

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 765 709**

51 Int. Cl.:

**C12Q 1/6886** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **29.09.2016 PCT/US2016/054501**

87 Fecha y número de publicación internacional: **06.04.2017 WO17059108**

96 Fecha de presentación y número de la solicitud europea: **29.09.2016 E 16782152 (9)**

97 Fecha y número de publicación de la concesión europea: **30.10.2019 EP 3356554**

54 Título: **Procedimientos para subtipificar el linfoma de células B difusas (DLBCL)**

30 Prioridad:

**29.09.2015 US 201562234491 P**

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

**10.06.2020**

73 Titular/es:

**HTG MOLECULAR DIAGNOSTICS, INC (100.0%)  
3430 E. Global Loop  
Tuscon, AZ 85706, US**

72 Inventor/es:

**LAFLEUR, BONNIE;  
LIU, QIAN;  
LUECKE, JOHN, W. y  
ROCHE, PATRICK, C.**

74 Agente/Representante:

**PONS ARIÑO, Ángel**

ES 2 765 709 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

Procedimientos para subtipificar el linfoma de células B difusas (DLBCL)

### 5 Campo

Esta descripción se refiere a clasificadores, matrices y kits para distinguir subtipos de linfoma difuso de células B grandes (DLBCL), así como a los usos de los resultados de clasificación para diagnóstico, pronóstico y/o selección de terapia.

10

### ANTECEDENTES

El linfoma es el cáncer de sangre más común. Las dos formas principales de linfoma son el linfoma de Hodgkin y el linfoma no Hodgkin. El linfoma difuso de células B grandes (DLBCL) es la forma más común de linfoma no Hodgkin y representa hasta el 30 por ciento de los casos recientemente diagnosticados. El DLBCL se puede caracterizar adicionalmente sobre la base de marcadores de expresión génica como subtipo de células B del centro germinal (GCB) o como un subtipo no tipo GCB (incluyendo el tipo de células B activadas (ABC)). Los subtipos GCB y no GCB (como ABC) pueden usarse para (i) predecir el pronóstico de un paciente con linfoma, con linfomas similares a GCB que tienen un mejor pronóstico que los subtipos no GCB (o ABC), y/o (ii) determinar tratamientos para pacientes con linfoma, por ejemplo, con terapias dirigidas a GCB o ABC, o sobre la base de la agresividad de la enfermedad (con el de ABC siendo generalmente más agresivo que la enfermedad de GCB).

No existe una prueba diagnóstica única y aceptada para distinguir los subtipos de DLBCL GCB y no GCB (por ejemplo, ABC) y/o para determinar el pronóstico de, o el tratamiento(s) para, pacientes con DLBCL. Algunas pruebas basadas en inmunohistoquímica (IHC) están disponibles en la práctica clínica, incluyendo las pruebas de Colomo y col. (Blood 101(1):78-84, 2003), Hans y col. (Blood 103:275-282, 2004), Muris y col. (J. Pathol. 208(5):714-23, 2006), Choi y col. (Mod. Pathol. 21:250A, 2008) y Tally (Meyer y col., J. Clin. Oncol. 29(2):200, 2011).

Debido a que el DLBCL tiene una amplia variedad de resultados clínicos, se cree que las bases moleculares de la enfermedad son igualmente complejas, y el descubrimiento de biomarcadores clínicamente útiles está en curso. Uno de los objetivos de dicho trabajo es "subdividir esta categoría de diagnóstico clínicamente heterogénea en enfermedades molecularmente distintas con comportamientos clínicos más homogéneos" (Alizadeh y col., Nature 403:503, 2000). En ese sentido, la IHC es una tecnología inherentemente limitada, al menos por la razón de que solo se puede co-detectar un número limitado de biomarcadores de proteínas. Del mismo modo, el genoma (*por ejemplo*, el análisis de ADN) es limitado al menos por la razón de que no está claro si se expresa que un evento genómico tiene un impacto funcional en el sistema biológico.

