagatc ctcggtgca ttgaaccgct attgattgat
 1801 gaagattact atgctagagt ggaaggccga gagatcattt ctaatttggc aaaggctgct
 1861 ggtctggcta ctatgatctc taccatgaga cctgatatag ataacatgga tgagtatgct

 1921 cgtaacacaa cagctagagc ttttgetggt gtagcctctg ccctgggcat tcttcttta
 1981 ttgcccttct taaaagctgt gtgcaaaagc aagaagtcct ggcaagcgag acacactggt
 2041 attagattg tacaacagat agctattcct atgggctgtg ccactctgcc acatcttaga
 2101 agtttagttg aatcattga acatggtcct gtggatgagc agcagaaagt tccgaccatc
 2161 agtgctttgg ccattgetgc cttggctgaa gcagcaactc cttatggtat cgaatctttt
 2221 gattctgtgt taaagccttt atggaagggg atccgccaac acagaggaaa gggtttggct
 2281 gctttcttga aggetattgg gtatcttatt cctcttatgg atgcagaata tgccaactac
 2341 tatactagag aagtgatggt aatccttatt cgagaattcc agtctcctga tgaggaaatg

ES 2 602 072 T3

```

2401 aaaaaaattg tgctgaaggt ggtaaaacag tgttgtggga cagatggtgt agaagcaaac
2461 tacattaaaa cagagattct tcctcccttt tttaaacact tctggcagca caggatggct
2521 ttggatagaa gaaattaccg acagttagtt gatactactg tggagttggc aaacaaagta
2581 ggtgcagcag aaattatata caggattgtg gatgatctga aagatgaagc cgaacagtac
2641 agaaaaatgg tgatggagac aattgagaaa attatgggta atttgggagc agcagatatt
2701 gatcataaac ttgaagaaca actgattgat ggtattcttt atgctttcca agaacagact
2761 acagaggact cagtaatggt gaacggcttt ggcacagtgg ttaatgctct tggcaaacga
2821 gtcaaaccat acttgcctca gatctgtggt acagttttgt ggcgtttaaa taacaaatct
2881 gctaaagtta ggcaacaggc agctgacttg atttctcgaa ctgctgtgt catgaagact
2941 tgtcaagagg aaaaattgat gggacacttg ggtgtgtat tgtatgagta tttgggtgaa
3001 gagtaccctg aagtattggg cagcattctt ggagcactga aggccattgt aaatgtcata
3061 ggtatgcata agatgactcc accaattaaa gatctgctgc ctgactcac ccccatctta
3121 aagaacagac atgaaaaagt acaagagaat tgtattgatc ttgttggctg tatttgcagc
3181 aggggagctg aatatgtatc tgcaagagag tggatgagga tttgcttga gcttttagag
3241 ctcttaaaag cccacaaaaa ggctattcgt agagccacag tcaacacatt tgggtatatt
3301 gcaaaggcca ttggccctca tgatgtattg gctacacttc tgaacaacct caaagttcaa
3361 gaaaggcaga acagagtttg taccactgta gcaatageta ttgttgaga aacatgttca
3421 ccctttacag tactccctgc cttaatgaat gaatacagag ttctgaaact gaatgttcaa
3481 aatggagtgt taaaatcgtt ttccctcttg tttgaatata ttggtgaaat gggaaaagac
3541 tacatttatg ccgtaaacacc gttacttgaa gatgctttaa tggatgagaa ccttggtagc
3601 agacagacgg ctagtgcagt ggtacagcac atgtcacttg gggtttatgg atttggttgt
3661 gaagattcgc tgaatcactt gttgaactat gtatggccca atgtatttga gacatctcct
3721 catgtaattc aggcagttat gggagcccta gagggcctga gagttgctat tggaccatgt
3781 agaatggtgc aatattgttt acagggctctg tttcaccag ccgggaaagt cagagatgta
3841 tattggaaaa ttacaaactc catctacatt ggttcccagg acgctctcat agcacattac
3901 ccaagaatct acaacgatga taagaacacc tatattcgtt atgaacttga ctatatctta
3961 taattttatt gtttattttg tgtttaatgc acagctactt cacaccttaa acttgctttg
4021 atttggtgat gtaaaccttt aaacattgca gatcagtgta gaactggtea tagaggaaga
4081 gctagaaatc cagtagcatg atttttaaat aacctgtctt tgtttttgat gttaaacagt
4141 aatgcccagt agtgaccaag aacacagtga ttatatacac tatactggag ggatttcatt
4201 ttttaattcat ctttatgaag atttagaact cattccttgt gtttaagggg aatgtttaat
4261 tgagaaataa acatttgtgt acaaaatgct aaaaaaaaaa aaaaaaaaaa aaaa

```

El factor de empalme 3b de *Homo sapiens*, subunidad 1, 155 kDa, *SF3B1*, transcrito variante 2, está codificado por la siguiente secuencia de ARNm (NM_001005 526, SEQ ID NO: 18):

5

```

1 ggaagttctt gggagcgeca gtcccgctctg tgtgttcgag tggacaaaat ggcgaagatc
61 gccaaagactc acgaagatat tgaagcacag attcgagaaa ttcaaggcaa gaaggcagct
121 cttgatgaag ctcaaggagt gggcctcgat tctacaggtt attatgacca ggaaatttat
181 ggtggaagtg acagcagatt tgctggatac gtgacatcaa ttgctgcaac tgaacttgaa
241 gatgatgacg atgactattc atcatctacg agtttgcctg gtcagaagaa gccaggatat
301 catgccccctg tggcattgct taatgatata ccacagtcaa cagaacagta tgatccattt
361 gctgagcaca gacctcaaaa gattgcagac cgggaagatg aatacaaaaa gcataggcgg
421 accatgataa tttcccaga gcgtcttgat ccttttgagc atggcttcta ttctgctgct
481 tgaagtcaga actgctgatg gagacaaagg cacgaaagtg tacgtattcc ggattagcaa
541 ccaggaacc catcacttct gaagactcta aactgtgctg tcattttggt tttatatgca
601 ttaaaatatt tgttttaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa

```

El factor de empalme 3b de *Homo sapiens*, subunidad 1, 155 kDa, *SF3B1*, transcrito variante 1, está codificado por la siguiente secuencia de aminoácidos (NP_036565.2, SEQ ID NO: 19) (la parte que contiene los dominios HEAT3, HEAT4 y HEAT5 está subrayada):

10

```

1 makiakthed ieaqireiqg kkaaldeagq vgldestgyd qeiyygsdsr fagyvtsiaa
61 teledddddyy ssstslgqk kpgyhapval lndipqsteg ydpfaehrpp kiadredeyk
121 khrrtmiisp erldpfadgg ktpdpkmnar tymdvmreqh ltkeereirq qlaekakage
181 lkvvngaaas qppskrkrrw dqtadqtpga tpkklsswdq aetpghtpsl rwdetpgrak
241 gsetpgatpg skiwdptpsh tpagaatpgr gdtpphatpg hggatssark nrwdetpkte
301 rdtppghsgw aetprtdrpg dsigetptpg askrksrwe tpasqmggst pvltpgktpi

```

```

361 gtpammatp tpghimsmtp eqlgawrwer eidernrpls deeldamfpe gykvlpppag
421 yvpirtpark ltatptplgg mtgfhmqted rtmksvndqp sgnlpflkpd diqyfdkllv
481 dvdestlspe eqkerkimkl llkikngtpp mrkaalrqit dkarefgagg lfnqilpllm
541 sptledqerh llvkvidril yklddlvrpy vkhilvviép llidedyyar vegreiisnl
601 akaaglatmi stmrpdidnm deyvrrnttar afavvasalg ipsllpflka vckskkswqa
661 rhtgikivqg iailmgcail phlrslveii ehglvdeqgk vrtisalaia alaeaatpyg
721 iesfdsvlkp lwkqirghrg kglaaflkai gyliplmdae yanyytrevm lilirefqsp
781 deemkkivlk vvkqccgtgd veanyikteí lppffkhfwq hrmaldrny rqlvdtvel
841 ankvgaeeii srivddlkdé aeqyrkmvme tiekingnlg aadidhkleé qlidgilyaf
901 qeqttedsvm lngfgtvvna lgkrvkpylp qicgtvlwrl nnksakvrqq aadlisrtav
961 vmktcqeekl mghlgvvlve ylgeeypevl gsilgalkai vnvigmhkmt ppikdllprl
1021 tpilknrhek vqencidlvq riadrgaeyv sarewmricf ellellkakh kairratvnt
1081 fgyiakaigp hdvlatllnn lkvqerqnrvt cttvaiaiva etcspftvlp almneyrvpe
1141 lnvqngvlks lsflfeyige mgkdyiyavt plledalmdr dlvrhrtasa vvqhmslgvy
1201 gfgcedslnh llnyvwpnvf etsphviqav mgaleglrva igpcrmlqyc lqglfhpark
1261 vrdvywkiyn siyigsqdal iahypriynd dkntyiryel dyil

```

El factor de empalme 3b de *Homo sapiens*, subunidad 1, 155 kDa, SF3B1, transcrito variante 2, está codificado por la siguiente secuencia de aminoácidos (NP_001005526, SEQ ID NO: 20):

5

```

1 makiakthed ieaqireiqg kkaaldeagg vgldestggyd qeiyygsdsr fagyvtsiaa
61 teledddddy ssstslggk kpgyhapval lndipqsteq ydpfaehrpp kiadredeyk
121 khrrtmiisp erldpfadgf ysaa

```

En algunas realizaciones no limitantes de la divulgación, se puede utilizar un método *in vitro* desvelado para analizar secuencias de SF3B1 y alteraciones de secuencia como se desvelan en el presente documento sin que incluya un acto de diagnóstico o tratamiento médico.

10

Otras características y ventajas de la divulgación serán evidentes a partir de y están englobadas en la siguiente descripción y las reivindicaciones.