El análisis de transcriptoma de ARNm tiene ventajas sobre el análisis de proteínas y de ADN porque la expresión de ARNm proporciona una instantánea de la actividad genómica y es susceptible de multiplexación para la detección de muchos biomarcadores en la misma muestra. Se han informado varios ARNm expresados diferencialmente en muestras de DLBCL (*por ejemplo*, Alizadeh y col., Nature, 403:503, 2000; Rosenwald y col., NEJM 346 (25):1937, 2002; Wright y col., Proc. Natl. Acad. Sci. 100:9991, 2003; y Roberts y col., Lab. Invest. 78:979, 2007). Sin embargo, quedan lagunas significativas e importantes en esta área por resolver.

El documento de los EE.UU. 2011/152115 (STAUDT y col., 2011) describe un procedimiento para clasificar un linfoma difuso de células B grandes (DLBCL) que comprende medir la expresión de 27 genes que comprenden ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME y PIM1. El procedimiento calcula una puntuación de probabilidad utilizado para clasificar las muestras en ABC, GCB o subtipos sin clasificar.

### 50 RESUMEN

La identificación de biomarcadores cuya expresión es característica de DLBCL en una variedad de tecnologías de detección es un primer paso importante (*por ejemplo*, Roberts y col., Lab. Invest. 78:979, 2007). Sin embargo, desde una perspectiva práctica y clínica, un desafío crítico que queda es sintetizar grandes cantidades de datos de expresión arcana en resultados significativos para el beneficio de las prácticas médicas y los pacientes con DLBCL.

Esta descripción describe las innovaciones para clasificar los subtipos de DLBCL, así como el uso de los resultados de la clasificación para el diagnóstico, el pronóstico y la selección de la terapia. Por ejemplo, un clasificador (que puede implementarse por ordenador) determina un subtipo de DLBCL basado en la expresión génica, como los niveles de expresión para dos o más genes de firma en una muestra de un sujeto con DLBCL. En algunos ejemplos, la expresión se mide por un número proporcional de lecturas de secuencia para un gen dado del número total de lecturas (donde una lectura es el número de copias complementarias de una secuencia específica (*por ejemplo*, de una sonda) que

puede leer un secuenciador). El clasificador puede clasificar eficazmente los subtipos de DLBCL y proporcionar resultados significativos para el beneficio de las prácticas médicas y los pacientes con DLBCL.

- Según un aspecto de las innovaciones, los procedimientos miden los niveles de expresión (*por ejemplo*, ácido nucleico o proteína) para una pluralidad de genes de firma de DLBCL en una muestra. Por ejemplo, el procedimiento puede medir la expresión de al menos dos de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; o al menos dos de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS (como medir la expresión de los 16 genes).
- 10 El valor de expresión resultante para cada gen se puede ponderar y los valores de expresión ponderados se pueden sumar. Se puede determinar una puntuación de probabilidad utilizando los valores de expresión ponderados sumados. El procedimiento o clasificador puntúa o compara la puntuación de probabilidad con los umbrales correspondientes para el clasificador de DLBCL. Después, basado al menos en parte en los resultados de la puntuación, el procedimiento clasifica la muestra de DLBCL como tipo células B activadas (ABC), como tipo células B del centro germinal (GCB) o sin clasificar. La clasificación de subtipos de DLBCL se puede implementar como parte de un procedimiento, como parte de un sistema informático adaptado para realizar el procedimiento, o como parte de un medio tangible legible por ordenador que almacena instrucciones ejecutables por ordenador para hacer que un sistema informático realice el procedimiento.
- 15
- 20 Las innovaciones descritas en esta invención también incluyen el uso de los resultados de la clasificación para el diagnóstico, pronóstico y/o selección de terapia para un sujeto con DLBCL. También se describen en esta invención conjuntos y kits que se pueden usar para determinar la expresión de genes de firma de DLBCL.
- Según un aspecto de la invención, se proporciona un procedimiento para clasificar un linfoma difuso de células B grandes (DLBCL), que comprende:
- 25

medir la expresión, opcionalmente la expresión de ácido nucleico, de CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, en una muestra obtenida de un sujeto, obteniendo así un valor de expresión para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2,