15 Breve descripción de los dibujos

La Figura 1 es un diagrama esquemático del gen SF3B1 humano (arriba) y la proteína (abajo) con sus dominios funcionales (dominio de unión PPP1R8 y repeticiones HEAT). Se proporciona el alineamiento múltiple correspondiente de las secuencias de aminoácidos HEAT3, HEAT4 y HEAT5 de la proteína humana SF3B1 con proteínas SF3B1 ortólogas (n = 15). Los aminoácidos que se conservan entre especies están resaltados (gris). Las formas codificadas por colores indican la posición de mutaciones que se encuentran en la CLL en el momento del diagnóstico (verde, mutaciones sin sentido correspondientes a K666E, K700E, y G740E; véase también la Tabla 4), en la CLL refractaria a fludarabina (rojo, mutaciones sin sentido correspondientes a R630S, H662Q, H662D, K666E, y K700E; eliminaciones en fase correspondientes a de1Q699_K700; véase también la Tabla 4), y en el síndrome de Richter (RS) (naranja, mutaciones sin sentido correspondientes con Q670E y K700E en el mismo, véase también la Tabla 4). Se proporcionan las siguientes secuencias:

20

25

```

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKQIRQHRGK

```

30 (*Homo sapiens*, SEQ ID NO: 1),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*P.*
trogodytes, SEQ ID NO: 2),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*C.*
familiaris, SEQ ID NO: 3),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*B.*
tauris, SEQ ID NO: 4),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*M.*
musculus, SEQ ID NO: 5),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*R.*
norvegicus, SEQ ID NO: 6),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*G.*
gallus, SEQ ID NO: 7),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLRSL
VEIIHGLVDEQQKVRTISALAI AALAEAATPYGIESFDSVLKPLWKGIRQHRGK (*D.*
rerio, SEQ ID NO: 8),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLKAL
VEIIHGLVDEQQKVRTITALAI AALAEAATPYGIESFDSVLKPLWKGIRTHRGK (*D.*
melanogaster, SEQ ID NO: 9),

ARAFAVVASALGIPSLLPFLKAVCKSKKSWQARHTGIKIVQQIAILMGCAILPHLKSL
VEIIHGLVDEQQKVRTITALAL AALAEAATPYGIESFDSVLKPLWKGIRTHRGK (*A.*
gambiae, SEQ ID NO: 10),

ARAFAVVASALGIPALLPFLKAVCKSKKSWQARHTGIKIVQQMAILMGCAVLPHLK
ALVDIVESGLDDEQQKVRTITALCLAALAEASSPYGIEAFDSVLKPLWKGIRMHRGK
(*C. elegans*, SEQ ID NO: 11),

ARAFSVVASALGVPALLPFLKAVCRSKKSWQARHTGVRIIQQIALLLGCSILPHLKNL
VDCIGHGLEDEQQKVRIMTALSLSALAEAATPYGIEAFDSVLKPLWSGVQRHRGK (*S.*

Pombe, SEQ ID NO: 12),

ARAFVVASALGIPALLPFLQAVCRSKKSWQARHTGVKIVQQIPLMGC AVLPHLKR
LVDCIGPNLNDEQTKVVRTVTS LAIAALAEAA NPYGIESFDDILNPLWTGARKQRGK

(*M. oryzae*, SEQ ID NO: 13),

ARAFVVASALGIPALLPFLRAVCRSKKSWQARHTGVKIVQQIPLMGC AVLPHLKQ
LVDCIGPNLNDEQTKVVRTVTS LAIAALAEASNPYGIESFDDILNPLWTGARKQRGK

(*M. crassa*, SEQ ID NO: 14),

ARAFSVVASALGIPALLPFLKAVCQSKR SWQARHTGIKIVQQIAILIGCAVLPHLRLSV
EIIHGLSDENQKVRTTITSLAALAEAAAPY GIESFDSVLKPLWKGIRSHRGK (*A.*

thaliana, SEQ ID NO: 15), and

ARAFSVVASALGTPALLPFLKAVCQSKK SWQARHTGIKIVQQIAILMGC AVLPHLKS
LVEIIHGLSDENQKVRTTITSLATLAEAAAPY GIESFDTVLKPLWKGIRSHRGK (*O.*

Sativa, SEQ ID NO: 16).

Las Figuras 2A-D es una serie de gráficos que representan la prevalencia, relación mutua con otras lesiones genéticas, e impacto clínico de las mutaciones de SF3B1 en la CLL. Panel A: prevalencia de las mutaciones SF3B1 en CLL en el momento del diagnóstico, en CLL refractaria a fludarabina, y síndrome de Richter; los números superiores indican el número actual de muestras mutadas respecto al número total analizado. Panel B: relación mutua de mutaciones SF3B1 con otras lesiones genéticas en CLL en el momento del diagnóstico y en CLL refractaria a fludarabina. En el mapa caliente, las flechas corresponden con genes idénticos, y las columnas representan pacientes individuales basándose en el código de color del estatus genético (blanco: tipo silvestre; rojo: mutaciones SF3B1, mutaciones de NOTCH1, mutaciones y/o eliminación de TP53, eliminación de ATM). Panel C: Estimaciones de Kaplan-Meier de supervivencia libre de tratamiento (TFS) y supervivencia total (OS) a partir del diagnóstico en la serie consecutiva de nuevos diagnosticados y CLL sin tratar previamente (n = 301). El SF3B1 tipo silvestre (SF3B1 ts) está representado por la línea azul. Los casos de SF3B1 mutado (SF3B1 M) están representados por la línea roja. Panel D: Niveles de expresión genética de BCL6, AICDA, BCL2, IRF4 y SF3B1 en subpoblaciones de células B normales (Nativo; Centroblastos, CB; Centrocitos, CC; Memoria) y muestras de CLL. Los niveles relativos de expresión genética se representan con una escala de color: el rojo representa el nivel de expresión más alto y el azul representa el nivel más bajo.

Descripción detallada de los modos para practicar la divulgación

Las lesiones genéticas que se identifican en la leucemia linfocítica crónica (CLL) no recapitulan completamente la patogénesis de la enfermedad y el desarrollo de complicaciones graves, tales como la quimiorrefractoriedad. Al investigar el genoma codificante de la CLL refractaria a fludarabina, se descubrió que las mutaciones de SF3B1, que codifican un factor de empalme y representan un componente crítico del empalmosoma celular, eran recurrentes en 10/59 (un 17 %) de casos refractarios a fludarabina, con una frecuencia significativamente más alta que la que se observa en una cohorte de CLL consecutiva muestreada en el momento del diagnóstico (17/301, un 5 %; p = 0,02). Las mutaciones que se adquirían somáticamente, estaban representadas por cambios de nucleótido sin sentido, agrupadas en repeticiones HEAT seleccionadas de la proteína SF3B1, dirigidas recurrentemente a tres puntos clave (codones 662, 666 y 700), y predecían un pronóstico malo. En la CLL refractaria a fludarabina, las mutaciones SF3B1 y la alteración de TP53 se distribuían de una manera mutuamente excluyente (p = 0,046). La identificación de las mutaciones SF3B1 indica que la regulación de corte y empalme es un nuevo mecanismo patogénico de relevancia crítica en la CLL.

El curso clínico de la leucemia linfocítica crónica (CLL) varía desde un trastorno muy indolente con una esperanza de vida normal, hasta una enfermedad rápidamente progresiva que da lugar a la muerte. De manera ocasional, la CLL sufre una transformación histológica hacia el síndrome de Richter (RS) (Müller-Hermelink HK et al. En: Swerdlow SH et al. eds. World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC; 2008:180-182; Hallek M, et al. Blood. 2008; 111(12):5446-5456; Rossi D, et al. Blood. 2011; 117(12):3391-3401). El curso clínico variable de CLL está dirigido, al menos en parte, por la heterogeneidad inmunogenética y molecular de la enfermedad (Chiorazzi N, et al. N Engl J Med. 2005; 352(8):804-815).

A pesar de los recientes avances, las lesiones genéticas identificadas hasta la fecha no recapitulan completamente la patogénesis molecular de CLL y no explican completamente el desarrollo de las graves complicaciones, tales como la quimiorrefractoriedad, que siguen representando una necesidad clínica no satisfecha (Kay NE, et al. *Leukemia* 2007; 21(9):1885-1891). La refractoriedad a la fludarabina es debida a la alteración de TP53 en ~40 % de casos refractarios, pero en una fracción considerable de pacientes, la base molecular de su fenotipo clínico agresivo sigue sin esclarecerse (Stilgenbauer S y Zenz T. *Hematology Am Soc Hematol Educ Program*. 2010; 2010:481-488).

Recientemente, dos estudios independientes del genoma codificante de CLL investigado en el momento de la presentación de la enfermedad han revelado un número restringido de genes mutados, que incluyen *NOTCH1* (Fabbri G, et al. *J Exp Med*. 2011; 208(7):1389-1401; Puente XS, et al. *Nature*. 2011; 475(7354):101-105). Estos estudios han proporcionado una prueba del concepto que, al igual que en otros cánceres malignos, el análisis mutacional amplio del genoma puede identificar nuevas lesiones de relevancia potencial biológica y clínica en CLL. Después de los hallazgos iniciales a partir de la secuenciación de todo el exoma del genoma codificante de CLL refractaria a fludarabina, se desvela en el presente documento la existencia de mutaciones recurrentes de *SF3B1*, un componente crítico del empalmosoma celular.

Las expresiones “ácido nucleico” y “polinucleótido” se utilizan en el presente documento de manera intercambiable para referirse a un ARN, ADN de cadena sencilla o doble, o polímeros mixtos. Los polinucleótidos pueden incluir secuencias genómicas, extra-genómicas y de plásmido, y segmentos genéticos modificados más pequeños que expresan, o se pueden adaptar para que expresen polipéptidos.

Un “ácido nucleico aislado” es un ácido nucleico que está separado sustancialmente de otras secuencias de ADN del genoma así como proteínas o complejos tales como ribosomas y polimerasas, que acompañan naturalmente a una secuencia nativa. El término abarca una secuencia de ácido nucleico que se ha retirado de su ambiente natural de origen, e incluyen aislados de ADN recombinante o clonado y análogos sintetizados químicamente o análogos sintetizados biológicamente por sistemas heterólogos. Un ácido nucleico sustancialmente puro incluye formas aisladas del ácido nucleico. Esto hace referencia al ácido nucleico como se aísla originalmente y no excluye genes o secuencias añadidas más tarde al ácido nucleico aislado por manipulación. Además de la preparación de las moléculas de ácido nucleico como se ha descrito anteriormente, la divulgación incluye la preparación de moléculas de ácido nucleico por inmovilización directa o indirecta en un soporte sólido o medio de fase sólida. La inmovilización directa puede ser mediante enlaces de hidrógeno, tal como en el caso de hibridación como un ejemplo no limitante, o estar mediada por uno o más enlaces covalentes. Ejemplos no limitantes incluyen la hibridación de moléculas de ácido nucleico a una sonda de polinucleótido en una micromatriz o una perla u otro soporte sólido para detectar una molécula de ácido nucleico de interés. Opcionalmente, las moléculas de ácido nucleico hibridadas pueden ser las que se han amplificado por PCR. La inmovilización indirecta de una molécula de ácido nucleico puede ser mediante la unión a una polimerasa inmovilizada, tal como una ARN polimerasa o ADN polimerasa. En realizaciones adicionales, una molécula de ácido nucleico se puede preparar por secuenciación por unión a una secuencia de ácido nucleico conocida o unión a un polinucleótido cebador por complementariedad de pares de bases. En algunas realizaciones, una molécula de ácido nucleico se puede secuenciar sin necesidad de amplificación o replicación. En realizaciones adicionales, una molécula de ácido nucleico que se prepara puede ser una molécula de ARN que se ha marcado para que pueda detectarse para ayudar a su análisis o una molécula de ARN que se ha convertido en una molécula de ADNc para su uso como se ha descrito en el presente documento.