30 PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;

- ponderar cada valor de expresión, generando así valores de expresión ponderados;
- sumar los valores de expresión ponderados, generando así un valor de expresión ponderado sumado;
- calcular una puntuación de probabilidad usando el valor de expresión ponderado sumado;
- 35 comparar la puntuación de probabilidad con los umbrales; y
- clasificar el DLBCL como tipo células B activadas (ABC) cuando la puntuación de probabilidad es un valor en un intervalo para ABC;
- clasificar el DLBCL como tipo células B del centro germinal (GCB) cuando la puntuación de probabilidad es un valor en un intervalo para GCB; o
- 40 clasificar el DLBCL como sin clasificar cuando la puntuación de probabilidad es un valor en un intervalo para la categoría sin clasificar.

En un procedimiento preferido de la invención, dicho ácido nucleico es ARNm y/o ARNm.

- 45 En un procedimiento preferido de la invención, dicha muestra es una muestra fija y la expresión se analiza opcionalmente usando un ensayo basado en nucleasa.

En un procedimiento preferido de la invención, dicha medición de la expresión comprende secuenciación, y donde el valor de la expresión es un recuento que representa un número de moléculas presentes en la muestra y,

50 opcionalmente, donde la expresión de medición comprende la cuantificación de la expresión.

En un procedimiento preferido de la invención, dicha ponderación de cada valor de expresión comprende multiplicar los valores de expresión por los coeficientes que se muestran en la Tabla 5.

- 55 En un procedimiento preferido de la invención, dicho cálculo de una puntuación de probabilidad usa la fórmula:

$$Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}}$$

(Ecuación 1)

donde  $\beta_0$  es una constante,  $\beta$  es un vector de ponderaciones para los genes clasificadores y  $x^T$  es un vector de valores

medidos de los genes clasificadores para una muestra dada.

En un procedimiento preferido de la invención, dichos umbrales son puntos límites t de puntuación de probabilidad.

- 5 En un procedimiento preferido de la invención, dichos puntos límites t de puntuación de probabilidad son <0,43 para ABC, > 0,57 para GCB y 0,43 - 0,57 para la categoría sin clasificar.

En un procedimiento preferido de la invención, dicho procedimiento clasifica menos del 6% de las muestras como sin clasificar.

10

En un procedimiento preferido de la invención, dichos valores de expresión se obtienen al:

poner en contacto la muestra con sondas de protección de nucleasa que comprenden una secuencia flanqueante (NPPF) en condiciones suficientes para que cada NPPF se una específicamente a su molécula de ácido nucleico diana; poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la secuencia flanqueante (CFS) en condiciones suficientes para que la secuencia flanqueante se una específicamente a la CFS;

15

poner en contacto la muestra con una nucleasa específica para moléculas de ácido nucleico de una sola cadena en condiciones suficientes para eliminar las moléculas de ácido nucleico no unidas, generando así una muestra digerida que comprende NPPF hibridadas para las moléculas de ácido nucleico diana y con la o las CFS; amplificar las NPPF en la muestra digerida con un cebador de amplificación, generando así amplicones de NPPF; secuenciar al menos una porción de los amplicones de NPPF; y

20

contar el número de amplicones de NPPF; determinando así los valores de expresión para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS en la muestra.

25

En un procedimiento preferido de la invención, dichos valores de expresión se obtienen al:

poner en contacto la muestra con NPPF específicas para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y moléculas de ácido nucleico diana TYMS,

30

donde cada NPPF comprende:

un extremo 5' y un extremo 3', una secuencia complementaria a una región de la molécula de ácido nucleico diana CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, que permite la unión específica entre NPPF y la molécula de ácido nucleico diana,

35

donde la secuencia flanqueante se ubica como 5', 3' o ambas, a la secuencia complementaria a la molécula de ácido nucleico diana, donde la secuencia flanqueante 5' se ubica como 5' en la secuencia complementaria a la molécula de ácido nucleico diana, y la secuencia flanqueante 3' se ubica como 3' en la secuencia complementaria a la molécula de ácido nucleico diana,

40

donde la secuencia flanqueante comprende al menos 12 nucleótidos contiguos que no se encuentran en una molécula de ácido nucleico presente en la muestra,

si la NPPF comprende una secuencia flanqueante 5', poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la secuencia flanqueante 5' (5CFS), un fosfato del extremo 5', en condiciones suficientes para que la secuencia flanqueante 5' se hibride específicamente con la 5CFS;

45

si la NPPF comprende una secuencia flanqueante 3', poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la secuencia flanqueante 3' (3CFS) en condiciones suficientes para que la secuencia flanqueante 3' se hibride específicamente con la 3CFS; donde al menos una de las 3CFS y 5CFS comprende un residuo de captura;

50

generar una NPPF hibridada con la molécula de ácido nucleico diana, hibridada con la 3CFS, hibridada con la 5CFS o hibridada con la 3CFS y la 5CFS;

poner en contacto la muestra con una nucleasa específica para moléculas de ácido nucleico de una sola cadena en condiciones suficientes para eliminar las moléculas de ácido nucleico no unidas, generando así una muestra digerida que comprende una NPPF hibridada con la molécula de ácido nucleico diana, hibridada con la 3CFS, hibridada con la 5CFS, o hibridada tanto con la 3CFS como con la 5CFS;

55

capturar la NPPF hibridada con la molécula de ácido nucleico diana, hibridada con la 3CFS, hibridada con la 5CFS o hibridada con la 3CFS y la 5CFS;

60

ligar el fosfato 5' de la 3CFS a un extremo 3' de la molécula de ácido nucleico diana, y ligar un extremo 3' de la 5CFS a un extremo 5' de la molécula de ácido nucleico diana, generando así una molécula de ácido nucleico diana ligada;

separar la NPPF de la molécula de ácido nucleico diana ligada, generando así una mezcla que comprende una

NPPF de cadena simple y una molécula de ácido nucleico diana ligada de cadena simple; y secuenciar al menos una porción de la molécula de ácido nucleico diana ligada de cadena simple, determinando así la secuencia de moléculas de ácido nucleico diana CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS en la muestra.

5

En un procedimiento preferido de la invención, dicha NPPF comprende al menos un dUTP, y el procedimiento comprende además poner en contacto la mezcla que comprende la NPPF de una sola cadena y una molécula de ácido nucleico diana ligada de una sola cadena con uracilo ADN glicosilasa (UDG) en condiciones suficientes para degradar la NPPF de una sola cadena, después de la desnaturalización y antes de la secuenciación.

10

Según un aspecto de la invención, se proporciona una cantidad terapéuticamente efectiva de (1) quimioterapia con ciclofosfamida, doxorubicina, vincristina y prednisona o prednisolona (CHOP), (2) una quimioterapia con rituximab más CHOP (R-CHOP) o (3) una quimioterapia con etopósido más R-CHOP (R-EPOCH) para usar en el tratamiento de un sujeto diagnosticado por el procedimiento según la invención cuando el DLBCL se clasifica como GCB.

15

De acuerdo con un aspecto de la invención, se proporciona una cantidad terapéuticamente eficaz de uno o más de bendamustina, pixantrona, gemcitabina/oxaliplatino, vincristina liposomal, mAb anti-CD20, mAb anti-CD22, mAb anti-CD74, mAb anti-CD40, una construcción de mAb anti-CD19 y CD3 biespecífica de cadena simple, I-131 tositumomab, Inotuzumab ozogamicina, 90Y-epratuzumab tetraxetán, talidomida, lenalidomida, Bortezomib, NPI-0052, Everolimus, Tamsulosin, Vorinostat, Oblimersen sodio, PF-3512676, 17-AAG, Bevacizumab, Aflibercept, CAL-101, ácido valproico, Dinaciclib, Fostamatinib, Dasatinib, Enzastaurin, PCI-32765, SB1518 y Sorafenib para usar en el tratamiento de un sujeto diagnosticado mediante el procedimiento según la invención cuando el DLBCL se clasifica como ABC.

20

De acuerdo con un aspecto de la invención, se proporcionan uno o más medios de almacenamiento legibles por ordenador que almacenan instrucciones ejecutables por ordenador para hacer que el sistema informático, cuando así se lo programe, realice el procedimiento según la invención.