El término “polipéptido” se utiliza en su significado convencional, es decir, como una secuencia de aminoácidos. Los polipéptidos no se limitan a una longitud específica del producto. Los péptidos, oligopéptidos y proteínas están incluidos en la definición de polipéptido, y dichos términos se pueden utilizar en el presente documento de manera intercambiable a menos de que específicamente se indique otra cosa. Este término tampoco es que no se refiera o excluya las modificaciones post-expresión del polipéptido, por ejemplo, glicosilaciones, acetilaciones, fosforilaciones, y similares, así como otras modificaciones conocidos en la técnica, tanto de origen natural como de origen no natural. Un polipéptido puede ser una proteína completa, o una sub-secuencia de la misma.

Un “polipéptido aislado” es el que se ha identificado y separado y/o recuperado de un componente de su entorno natural. En algunas realizaciones, el polipéptido aislado estará purificado (1) hasta más de un 95 % por peso de polipéptido como se determina por el método de Lowry, y más preferentemente más del 99 % por peso, (2) hasta un grado suficiente para obtener al menos 15 restos de secuencia de aminoácidos del extremo N o interna por el uso de un secuenciador de copa giratoria, o (3) hasta la homogeneidad por SDS-PAGE en condiciones reductoras o no reductoras utilizando azul Coomassie o, preferentemente tinte de plata. El polipéptido aislado incluye el polipéptido *in situ* en células recombinantes en las que al menos un componente del entorno natural de polipéptido no está presente. Habitualmente, sin embargo, el polipéptido aislado se preparará mediante al menos una etapa de purificación.

Un polinucleótido de “secuencia nativa” es el que tiene la misma secuencia de nucleótidos que el polinucleótido derivado de la naturaleza. Un polipéptido de “secuencia nativa” es el que tiene la misma secuencia de aminoácidos que un polipéptido (por ejemplo, una subunidad proteica) derivado de la naturaleza (por ejemplo, de cualquier especie). Dichos polinucleótidos y polipéptidos de secuencia nativa se pueden aislar de la naturaleza o se pueden producir por medios recombinantes o sintéticos.

Un polinucleótido “mutante o mutado”, como se utiliza la expresión en el presente documento, es un polinucleótido que normalmente se diferencia de un polinucleótido que se desvela específicamente en el presente documento en una o más sustituciones, eliminaciones, adiciones y/o inserciones. Dichas variantes pueden ser de origen natural o pueden generarse sintéticamente, por ejemplo, modificando una o más de las secuencias de polinucleótido de la divulgación y evaluando una o más actividades biológicas del polipéptido codificado como se describe en el presente documento y/o utilizando cualquiera de varias técnicas bien conocidas en la técnica. El polinucleótido “mutante” de la divulgación contiene una o más sustituciones, eliminaciones, adiciones y/o inserciones que alteran la función del polipéptido resultante codificado por el mismo. De manera alternativa, o adicionalmente, los polinucleótidos “modificados” de la divulgación contienen una o más sustituciones, eliminaciones, adiciones y/o inserciones que no alteran la función del polipéptido resultante codificado por el mismo. En algunas realizaciones, un polinucleótido “mutante o mutado” se define en referencia a una secuencia de tipo silvestre como se ha desvelado en el presente documento. Adicionalmente, un polinucleótido “mutante o mutado” se puede preparar, y opcionalmente detectar, de la misma manera que otros polinucleótidos desvelados en el presente documento.

Un polipéptido “mutante o mutado”, como se utiliza la expresión en el presente documento, es un polipéptido que se diferencia de un polipéptido desvelado específicamente en el presente documento en una o más sustituciones, eliminaciones, adiciones y/o inserciones. Dichas variantes pueden producirse naturalmente o generarse sintéticamente, por ejemplo, modificando una o más de las secuencias polipeptídicas anteriores de la divulgación y evaluando una o más actividades biológicas del polipéptido como se ha descrito en el presente documento y/o utilizando cualquiera de las distintas técnicas bien conocidas en la técnica. Los polipéptidos “mutantes” de la divulgación contienen una o más sustituciones, eliminaciones, adiciones y/o inserciones que alteran la función del polipéptido resultante. De manera alternativa, o adicionalmente, los polipéptidos “modificados” de la divulgación contienen una o más sustituciones, eliminaciones, adiciones y/o inserciones que no alteran la función del polipéptido resultante. En algunas realizaciones, un polipéptido “mutante o mutado” se define en referencia a una secuencia de tipo silvestre como se desvela en el presente documento. Adicionalmente, un polipéptido “mutante o mutado” se puede analizar o detectar por cualquier método conocido por el experto. Ejemplos no limitantes incluyen secuenciación peptídica, análisis por espectrometría de masas, y unión con anticuerpos o receptores.

Se pueden hacer modificaciones en la estructura de los polinucleótidos y polipéptidos, de tipo silvestre o mutantes, de la presente divulgación y seguir obteniendo una molécula funcional que codifique una variante o derivado polipeptídico con características deseables. Cuando se desea alterar la secuencia de aminoácidos de un polipéptido para crear un equivalente, o incluso una mejora, variante o parte de un polipéptido de la divulgación, un experto en la técnica normalmente cambiará uno o más de los codones de la secuencia de ADN codificante.

Por ejemplo, se pueden sustituir ciertos aminoácidos por otros aminoácidos en una estructura proteica sin una pérdida apreciable de su capacidad para unirse con otros polipéptidos o células. Debido a que es la capacidad de unión y la naturaleza de una proteína lo que define la actividad funcional biológica de una proteína, se pueden hacer sustituciones de ciertos aminoácidos en la secuencia de una secuencia proteica, y, por supuesto en la secuencia de ADN codificante subyacente de la proteína, y sin embargo obtener una proteína con propiedades parecidas. Se contempla por lo tanto, que se pueden hacer varios cambios en las secuencias peptídicas de las composiciones desveladas, o las secuencias de ADN correspondientes que codifican dichos péptidos sin pérdida apreciable de su utilidad o actividad biológica.

En muchos casos, un polipéptido modificado contendrá una o más sustituciones conservadoras. Una “sustitución conservadora” es en la que se sustituye un aminoácido por otro aminoácido que tiene propiedades similares, de manera que un experto en la técnica de la química de péptidos esperaría que la estructura secundaria y la naturaleza hidropática del polipéptido no cambien sustancialmente.

Al hacer dichos cambios, se puede considerar el índice hidropático de los aminoácidos. La importancia del índice hidropático de los aminoácidos para conferir la función biológica interactiva de una proteína se entiende en general en la técnica (Kyte y Doolittle, 1982). Se acepta que el carácter hidropático relativo del aminoácido contribuye a la estructura secundaria de la proteína resultante, lo que a su vez define la interacción de la proteína con otras moléculas, por ejemplo, enzimas, sustratos, receptores, ADN, anticuerpos, antígenos, y similares. A cada aminoácido se le ha asignado un índice hidropático basándose en sus características de hidrofobia y de carga (Kyte y Doolittle, 1982). Estos valores son: isoleucina (+4,5); valina (+4,2); leucina (+3,8); fenilalanina (+2,8); cisteína/cistina (+2,5); metionina (+1,9); alanina (+1,8); glicina (-0,4); treonina (-0,7); serina (-0,8); triptófano (-0,9); tirosina (-1,3); prolina (-1,6); histidina (-3,2); glutamato (-3,5); glutamina (-3,5); aspartato (-3,5); asparagina (-3,5); lisina (-3,9); y arginina (-4,5).

Se sabe en la técnica que se pueden sustituir ciertos aminoácidos por otros aminoácidos que tienen un índice hidropático o valor similar y seguir dando como resultado una proteína con una actividad biológica similar, es decir, se sigue obteniendo una proteína funcional biológicamente equivalente. Al hacer dichos cambios, se prefiere la sustitución de aminoácidos cuyos índices hidropáticos tienen ± 2 , los que tienen ± 1 son particularmente preferidos, y los que tienen $\pm 0,5$ son incluso más particularmente preferidos. También se entiende en la técnica que la sustitución de aminoácidos parecidos se puede hacer eficazmente basándose en la hidrofilia. La patente de EE. UU. 4.554.101 establece que la media local mayor de hidrofilia de una proteína, que se regula por la hidrofilia de sus aminoácidos

adyacentes, se correlaciona con una propiedad biológica de la proteína.

Como se detalla en la Patente de EE. UU. 4.554.101, se han asignado los siguientes valores de hidrofilia a los restos de aminoácidos: arginina (+3,0); lisina (+3,0); aspartato (+3,0 ± 1); glutamato (+3,0 ± 1); serina (+0,3); asparagina (+0,2); glutamina (+0,2); glicina (0); treonina (-0,4); prolina (-0,5 ± 1); alanina (-0,5); histidina (-0,5); cisteína (-1,0); metionina (-1,3); valina (-1,5); leucina (-1,8); isoleucina (-1,8); tirosina (-2,3); fenilalanina (-2,5); triptófano (-3,4). Se entiende que un aminoácido se puede sustituir por otro que tenga un valor de hidrofilia similar y seguir obteniendo una proteína biológicamente equivalente, y en particular, una proteína inmunológicamente equivalente. En dichos cambios, la sustitución de aminoácidos cuyos valores de hidrofilia tienen ±2 se prefieren, los que tienen ±1 son particularmente preferidos, y los que tienen ±0,5 son incluso más particularmente preferidos.

Como se ha descrito anteriormente, las sustituciones de aminoácidos se basan generalmente por lo tanto en la similitud relativa de los sustituyentes de cadenas laterales del aminoácido, por ejemplo, su hidrofobia, hidrofilia, carga, tamaño y similares. Las sustituciones ejemplares que toman en consideración varias de las características anteriores son bien conocidas por los expertos en la técnica e incluyen: arginina y lisina; glutamato y aspartato; serina y treonina; glutamina y asparagina; y valina, leucina e isoleucina.