25

Según un aspecto de la invención, se proporciona un sistema informático adaptado para realizar el procedimiento según la invención.

30

Según un aspecto de la invención, se proporciona un kit que comprende:

un contenedor que comprende al menos 16 NPPF diferentes y CFS correspondientes, donde las al menos 16 NPPF diferentes son específicas para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;

35

uno o más de:

un recipiente que comprende microesferas que pueden unirse específicamente a amplicones de al menos 16 NPPF diferentes;

40

un recipiente que comprende una nucleasa específica para ácidos nucleicos de una sola cadena;

un recipiente que comprende un tampón de lisis;

un recipiente que comprende un tampón de lavado;

un recipiente que comprende regentes para PCR;

45

un recipiente que comprende etanol;

un recipiente que comprende una ligasa;

un recipiente que comprende un tampón de ligadura;

un recipiente que comprende aceite de desnaturalización;

un recipiente que comprende proteinasa K; y opcionalmente además comprende que el contenedor comprenda al

50

menos 16 NPPF diferentes y CFS correspondientes,

donde las al menos 16 NPPF diferentes son específicas para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;

55

una nucleasa específica para ácidos nucleicos de una sola cadena;

tampón de lisis; y

un tampón de terminación de nucleasa, y

un sistema informático según la invención que implementa un clasificador para subtipos de DLBCL y opcionalmente

donde el sistema informático comprende un software o un medio legible por ordenador que recibe valores de

60

expresión para dos o más genes de firma de DLBCL, califica los valores múltiples contra los umbrales correspondientes para cada gen y clasifica una muestra en un marco que indica el subtipo de DLBCL de la muestra.

Las características anteriores y otras de la descripción se volverán más aparentes a partir de la descripción detallada, la cual se encuentra a continuación, con referencia a las figuras adjuntas.

#### BREVE DESCRIPCIÓN DE LOS DIBUJOS

- 5 **La FIG. 1** es un diagrama de flujo que ilustra una técnica generalizada para clasificar el subtipo de DLBCL.
- La FIG. 2** es un diagrama de flujo que muestra los casos utilizados para el entrenamiento y validación del modelo clasificador de DLBCL.
- 10 **La FIG. 3** es un gráfico que muestra la distribución de la expresión a través de las muestras para los 96 genes incluidos en el ensayo de DLBCL utilizado para entrenar y validar una realización de clasificador.
- La FIG. 4** es un gráfico que muestra un medio para determinar un límite inferior para muestras sin clasificar.
- 15 **La FIG. 5** es un gráfico que muestra la clasificación (*es decir*, ABC, GCB o sin clasificar) de un conjunto de validación de muestras de DLBCL, que previamente se había caracterizado por procedimientos clínicamente aceptados como ABC (puntos negros) o GCB (puntos rojos). Las muestras se identifican a lo largo del eje x. El eje y muestra la probabilidad pronosticada para cada caso según el perfil de expresión génica. En esta realización, el punto límite para la probabilidad estimada para una muestra de GCB fue de 0,57 (la probabilidad pronosticada > 0,57 es GCB), el punto límite para la clasificación ABC fue inferior a 0,43 (probabilidad pronosticada < 0,43) y las muestras sin clasificar cayeron entre, e incluyen, el ABC y los puntos límites de GCB ( $0,43 \leq$  probabilidad pronosticada  $\leq 0,57$ ).
- 20 **La FIG. 6** es un gráfico que muestra el Log2 CPM para CD274 (PDL-1) y PDCD1 (PD1) en un lisado preparado a partir de una muestra FFPE de GCB.
- La FIG. 7** es un gráfico que muestra Log2 CPM para CD274 (PDL-1) y PDCD1 (PD1) en un lisado preparado a partir de una muestra FFPE de ABC.
- 30 **La FIG. 8** es un gráfico que muestra un plan de clasificación para cada réplica y condición.