Las sustituciones de aminoácidos se pueden hacer adicionalmente basándose en la similitud de la polaridad, carga, solubilidad, hidrofobia, hidrofilia y/o la naturaleza anfipática de los restos. Por ejemplo los aminoácidos cargados negativamente incluyen, el ácido aspártico y ácido glutámico; los aminoácidos cargados positivamente incluyen lisina y arginina; y los aminoácidos con grupos en cabeza sin carga polar que tienen valores de hidrofilia similares incluyen leucina, isoleucina y valina; glicina y alanina; asparagina y glutamina; y serina, treonina, fenilalanina y tirosina. Otros grupos de aminoácidos que pueden representar cambios conservadores incluyen: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; y (5) phe, tyr, trp, his. Un polipéptido modificado puede también, o alternativamente, contener cambios no conservadores. En una realización preferida, las variantes de polipéptido se diferencian de una secuencia nativa por sustitución, eliminación o adición de cinco aminoácidos o menos. Los polipéptidos también (o alternativamente) se pueden modificar, por ejemplo, por eliminación o adición de aminoácidos que tengan una influencia mínima sobre la inmunogenicidad, estructura secundaria y naturaleza hidropática del polipéptido.

Los polipéptidos pueden comprender una secuencia de señal (o líder) en el extremo N de la proteína, que dirige co-traduccionally o post-traduccionally la transferencia de la proteína. El polipéptido también se puede conjugar con un enlazador u otra secuencia para facilitar la síntesis, purificación o identificación del polipéptido (por ejemplo, poli-His), o para aumentar la unión del polipéptido a un soporte sólido. Por ejemplo, se puede conjugar un polipéptido a una biotina, estreptavidina, o Fc de inmunoglobulina.

Cuando se comparan secuencias de polinucleótido y polipéptido, se dice que dos secuencias son "idénticas" si la secuencia de nucleótidos o aminoácidos de las dos secuencias es la misma cuando se alinean con una correspondencia máxima, como se describe posteriormente. Las comparaciones entre dos secuencias se llevan a cabo normalmente comparando las secuencias en una ventana de comparación para identificar y comparar regiones locales de similitud de secuencia. Una "ventana de comparación" como se utiliza en el presente documento, se refiere a un segmento de al menos aproximadamente 20 posiciones contiguas, habitualmente de 30 a aproximadamente 75, de 40 a aproximadamente 50, en el que se puede comparar una secuencia con una secuencia de referencia del mismo número de posiciones contiguas tras alinear óptimamente las dos secuencias.

El alineamiento óptimo de secuencias para la comparación se puede llevar a cabo utilizando el programa Megalign del paquete de software bioinformático Lasergene (DNASTAR, Inc. Madison, WI), utilizando los parámetros por defecto. Este programa lleva a cabo varios esquemas de alineamiento que se describen en las siguientes referencias: Dayhoff, M.O. (1978) A model of evolutionary change in proteins - Matrices for detecting distant relationships. En Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenesis pp. 626-645 Methods. En Enzymology vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. y Sharp, P.M. (1989) CABIOS 5:151-153; Myers, E.W. y Muller W. (1988) CABIOS 4:11-17; Robinson, E.D. (1971) Comb. Theor 11:105; Santou, N. Nes, M. (1987) Mol. Biol. Evol. 4:406-425; Sneath, P.H.A. y Sokal, R.R. (1973) Numerical Taxonomy - the Principles and Practice of Numerical Taxonomy, Freeman Press, San Francisco, CA; Wilbur, W.J. y Lipman, D.J. (1983) Proc. Natl. Acad. Sci. USA 80:726-730.

De manera alternativa, el alineamiento óptimo de secuencias para la comparación se puede llevar a cabo por el algoritmo de identidad local de Smith y Waterman (1981) Add. APL. Match 2:482, por el algoritmo de identidad de alineamiento de Needleman y Wunsch (1970) J. Mol. Biol. 48:443, por los métodos de búsqueda de similitud de Pearson y Lipman (1988) Proc. Natl. Acad. Sci. USA 85: 2444, por implementación computarizada de estos algoritmos (GAP, BESTFIT, BLAST, FASTA, y TFASTA del paquete de software de Wisconsin Genetics, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), o por inspección.

Un ejemplo no limitante de algoritmos que son adecuados para determinar el porcentaje de identidad de secuencia y similitud de secuencia son los algoritmos BLAST y BLAST 2.0, que se describen en Altschul et al. (1977) Nucl. Acids

Res. 25:3389-3402 y Altschul et al. (1990) J. Mol. Biol. 215:403-410, respectivamente. Se pueden utilizar el BLAST y BLAST 2.0, por ejemplo, con los parámetros descritos en el presente documento, para determinar el porcentaje de identidad de secuencia para los polinucleótidos y polipéptidos de la divulgación. El software para llevar a cabo los análisis BLAST está disponible para el público por medio de la información del Centro Nacional de Biotecnología.

5 En un ejemplo ilustrativo, se pueden calcular los valores acumulados utilizando, para las secuencias de nucleótidos, los parámetros M (valor de búsqueda para un par de restos coincidentes; siempre > 0) y N (valor de penalización por restos no coincidentes; siempre < 0). La extensión del alcance de la palabra en cada dirección se detiene cuando: el valor de alineamiento acumulado cae por debajo de una cantidad X de su valor máximo alcanzado; el valor acumulado llega a 0 o por debajo, debido a la acumulación de uno o más alineamientos de restos de valor negativo; 10 o cuando se alcanza el final de secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y velocidad del alineamiento. El programa BLASTN (para secuencias de nucleótidos) utiliza como parámetros por defecto una longitud de palabra (W) de 11, y una expectativa (E) de 10, y la matriz de valoración de alineamiento BLOSUM62 (véase Henikoff y Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915), (B) de 50, expectativas (E) de 15 10, M = 5, N = 4 y una comparación de ambas cadenas.

Para las secuencias de aminoácidos, se puede utilizar una matriz de valoración para calcular la valoración acumulada. La extensión del alcance de la palabra en cada dirección se detiene cuando: el valor acumulado de alineamiento cae por debajo de una cantidad X de su valor máximo alcanzado; el valor acumulado llega a cero o por 20 debajo, debido a la acumulación de uno o más alineamientos de restos negativos; o se alcanza el final de cada secuencia. Los parámetros del algoritmo BLAST W, T y X determinan la sensibilidad y la velocidad del alineamiento.

En una estrategia, el "porcentaje de identidad de secuencia" se determina comparando dos secuencias alineadas óptimamente en una ventana de comparación de al menos 20 posiciones, donde la parte de la secuencia de polinucleótido o polipéptido en la ventana de comparación puede comprender adiciones o eliminaciones (es decir, huecos) de un 20 por ciento o menos, habitualmente de un 5 a 15 por ciento, o un 10 a 12 por ciento, en comparación con secuencias de referencia (las cuales no comprenden adiciones o eliminaciones) por el alineamiento óptimo de las dos secuencias. El porcentaje se calcula determinando el número de posiciones en las que existen bases de ácido nucleico o restos de aminoácidos idénticos en ambas secuencias para dar lugar al 30 número de posiciones coincidentes, dividiendo el número de posiciones coincidentes por el número total de posiciones de la secuencia de referencia (es decir, el tamaño de la ventana) y multiplicando los resultados por 100 para dar lugar al porcentaje de identidad de secuencia.

"Homología" se refiere al porcentaje de restos en la variante de secuencia de polinucleótido o polipéptido que es idéntico a la secuencia no variante tras el alineamiento de las secuencias e introduciendo huecos, si fuera necesario, para conseguir el máximo porcentaje de homología. En realizaciones particulares, las variantes de polinucleótidos y polipéptidos tienen al menos un 70 %, al menos un 75 %, al menos un 80 %, al menos un 90 %, al menos un 95 %, al menos un 98 %, o al menos un 99 % de homología de polinucleótido o polipéptido con un polinucleótido o polipéptido descrito en el presente documento.

40 Será evidente para los expertos en la técnica que, como resultado de la degeneración del código genético, hay múltiples secuencias de nucleótidos que codifican un polipéptido que se describe en el presente documento. Algunos de estos polinucleótidos tienen una homología mínima con la secuencia de nucleótidos de cualquier gen nativo. Sin embargo, los polinucleótidos que codifican un polipéptido de la presente divulgación pero que varían debido a diferencias en el uso del codón se contemplan específicamente en la divulgación. Además, los alelos de los genes que incluyen secuencias de polinucleótido que se proporcionan en el presente documento están en el alcance de la divulgación. Los alelos son genes endógenos que se alteran como resultado de una o más mutaciones, tales como eliminaciones, adiciones y/o sustituciones de nucleótidos. El ARNm resultante y cualquier proteína pueden, pero no necesariamente, tener una estructura o función alteradas. Los alelos se pueden identificar utilizando técnicas convencionales (tales como hibridación, amplificación y/o comparación de secuencias de bases de datos).

En ciertas realizaciones de la presente divulgación, la mutagénesis de las secuencias de polinucleótido desveladas se lleva a cabo con el fin de alterar una o más propiedades del polipéptido codificado, tales como su especificidad de unión o fuerza de unión. Las técnicas de mutagénesis son bien conocidas en la técnica, y se utilizan ampliamente 55 para crear variantes de polipéptidos y polinucleótidos. Se emplea una estrategia de mutagénesis, tal como la mutagénesis específica del sitio, para la preparación de variantes y/o derivados de los polipéptidos descritos en el presente documento. Con esta estrategia, se hacen modificaciones específicas en una secuencia de polipéptido por medio de mutagénesis de los polinucleótidos subyacentes que la codifican. Estas técnicas proporcionan una estrategia directa para preparar y ensayar variantes de secuencia, por ejemplo, incorporando una o más de las consideraciones anteriores, introduciendo uno o más cambios de secuencia en el polinucleótido.

La mutagénesis específica del sitio permite la producción de mutantes por medio del uso de secuencias de oligonucleótidos específicas que incluyen la secuencia de nucleótidos de la mutación deseada, así como un número suficiente de nucleótidos adyacentes, para proporcionar una secuencia de cebador de suficiente tamaño y 65 complejidad de secuencia para formar un dúplex estable en ambos sitios de la eliminación de unión al ser atravesada. Las mutaciones se emplean en una secuencia de polinucleótido seleccionada para mejorar, alterar,

disminuir, modificar, o cambiar de otra manera las propiedades propias del polinucleótido, y/o alterar las propiedades, actividad, composición, estabilidad, o secuencia primaria del polipéptido codificado.

5 En otras realizaciones de la presente divulgación, las secuencias de polinucleótido que se proporcionan en el presente documento se utilizan como sondas o cebadores para la hibridación de ácidos nucleicos, por ejemplo, cebadores de PCR. La capacidad de dichas sondas de ácido nucleico para hibridarse específicamente a una secuencia de interés las capacita para detectar la presencia de secuencias complementarias en una muestra determinada. Sin embargo, también se engloban otros usos en la divulgación, tal como el uso de la información de secuencia para la preparación de especies de cebadores mutantes, o cebadores para su uso en la preparación de
10 otras construcciones genéticas. Como tales, los segmentos de ácido nucleico de la divulgación que incluyen una región de secuencia de una secuencia contigua de al menos aproximadamente 15 nucleótidos de larga que tiene la misma secuencia, o es complementaria de una secuencia contigua de 15 nucleótidos de larga que se desvela en el presente documento son particularmente útiles. Las secuencias contiguas más largas idénticas o complementarias, por ejemplo, la de aproximadamente 20, 30, 40, 50, 100, 200, 500, 1000 (incluyendo las longitudes intermedias) incluyendo las secuencias de longitud completa, y todas las longitudes entre ellas, también se utilizan en ciertas realizaciones. La divulgación por lo tanto, incluye el uso de secuencias desveladas en el diseño y la preparación de
15 cebadores y sondas de ácido nucleico, tal como para el uso en la amplificación y detección de ácido nucleico como ejemplos no limitantes. En algunas realizaciones, los cebadores se pueden utilizar para secuenciar ácidos nucleicos para detectar una secuencia divulgada, tal como una secuencia mutante.

20 Las moléculas de polinucleótido que tienen regiones de secuencia que consisten en tramos de nucleótidos contiguos de 10-14, 15-20, 30, 50, o incluso de 100-200 nucleótidos o así (incluyendo las longitudes intermedias también), idénticos o complementarios a una secuencia de polinucleótido desvelada en el presente documento, se contemplan particularmente como sondas de hibridación para su uso, por ejemplo en la transferencia de Southern o Northern, y/o cebadores para su uso, por ejemplo, en la reacción en cadena de polimerasa (PCR), PCR cuantitativa, o PCR en
25 tiempo real. El tamaño total del fragmento así como el tamaño del tramo(s), depende en último término del uso que se pretende o la aplicación del segmento de ácido nucleico en particular. Los fragmentos más pequeños se utilizan generalmente en realizaciones de hibridación, donde la longitud de la región complementaria contigua puede variarse, tal como entre aproximadamente 15 y aproximadamente 100 nucleótidos, pero se pueden utilizar tramos contiguos complementarios más grandes, de acuerdo con la longitud de secuencias complementarias que se desea detectar.

30 El uso de una sonda de hibridación de aproximadamente 15-25 nucleótidos de longitud permite la formación de una molécula dúplex que es tanto estable como selectiva. Generalmente se prefieren las moléculas que tienen secuencias complementarias contiguas sobre tramos mayores de 12 bases de longitud, aunque, con el fin de aumentar la estabilidad y selectividad del híbrido, y por lo tanto mejorar la calidad y el grado de moléculas híbridas específicas, se prefieren generalmente las moléculas de ácido nucleico que tienen tramos genéticos complementarios de 15 a 25 nucleótidos contiguos, o incluso más largos si se desea.

35 40 Las sondas de hibridación se seleccionan de cualquier parte de las secuencias desveladas en el presente documento. Todo lo que se necesita es revisar las secuencias que se exponen en el presente documento, o cualquier parte contigua de las secuencias, de aproximadamente 15-25 nucleótidos de longitud hasta la secuencia de longitud completa inclusive, que se desee utilizar como una sonda o cebador. La elección de las secuencias de sonda o cebador está regulada por varios factores. Por ejemplo, se puede desear el empleo de cebadores desde y
45 hacia los extremos de la secuencia total.

50 El polinucleótido de la presente divulgación, o los fragmentos o secuencias modificadas del mismo, se preparan fácilmente, por ejemplo, sintetizando directamente el fragmento por medios químicos, como se practica comúnmente utilizando un sintetizador automático de oligonucleótidos. También, se obtienen los fragmentos por aplicación de tecnología de reproducción de ácidos nucleicos, tal como por la tecnología de PCR™ de la Patente de EE. UU. 4.683.202, introduciendo secuencias seleccionadas en vectores recombinantes para la producción recombinante, y por otras técnicas de ADN recombinante generalmente conocidas por los expertos en la técnica de biología molecular.

55 En algunas realizaciones, los polinucleótidos que se desvelan en el presente documento se pueden preparar por la determinación de secuencia o detección por cualquier método conocido por el experto. Ejemplos no limitantes incluyen la secuenciación basada en: 1) terminadores-colorantes reversibles y anclaje de moléculas de ADN a cebadores en un portaobjetos con amplificación utilizando cuatro tipos de bases de terminador reversibles para extender el ADN solo en un nucleótido cada vez, seguido por la eliminación del colorante junto con el bloqueador del extremo 3' para permitir el nuevo ciclo de extensión; 2) unión de oligonucleótidos inmovilizados de secuencias conocidas seguida por una PCR (opcionalmente PCR en emulsión) y secuenciación; 3) liberación de iones hidrógeno debido a la extensión del nucleótido con detección por semiconductor; 4) secuenciación en nanoesferas; 5) adición de adaptadores de cola de poliA seguida por extensión de nucleótidos y secuenciación; 6) secuenciación de molécula única en tiempo real (SMRT) utilizando polimerasa inmovilizada; 7) secuenciación de señal paralela
60 masivamente (MPSS); 8) secuenciación de Polony; 9) pirosecuenciación por medio de matrices únicas de ADN hibridadas con perlas revestidas de cebador único; y 10) secuenciación mediada por ARN polimerasa (RNAP).

El análisis de la divulgación puede estar precedido o seguido por una variedad de acciones relacionadas. En algunas realizaciones, un análisis está precedido por una determinación o diagnóstico de un sujeto humano que necesita el análisis. El análisis puede estar precedido por una determinación de la necesidad del análisis, tal como por un doctor médico, enfermera u otro empleado o profesional de la salud, o los que trabajan bajo su jurisdicción, o el personal de una aseguradora de salud u organización de mantenimiento que apruebe la realización de la medición en base a un requerimiento de reembolso o pago de la realización. En algunas realizaciones, el análisis puede estar seguido por el pago de la realización de un método desvelado.

El análisis de la divulgación puede estar precedido también por actos de preparación necesarios para el análisis actual. Ejemplos no limitantes incluyen la obtención actual de una muestra de un ser humano que contiene una célula o que contiene un ácido nucleico o que contiene un polipéptido; o la recepción de dicha muestra; o la sección de una célula que contiene la muestra; o el aislamiento de las células de una muestra que contiene células; o la preparación de moléculas de ácido nucleico de células de una muestra que contiene células; o la transcripción inversa de ARN a partir de las células de una muestra que contiene células.

La divulgación proporciona además kits para la práctica de cualquier método desvelado que se describe en el presente documento. Un kit comprenderá normalmente uno o más reactivos para detectar secuencias de ácido nucleico o secuencias de polipéptido que se describen en el presente documento para la práctica de la presente divulgación. Ejemplos no limitantes incluyen sondas o cebadores de polinucleótido para la detección de los niveles de expresión, una o más enzimas que se utilizan en los métodos de la divulgación, y uno o más envases o soportes sólidos o un medio sólido para su uso en la práctica de la divulgación. En algunas realizaciones, el kit incluirá una matriz u otro medio sólido, que incluye perlas como ejemplo no limitante, para la detección de secuencias como se describe en el presente documento. En otras realizaciones, el kit puede comprender uno o más anticuerpos que son inmunorreactivos con los epítopos presentes en un polipéptido lo que indica la presencia de una alteración de secuencia genética como se desvela en el presente documento. En algunas realizaciones, el anticuerpo puede ser un fragmento de anticuerpo.

Un kit de la divulgación puede incluir también, materiales de instrucciones desvelando o describiendo el uso del kit o un cebador o sonda de la presente divulgación en un método de la divulgación como se proporciona en el presente documento. Un kit puede incluir también componentes adicionales para facilitar la aplicación particular para la que se diseñó el kit. Por lo tanto, por ejemplo, un kit puede contener adicionalmente medios para detectar el marcador (por ejemplo, sustratos enzimáticos para marcadores enzimáticos, juegos de filtros para detectar los marcadores fluorescentes, marcadores secundarios adecuados tales como anti-HRP de ratón de oveja, o similares). Un kit puede incluir adicionalmente tampones y otros reactivos reconocidos para su uso en un método de la divulgación. En algunas realizaciones, un kit se puede diseñar para su uso para el diagnóstico *in vitro*.

Habiendo ya proporcionado en general la divulgación, la misma se entenderá más fácilmente en referencia a los siguientes ejemplos que se proporcionan a modo de ilustración, y no se pretende que sean limitantes de la divulgación, a menos de que se especifique.

Ejemplos

Ejemplo 1: Pacientes y métodos

Pacientes

La población de estudio incluye tres cohortes clínicas representativas de diferentes fases de la enfermedad; *i*) CLL refractaria a fludarabina (n = 59), incluyendo casos (n = 11) sometidos a secuenciación del exoma completo (Tabla 1); *ii*) una serie consecutiva de CLL diagnosticados por primera vez y sin tratamiento previo (n = 301) (Tabla 2); y *iii*) RS relacionado clónicamente (n = 33; todos con linfomas de células B grandes difusos) (Tabla 3). El diagnóstico de CLL se basaba en el criterio de IWCLL-NCI (Hallek M, et al. Blood. 2008; 111(12):5446-5456); el diagnóstico de refractariedad a fludarabina era según las directrices (Hallek M, et al. Blood. 2008; 111(12):5446-5456); el RS se basaba en criterios histológicos (Muller-Hermelink HK, et al. Swerdlow SH et al eds. World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC; 2008:180-182; Stein H et al. Swerdlow SH et al. eds. World Health Organization Classification of Tumours, Pathology and Genetics of Tumours of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC; 2008:233-237). Las muestras tumorales en sangre periférica se obtuvieron de la siguiente manera: *i*) para la CLL refractaria a fludarabina, inmediatamente antes de iniciar el tratamiento al que el paciente eventualmente fallaba en la respuesta; *ii*) para los recién diagnosticados y la CLL previamente sin tratar, en la presentación de enfermedad. Todos los estudios de RS se llevaron a cabo en biopsias diagnósticas de RS. Los ADN normales de los mismos pacientes se obtuvieron de saliva o a partir de granulocitos purificados y se confirmó que estaban libres de tumor por PCR de redistribuciones de *IGHV-D-J* específico del tumor. Los pacientes proporcionaban un consentimiento informado de acuerdo con los requerimientos del IRB local y la Declaración de Helsinki. El estudio fue aprobado por el Comité de Ética del Ospedale Maggiore della Carità di Novara asociado con la Universidad Amedeo Avogadro del Piamonte oriental (Código de Protocolo 59/CE; Número de Estudio CE 8/11).