#### LISTADO DE SECUENCIAS

Las secuencias de ácido nucleico enumeradas aquí se muestran usando abreviaturas de letras estándar para las bases de nucleótidos, como se define en el inciso 37 del No. 1.822 del Código de Reglas Federales. Solamente se muestra una cadena de cada secuencia de nucleótidos, pero se entiende que la cadena complementaria se incluye mediante cualquier referencia a la cadena que se muestra. La lista de secuencias presentada aquí se generó el 1 de septiembre de 2016. En las secuencias proporcionadas: SEQ ID NO: 1-16 son ejemplos de secuencias de ácido nucleico de sonda.

#### 40 Descripción detallada

A menos que se indique lo contrario, los términos técnicos se utilizan según el uso convencional. Las definiciones de términos comunes en la biología molecular pueden encontrarse en Benjamin Lewin, *Genes V*, publicado por Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew y col. (ed.), *The Encyclopedia of Molecular Biology*, publicado por Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); y Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, publicado por VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Los términos singulares "un", "uno(a)" y "el(la)" incluyen referentes plurales a menos que el contexto indique claramente lo contrario. Del mismo modo, la palabra "o" pretende incluir "y" a menos que el contexto indique claramente lo contrario. "Comprender" significa "incluir". Por lo tanto, "que comprende A o B" significa que incluye A, o B, o A y B. Debe entenderse además que todos los tamaños de bases o tamaños de aminoácidos, y todos los valores de peso molecular o masa molecular, dados para ácidos nucleicos o polipéptidos son aproximados, y se proporcionan para su descripción. Aunque se pueden usar métodos y materiales similares o equivalentes a los descritos en esta invención en la práctica o ensayo de la presente descripción, a continuación se describen métodos y materiales adecuados. En caso de conflicto, prevalecerá la presente memoria descriptiva, incluyendo las explicaciones de los términos. Además, los materiales, procedimientos y ejemplos son sólo ilustrativos y no pretenden ser limitativos.

Salvo que se indique lo contrario, los procedimientos y técnicas de la presente descripción se realizan generalmente según procedimientos convencionales bien conocidos en la técnica y como se describe en varias referencias generales y más específicas que se citan y analizan a lo largo de la presente memoria descriptiva. Véase, *por ejemplo*, Sambrook y col., *Molecular Cloning: A Laboratory Manual*, 2da ed., Cold Spring Harbor Laboratory Press, 1989; Sambrook y col.,

Molecular Cloning: A Laboratory Manual, 3ra ed., Cold Spring Harbor Press, 2001; Ausubel y col., Current Protocols in Molecular Biology, Greene Publishing Associates, 1992 (y Suplementos para el 2000); Ausubel y col., Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 4ta ed., Wiley & Sons, 1999; Harlow y Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1990; y Harlow y Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 1999.

## Resumen

El DLBCL es el tipo más común de linfoma no Hodgkin en adultos. Hay dos subtipos principales de DLBCL, de tipo 10 células B del centro germinal (GCB) y no GCB (generalmente de tipo células B activadas (ABC)). El GCB generalmente se asocia con un pronóstico favorable, mientras que el ABC o no GCB generalmente se asocia con un peor pronóstico ( *por ejemplo*, es una enfermedad más agresiva). Algunos médicos consideran que el linfoma mediastínico primario de células B (PMBCL) es otro subtipo de DLBCL (Lossos, J. Clin. Oncol. 23:6351-6357, 2005). El PMBCL surge en el timo y generalmente se presenta como una masa en el mediastino. Los subtipos de DLBCL que no son GCB incluyen 15 todos los subtipos de DLBCL, excepto el subtipo GCB. De manera similar, los subtipos no ABC de DLBCL incluyen todos los subtipos de DLBCL, excepto el subtipo ABC.

En esta invención, se describen innovaciones para clasificar los subtipos de DLBCL, así como también el uso de los resultados de la clasificación para el diagnóstico, el pronóstico y/o la selección de la terapia. Los clasificadores 20 ejemplares determinan un subtipo de DLBCL basado en niveles de expresión para dos o más genes de firma en una muestra de un sujeto con DLBCL. Al determinar efectivamente el subtipo de DLBCL, un clasificador proporciona resultados significativos para el beneficio de las prácticas médicas y los pacientes con DLBCL.