[75] **Tabla 1.** Características clínicas y biológicas de CLL refractaria a fludarabina

	Todos (n = 59)		SF3B1 mutado (n = 10)		SF3B1 ts (n = 49)		P
	Número	%	Número	%	Número	%	
Edad >65 años	37	62,7	6	60,0	31	63,0	1,000
Varón	40	67,8	7	70,0	33	67,3	1,000
Estadio Rai III-IV	28	47,5	7	70,0	21	42,9	0,168
Número de terapias anteriores							0,264
0	26	44,1	4	40,0	22	44,9	
1	24	40,7	6	60,0	18	36,7	
>1	9	15,3	0	0	9	18,4	
Régimen de tratamiento cuando refractariedad							0,750
FCR	17	28,8	4	40,0	13	26,5	
FR	3	5,1	0	0	3	6,1	
FC	19	32,2	2	20,0	17	34,7	
F	20	33,8	4	40,0	16	32,7	
Identidad IGHV ≥ 98 %	48	81,4	8	80,0	40	81,6	1,000
CD38 ≥ 30 %	34	57,6	6	60,0	28	57,1	1,000
ZAP70 ≥ 20 %	39	66,1	6	60,0	33	67,3	0,721
Alteración TP53	23	39,0	1	10,0	22	44,9	0,072
Mutaciones NOTCH1	14	23,7	1	10,0	13	26,5	0,425
Eliminación 11q22-q23	15	25,4	3	30,0	12	24,5	0,704
Trisomía 12	16	27,1	0	0	16	32,7	0,049
Eliminación 13q14	31	52,5	6	60,0	25	51,1	0,734
FISH normal	10	16,9	5	50,0	5	10,2	0,008

^a ts, tipo silvestre; FCR, fludarabina, ciclofosfamida, rituximab; FR, fludarabina, rituximab; FC, fludarabina, ciclofosfamida; F, fludarabina; IGHV, gen de región variable de cadena pesada de inmunoglobulina; FISH, hibridación fluorescente in situ

[76] **Tabla 2.** Características clínicas y biológicas de la serie consecutiva de CLL recién diagnosticada y no tratada previamente

	Todos		SF3B1 mutado		SF3B1 ts		P
	Número	%	Número	%	Número	%	
Edad >65 años	183/301	60,8	13/17	76,5	170/284	59,9	0,173
Varón	163/301	54,2	13/17	76,5	150/284	52,8	0,057
Estadio Rai III-IV	33/301	11,0	7/17	41,2	26/284	9,2	0,001
Identidad IGHV ≥ 98 %	100/294	34,0	8/17	47,1	92/277	33,2	0,242
CD38 ≥ 30 %	81/298	27,2	7/17	41,2	74/281	26,3	0,259
ZAP70 ≥ 20 %	77/253	30,0	8/13	61,5	69/240	28,7	0,025
Alteración TP53	30/301	10,0	1/17	5,9	29/284	10,2	1,000
Mutaciones NOTCH1	34/301	11,3	1/17	5,9	33/284	11,6	0,704
Eliminación 11q22-q23	21/301	7,0	2/17	11,8	19/284	6,7	0,336
Trisomía 12	58/301	19,3	1/17	5,9	57/284	20,1	0,211
Eliminación 13q14	157/301	52,2	8/17	47,1	149/284	52,5	0,665
FISH normal	89/301	29,6	8/17	47,1	81/284	28,5	0,104

^a ts, tipo silvestre; IGHV, gen de la región variable de cadena pesada de inmunoglobulina; FISH, hibridación fluorescente in situ

5

[77] **Tabla 3.** Características clínicas y biológicas de la cohorte^a RS

	Número (n=33)	%
Características clínicas en el momento del diagnóstico de RS		
Edad >65 años	19	57,6
Varón	22	66,7
ECOG PS > 1	13	39,3
Estadio Ann Arbor III-IV	33	100
Estadio Rai III-IV	14	42,4
Síntomas B	13	39,3
Tamaño del tumor > 5 cm	24	72,7
Plaquetas < 100 x 10 ⁹ /l	7	21,2
LDH > 1,5 ULN	17	51,5

	Número (n=33)	%
Antes de terapias para CLL > 1	7	21,2
Características patológicas en el momento del diagnóstico de RS		
Fenotipo No-GC	32	96,9
Infección VBE	0	0
Características genéticas		
Alteración <i>TP53</i>	18	54,5
Aberraciones <i>c-MYC</i>	5	15,1
Mutaciones <i>NOTCH1</i>	13	39,4
Identidad <i>IGHV</i> ≥98 %	24	72,7

^a ULN, límite superior de lo normal; GC, centro germinal; *IGHV*, gen de región variable de cadena pesada de inmunoglobulina

Análisis de mutación de *SF3B1*

5 El análisis mutacional de *SF3B1* (exones 1-25, incluyendo los sitios de corte y empalme. RefSeq o N° de registro GenBank NM_012433.2) se llevó a cabo en amplímeros de PCR obtenidos a partir del ADN genómico por una combinación de secuenciación de Sanger (llevada a cabo en un Analizador Genético ABI PRISM 3100, Applied Biosystems) y secuenciación dirigida de próxima generación (llevada a cabo en un Secuenciador de Genoma Junior, 454 Life Sciences, Roche, Branford, CT; media de cobertura ~ 200X). Se compararon las secuencias de Sanger con la correspondiente línea germinal de la RefSeq utilizando el Mutation Surveyor Version 2.41 (SoftGenetics, State College, PA) tras la curación tanto automática como manual. Las lecturas de la secuenciación obtenidas por 10 secuenciación de próxima generación se confirmaron posteriormente por secuenciación de Sanger a partir de ambas cadenas sobre amplímeros independientes. Las mutaciones sinónimas, polimorfismos de la línea germinal conocidos de bases de datos (dbSNP132, Ensembl Database, UCSC Genome Browser), y los cambios presentes en el ADN normal coincidente se retiraron para el análisis. Se llevaron a cabo los estudios moleculares con ocultación 15 de los datos clínicos. La predicción de los efectos funcionales de las sustituciones de aminoácidos se llevaron a cabo utilizando el algoritmo PolyPhen-2 (software versión 2.1, genetics.bwh.harvard.edu/pph2) (Adzhubei IA, et al. Nat Methods. 2010; 7(4):248-249).

Análisis de cariotipo por FISH y de mutaciones *IGHV*, *TP53* y *NOTCH1*

20 Se llevó a cabo el análisis FISH como se ha expuesto utilizando las sondas LSI13 y LSI13S319, CEP12, LSIp53, y LSIATM (Abbott, Roma, Italia) (Rossi D, et al. Clin Cancer Res. 2009; 15(3):995-1004). Se investigó el estado mutacional de *IGHV* como se había expuesto previamente (Rossi D, et al. Clin Cancer Res. 2009; 15(13):4415-4422). Las secuencias se alinearon con el directorio de secuencias ImMunoGeneTics y se consideraban mutadas si 25 la identidad con los genes de la línea germinal correspondientes era < 98 % (Hamblin TJ, et al. Blood. 1999; 94(6):1848-1854; Damle RN, et al. Blood. 1999; 94(6):1840-1847). Se analizaron las mutaciones de *TP53* y *NOTCH1* como se había expuesto (Rossi D, et al. Clin Cancer Res. 2009; 15(3):995-1004; Fabbri G, et al. J Exp Med. 2011; 208(7):1389-1401).

30 Análisis del número de copias

Se obtuvieron los perfiles de ADN de todo el genoma a partir del ADN genómico de alto peso molecular de pacientes con CLL utilizando la Matriz 6.0 SNP de todo el genoma humano Affymetrix (Affymetrix, Santa Clara, CA, USA), siguiendo las instrucciones del fabricante. El canal bioinformático utilizado para la identificación del número de 35 alteraciones de copias se había descrito anteriormente (Pasqualucci L, et al. Nat Genet. 2011; doi: 10.1038/ng.892; Rinaldi A, et al. Br J Haematol. 2011; doi: 10.1111/j.1365-2141.2011.08789.x).

Análisis del perfil de expresión genética

40 El análisis del perfil de expresión genética en las muestras de subpoblaciones de células B normales purificadas y las de CLL se llevó a cabo utilizando matrices HG-U133_plus2 Affymetrix como parte de un estudio independiente (base de datos GEO GSE12195). Las sondas que se utilizaron en la Fig. 2D eran las siguientes: 228758_at, 203140_at, y 215990_s_at (para *BCL6*); 219841_at y 224499_s_at (para *AICDA*); 203684_s_at y 203685_at (para *BCL2*); 204562_at y 216986_s_at (para *IRF4*); y 201070_x_at, 201071_x_at, 211185_s_at, y 214305_s_at (para *SF3B1*). 45

Análisis estadístico

50 Se midió la supervivencia total a partir de la fecha del diagnóstico a la fecha de la muerte (suceso) o del último seguimiento (fin del estudio). Se midió la supervivencia libre de tratamiento a partir de la fecha del diagnóstico hasta la fecha de la enfermedad progresiva y sintomática que necesitaba tratamiento de acuerdo con las directrices del

IWCLL-NCI (suceso), muerte, o el último seguimiento (fin del estudio) (Hallek M, et al. Blood. 2008; 111(12):5446-5456). Se estimó la supervivencia por el método de Kaplan-Meier (Kaplan EL y Meier P. Am Stat Assoc. 1958; 53:457-481). Se estimó la asociación en bruto de las mutaciones *SF3B1* y la supervivencia por análisis del rango logarítmico (Kaplan EL y Meier P. Am Stat Assoc. 1958; 53:457-481).

Se valoraron las variables categóricas por el ensayo de chi cuadrado y ensayos exactos cuando era apropiado. Todos los ensayos estadísticos eran de dos lados. La significación estadística se definió como el valor de $p < 0,05$. Los análisis se llevaron a cabo con el software del Paquete Estadístico para Ciencias Sociales (SPSS) v.18.0 (Chicago, IL).

Ejemplo 2: Las mutaciones en el Factor de empalme SF3B1 afectan la progresión y refractariedad a la fludarabina en leucemia linfocítica crónica

Después de la observación inicial de las mutaciones *SF3B1* recurrentes en 3/11 CLL refractarias a fludarabina analizados por secuenciación del exoma completo, se llevó a cabo la re-secuenciación dirigida de la secuencia codificante *SF3B1* y los sitios de corte y empalme en 48 casos adicionales de CLL progresiva y refractaria a fludarabina (número total de casos analizados: 59), que se recolectaron en el momento de la progresión inmediatamente antes de comenzar el tratamiento al que el paciente eventualmente fallaba en la respuesta (Tabla 1). El *SF3B1* estaba alterado en 10/59 (un 17 %) de CLL refractaria a fludarabina por mutaciones sin sentido ($n = 9$) o por eliminaciones en fase ($n = 1$) agrupadas en las repeticiones HEAT3, HEAT4 y HEAT5 de la proteína *SF3B1* (Fig. 1 y Fig. 2A). Hay dos sitios altamente conservados entre especies (codón 662 y codón 700) que estaban mutados recurrentemente en 3 y 5 casos respectivamente (Fig. 1). Las mutaciones en *SF3B1* eran monoalélicas y se preveía que eran significativas funcionalmente de acuerdo con el algoritmo PolyPhen-2 (Tabla 4) (Adzhubei IA, et al. Nat Methods. 2010; 7(4):248-249). Estos datos documentan que las mutaciones de *SF3B1*, un factor de corte y empalme que es un componente crítico del empalmosoma, se asocian recurrentemente con la CLL refractaria a fludarabina.

Las características biológicas de la CLL refractaria a la fludarabina que alberga mutaciones *SF3B1* se resumen en la Tabla 1. Las mutaciones se producen sin relación con el estado de mutación en *IGHV*, la expresión de CD38 y la expresión de ZAP70. En el momento de la refractariedad a fludarabina, las mutaciones *SF3B1* se enriquecían en los casos que albergaban un cariotipo FISH normal ($p = 0,008$; Tabla 1). También, las mutaciones *SF3B1* se distribuían de una manera mutuamente excluyente en comparación con la alteración de *TP53* ensayada por eliminación y/o mutación (información mutua $I = 0,0609$; $p = 0,046$; Fig. 2B). Combinando las mutaciones *SF3B1* con otras lesiones genéticas enriquecidas en los casos quimiorrefractarios (alteración *TP53*, mutaciones *NOTCH1*, eliminación *ATM*) (Fabbri G, et al. J Exp Med. 2011; 208(7):1389-1401; Döhner H, et al. N Engl J Med. 2000; 343(26):1910-1916; Rossi D, et al. Clin Cancer Res. 2009; 15(3):995-1004; Zenz T, et al. Blood. 2009; 114(13):2589-2597; Stilgenbauer S, et al. J Clin Oncol. 2009; 27(24):3994-4001), la CLL refractaria a fludarabina parecía que se caracterizaba por múltiples alteraciones moleculares que, hasta cierto punto, son excluyentes mutuamente (Fig. 2B).

Para investigar si las mutaciones *SF3B1* se restringen a los casos quimiorrefractarios, se comparó la prevalencia de las mutaciones que se observaban a la vez que la refractariedad a fludarabina con la prevalencia de las mutaciones observadas en otras fases de la enfermedad. En una serie consecutiva evaluada en el momento del diagnóstico de CLL, las mutaciones *SF3B1* eran raras (17/301; un 5 %) (Fig. 2A; Tabla 4), y mostraban una asociación bruta con una corta supervivencia libre de tratamiento ($p < 0,001$) y supervivencia total ($p = 0,011$) (Fig. 2C). Extraordinariamente, 5/17 (29 %) de CLL mutadas en el diagnóstico eran pacientes refractarios primariamente a la fludarabina. Un paciente con alelos *SF3B1* de tipo silvestre en el momento del diagnóstico adquirió posteriormente una mutación en *SF3B1* concomitante con la refractariedad a la fludarabina (caso 7915 en la Tabla 4). En CLL investigadas en el momento del diagnóstico, la distribución de los puntos clave y el espectro molecular de las mutaciones de *SF3B1*, así como su relación mutua con otras lesiones genéticas, eran similares a las observadas en la CLL refractaria a fludarabina (Fig. 1 y 2B; Tabla 4). Las mutaciones *SF3B1* solo se encontraron en 2/33 (un 6,0 %) de RS relacionado clónicamente (Fig. 1 y 2A; Tabla 4). A lo largo de las diferentes fases de la enfermedad investigadas, se confirmó que las mutaciones se adquirían somáticamente en todos los casos ($n = 18$) para los que estaba disponible el ADN de la línea germinal (Tabla 4). Aunque la expresión relativa de *SF3B1* en CLL era mayor en comparación con los subgrupos de células B normales (Figura 2D), la investigación extensa por análisis de matriz SNP ejecutado sobre las anomalías del número de copias focales de *SF3B1* en esta leucemia ($n = 0/323$). Estos datos documentan que las mutaciones *SF3B1*: i) se producen con una tasa baja en el momento de la presentación de CLL, mientras que se enriquecen en los casos refractarios a fludarabina; ii) tienen un papel mínimo en la transformación en RS, corroborando la noción de que el cambio histológico de CLL es molecularmente distinto de la progresión quimiorrefractaria sin transformación en RS (Rossi D, et al. Blood. 2011; 117(12):3391-3401).

La identificación de mutaciones *SF3B1* señala la implicación de la regulación de corte y empalme como un nuevo mecanismo patógeno en CLL. El *SF3B1* es un componente crítico del empalmosoma tanto principal (tipo U2) como menor (tipo U12) (Luke MM, et al. Mol Cell Biol. 1996; 16(6):2744-2755; Wang C, et al. Genes Dev. 1998; 12(10):1409-1414; Das BK, et al. Mol Cell Biol. 1999; 19(10):6796-6802), los cuales representan la escisión precisa de los intrones a partir del pre-ARNm (Wahl MC, et al. Cell. 2009; 136(4):701-718; David CJ y Manley JL. Genes Dev. 2010; 24(21):2343-2364; Ward AJ y Cooper TA. J Pathol. 2010; 220(2): 152-163. El papel biológico preciso de las mutaciones *SF3B1* en la CLL actualmente es esquivo. La patogenicidad de las mutaciones *SF3B1* en CLL está

- fuertemente sustentada por el agrupamiento de estas mutaciones en puntos clave conservados durante la evolución localizados en los dominios HEAT, que se disponen en tándem en estructuras tipo floritura que sirven como un almacén flexible al que se pueden unir otros componentes (Andrade MA y Bork P. *Nat Genet.* 1995; 11(2): 115-116; Andrade MA, et al. *J Struct Biol.* 2001; 134(2-3):117-13 1). También, la observación de que SF3B1 regula el programa de corte y empalme alternativo de genes que controlan la progresión del ciclo celular y la apoptosis señala la contribución de las mutaciones *SF3B1* en la modulación de la proliferación de células tumorales y la supervivencia (David CJ y Manley JL. *Genes Dev.* 2010; 24(21):2343-2364; Kaida D, et al. *Nat Chem Biol.* 2007; 3(9):576-583; Corrionero A, et al. *Genes Dev.* 2011; 25(5):445-459).
- 5
- 10 Además de las implicaciones patógenas, las mutaciones *SF3B1* también proporcionan una diana terapéutica para los inhibidores de *SF3B1* (Kaida D, et al. *Nat Chem Biol.* 2007; 3(9):576-583; Corrionero A, et al. *Genes Dev.* 2011; 25(5):445-459), que están actualmente en desarrollo pre-clínico como fármacos anti-cáncer.

[89] Tabla 4. Mutaciones SF3B1 en CLL y RS

ID Muestra	Fase de la enfermedad	Cambio de nucleótidos ^c	Cambio de aminoácidos ^d	Dominio afectado	Sitio conservado ^e	PolyPhen-2 ^f	Valor COSMIC v54 ^g
7040 ^c	Diagnóstico de CLL	c.2044A>G	p.K666E	HEAT 4	No	Perjudicial	1,000
11772 ^a	Diagnóstico de CLL	c.2044A>G	p.K666E	HEAT 4	No	Perjudicial	1,000
9094	Diagnóstico de CLL	c.2046G>T	p.K666N	HEAT 4	No	Perjudicial	1,000
4602 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
4681 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
7561 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
10676 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
11196 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
11197 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
11489 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
11785 ^a	Diagnóstico de CLL	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
3950 ^a	Diagnóstico de CLL	c.2267G>A	p.G740E	-	Si	Perjudicial	0,949
4845 ^a	CLL refractaria a fludarabina ^b	c.1938A>T	p.R630S	HEAT 3	Si	Perjudicial	1,000
7425 ^a	CLL refractaria a fludarabina ^b	c.2034C>A	p.II662Q	HEAT 4	Si	Perjudicial	1,000
7228	CLL refractaria a fludarabina ^b	c.2034C>A	p.II662Q	HEAT 4	Si	Perjudicial	1,000
12627	CLL refractaria a fludarabina	c.2032C>G	p.II662D	HEAT 4	Si	Perjudicial	1,000
7915 ^a	CLL refractaria a fludarabina	c.2044A>G	p.K666E	HEAT 4	No	Perjudicial	1,000
12571	CLL refractaria a fludarabina	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
12631	CLL refractaria a fludarabina	c.2145A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
14220_R ^e	CLL refractaria a fludarabina	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
3981 ^a	CLL refractaria a fludarabina ^b	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000
5565 ^a	CLL refractaria a fludarabina ^b	c.2143_2148delCAGAAA	p.delQ699_K700	HEAT 5	Si	na	na
8343	Síndrome de Richter	c.2056C>G	p.Q670E	HEAT 4	Si	Perjudicial	0,999
7509 ^a	Síndrome de Richter	c.2146A>G	p.K700E	HEAT 5	Si	Perjudicial	1,000

^a Para estos pacientes, estaba disponible el ADN emparejado normal y se confirmó el origen somático de la mutación

^b En estos pacientes, el tiempo de refractariedad a la fludarabina era concomitante con el diagnóstico clínico

^c Numeración de acuerdo con el N° de registro GenBank NM_012433.2

^d Numeración de acuerdo con el N° de registro GenBank NP_036565.2

^e Posición conservada entre los ortólogos de SF3B1

^f na, no aplicable, ya que el algoritmo PolyPhen-2 solo predice el impacto de sustituciones de aminoácidos

^g Mutaciones enumeradas en la edición v.54 de la base de datos del Catálogo de Mutaciones Somáticas en Cáncer (COSMIC) (<http://www.sanger.ac.uk/genetics/CGP/cosmic/>)

La mención de los documentos en el presente documento no se considera que refleje la admisión de que sean técnicas anteriores relevantes. Además, su mención no es una indicación de una búsqueda de divulgaciones relevantes. Todo lo que se establece con respecto a los datos o contenido de los documentos se basa en información disponible y no es una admisión de su precisión o corrección.

5

REIVINDICACIONES

1. Un método *in vitro* para pronosticar un sujeto con leucemia linfocítica crónica (CLL), que comprende:

- 5 (a) determinar la secuencia de una parte del gen *SF3B1* en una muestra biológica obtenida del sujeto, donde la parte del gen *SF3B1* comprende una secuencia que codifica un dominio HEAT3, HEAT4 o HEAT5; y
 (b) analizar la secuencia del gen *SF3B1* en cuanto a una mutación,

10 donde la presencia de la mutación en la secuencia del gen *SF3B1* predice una disminución de la supervivencia del sujeto, pronosticando de esta manera al sujeto con leucemia linfocítica crónica (CLL).

2. Un método *in vitro* para determinar la respuesta de un sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina, que comprende:

- 15 (a) determinar la secuencia de una parte del gen *SF3B1* en una muestra biológica obtenida del sujeto, donde la parte del gen *SF3B1* comprende una secuencia de codifica el dominio HEAT3, HEAT4 o HEAT5; y
 (b) analizar la secuencia del gen *SF3B1* en cuanto a una mutación,

20 donde la presencia de la mutación en la secuencia del gen *SF3B1* indica que el sujeto es resistente o refractario a la fludarabina, determinando de esta manera la respuesta del sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina.

3. Un método *in vitro* para pronosticar un sujeto con leucemia linfocítica crónica (CLL), que comprende:

- 25 (a) determinar la secuencia de una parte del polipéptido SF3B1 en una muestra biológica obtenida del sujeto, donde la parte del polipéptido SF3B1 comprende la secuencia de un dominio HEAT3, HEAT4 o HEAT5; y
 (b) analizar la secuencia del polipéptido SF3B1 en cuanto a una mutación,

30 donde la presencia de la mutación en la secuencia del polipéptido SF3B1 predice una disminución de la supervivencia del sujeto, pronosticando de esta manera el sujeto con leucemia linfocítica crónica (CLL).

4. Un método *in vitro* para determinar la respuesta de un sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina, que comprende:

- 35 (a) determinar la secuencia de una parte del polipéptido SF3B1 en una muestra biológica obtenida del sujeto, donde la parte del polipéptido SF3B1 comprende una secuencia de un dominio HEAT3, HEAT4 o HEAT5; y
 (b) analizar la secuencia del polipéptido SF3B1 en cuanto a una mutación,

40 donde la presencia de la mutación en la secuencia del polipéptido SF3B1 indica que el sujeto es resistente o refractario a la fludarabina, determinando de esta manera la respuesta del sujeto con leucemia linfocítica crónica (CLL) al tratamiento con fludarabina.

45 5. El método de la reivindicación 1 o 3, donde la disminución de la supervivencia comprende la supervivencia libre de tratamiento o la supervivencia total.

50 6. El método de la reivindicación 4 donde el tratamiento con fludarabina se va a interrumpir o se va a sustituir por el tratamiento con clorambucilo, ciclofosfamida, rituximab, alemtuzumab, bendamustina, o una combinación de los mismos.

55 7. El método de la reivindicación 4, donde el sujeto se va a tratar con clorambucilo, ciclofosfamida, rituximab, alemtuzumab, bendamustina, o una combinación de los mismos.

8. El método de una cualquiera de las reivindicaciones 1-4, donde la etapa de análisis comprende además una reacción en cadena de la polimerasa (PCR), secuenciación de Sanger, secuenciación de próxima generación, o una combinación de las mismas.

60 9. El método de una cualquiera de las reivindicaciones 1-4, donde el sujeto se ha diagnosticado de leucemia linfocítica crónica (CLL) y/o donde la mutación es una mutación sin sentido o una eliminación en fase.

65 10. El método de la reivindicación 1 o 2, donde la muestra biológica comprende una molécula de ADN genómico, ADNc, o ARN aislada o purificada, o el método de la reivindicación 3 o 4 donde la muestra biológica comprende una molécula de polipéptido aislada o purificada.

11. El método de la reivindicación 3 o 4, donde la parte del polipéptido SF3B1 que comprende la secuencia de un dominio HEAT3, HEAT4 o HEAT5 se selecciona de entre el grupo que consiste en SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, y 19, y es preferentemente SEQ ID NO: 1 o 19.

5 12. El método de la reivindicación 1, donde la mutación es:

una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2044 o 2146 de SEQ ID NO: 17; o

10 una sustitución de una Timina (T) por una Guanina (G) en la posición de base de nucleótido 2046 de SEQ ID NO: 17; o

una sustitución de Adenina (A) por una Guanina (G) en la posición de base de nucleótido 2267 de SEQ ID NO: 17.

13. El método de la reivindicación 2, donde la mutación es:

15 una sustitución de una Timina (T) por una Adenina (A) en la posición de base de nucleótido 1938 de SEQ ID NO: 17; o

una sustitución de una Adenina (A) por una Citosina (C) en la posición de base de nucleótido 2034 de SEQ ID NO: 17; o

20 una sustitución de una Guanina (G) por una Citosina (C) en la posición de base de nucleótido 2032 de SEQ ID NO: 17; o

una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2044 de SEQ ID NO: 17; o

una sustitución de una Guanina (G) por una Adenina (A) en la posición de base de nucleótido 2146 de SEQ ID NO: 17; o

25 una eliminación de la secuencia de nucleótidos CAGAAA correspondiente con las posiciones de bases de nucleótido 2143 a 2148 de SEQ ID NO: 17; o

una sustitución de una Guanina (G) por una Citosina (C) en la posición de base de nucleótido 2056 de SEQ ID NO: 17.

30 14. El método de la reivindicación 3, donde la mutación da como resultado:

una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Lisina (Lys o K) en el codón 666 o 700 de SEQ ID NO: 19; o

35 una sustitución de un resto de Asparagina (Asn o N) por un resto de Lisina (Lys o K) en el codón 666 de SEQ ID NO: 19; o

una sustitución de Ácido Glutámico (Glu o E) por un resto de Glicina (Gly o G) en el codón 740 de SEQ ID NO: 19.

15. El método de la reivindicación 4, donde la mutación da como resultado:

40 una sustitución de un resto de Serina (Ser o S) por un resto de Arginina (Arg o R) en el codón 630 de SEQ ID NO: 19; o

una sustitución de un resto de Glutamina (Gln o Q) por un resto de Histidina (His o H) en el codón 662 de SEQ ID NO: 19; o

45 una sustitución de un resto de Ácido Aspártico (Asp o D) por un resto de Histidina (His o H) en el codón 662 de SEQ ID NO: 19; o

una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Lisina (Lys o K) en el codón 666 o 700 de SEQ ID NO: 19; o

50 una eliminación de un resto de Glutamina (Gln o Q) en el codón 699 y un resto de Lisina (Lys o K) en el codón 700 de SEQ ID NO: 19; o

una sustitución de un resto de Ácido Glutámico (Glu o E) por un resto de Glutamina (Gln o Q) en el codón 670 de SEQ ID NO: 19.

FIGURA 1

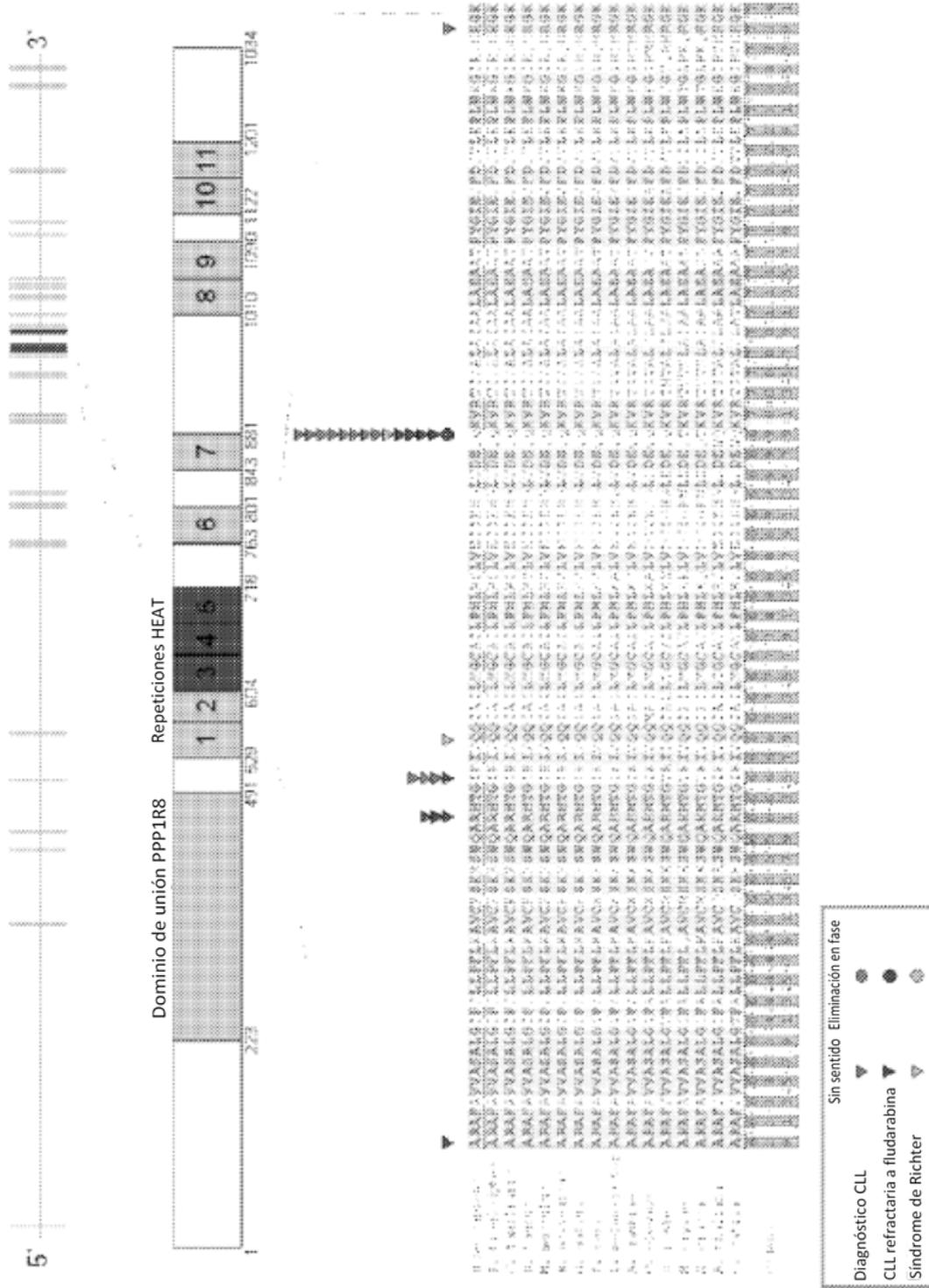


FIGURA 2