Las innovaciones descritas en esta invención incluyen procedimientos de clasificación (que pueden implementarse 25 total o parcialmente por ordenador), medios legibles por ordenador que almacenan instrucciones ejecutables por ordenador para realizar dichos procedimientos, y sistemas informáticos adaptados para realizar dichos procedimientos. Las innovaciones descritas en esta invención incluyen además el uso de los resultados de dicha clasificación para el pronóstico, la selección de terapia y/u otros propósitos, así como las matrices y kits que proporcionan los valores que cuantifican los niveles de expresión de genes de firma como entrada para la clasificación. 30

Son posibles varias alternativas a los ejemplos descritos en esta invención. Por ejemplo, cualquiera de los procedimientos descritos en esta invención puede alterarse cambiando el orden de los actos del procedimiento descrito, dividiendo, repitiendo u omitiendo ciertos actos del procedimiento, etc. Los diversos aspectos de la tecnología descrita pueden usarse en combinación o por separado. Las diferentes realizaciones usan una o más de las 35 innovaciones descritas. Algunas de las innovaciones aquí descritas abordan uno o más de los problemas señalados en el fondo. Generalmente, una técnica/herramienta dada no resuelve todos esos problemas.

## Procedimientos para el subtipado del DLBCL

40 En esta invención se describen procedimientos para determinar un subtipo de DLBCL (*por ejemplo*, GCB o subtipos no GCB (como ABC) basados en la expresión génica de dos o más genes (como dos o más genes de firma de DLBCL) en una muestra de un sujeto con DLBCL. En general, a los datos de expresión para cada gen de firma evaluado se les asigna un valor (por ejemplo, basado o representado como una cantidad proporcional del total de lecturas para una muestra dada). Los valores asignados a los genes de firma evaluados se ponderan y se combinan para 45 proporcionar una puntuación (o puntuación de probabilidad). Después se determina un subtipo de DLBCL (*por ejemplo*, una indicación de GCB o no GCB) de la puntuación basada en los puntos límites, los umbrales u otros criterios.

En un ejemplo, la descripción proporciona procedimientos para clasificar un linfoma difuso de células B grandes (DLBCL). Dichos procedimientos pueden incluir la detección o medición directa o indirecta (como la expresión de ácido 50 nucleico o la expresión de proteínas) de una pluralidad de genes característicos de DLBCL, como al menos dos, tres, cuatro o la totalidad de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; o al menos dos, tres, cuatro o la totalidad de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, en una muestra (*por ejemplo*, como una muestra dFFPE fija) obtenida de un sujeto. Por lo tanto, se obtiene un valor de expresión (que puede ser cuantitativo) para cada gen 55 de firma de DLBCL, como un valor de expresión para al menos dos de (o todos) los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8; o para al menos dos de (o todos) los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. El valor de expresión resultante para cada gen de firma de DLBCL se puede ponderar (por ejemplo, multiplicando por un valor de coeficiente, como los que se muestran en la Tabla 5), generando así al menos dos valores de expresión ponderados. Los valores 60 de expresión ponderados resultantes se suman o adicionan, generando así un valor de expresión ponderada sumada. El valor de la expresión ponderada sumada se utiliza (*por ejemplo*, en un modelo de probabilidad) para calcular una puntuación de probabilidad, que se compara con umbrales o valores límite. En un ejemplo, el cálculo de una puntuación

de probabilidad usa la fórmula

$$Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}}$$

(Ecuación 1)

5 donde  $\beta_0$  es una constante,  $\beta$  es el peso para un gen de firma (como uno de la Tabla 1) y  $x^T$  es el valor para tal gen de firma. En un ejemplo, la muestra de DLBCL se clasifica como tipo células B activadas (ABC) cuando la puntuación de probabilidad está por debajo de la puntuación establecida para el ABC, como tipo células B del centro germinal (GCB) cuando la puntuación de probabilidad está por encima del límite establecido para el GCB; o como sin clasificar cuando la puntuación de probabilidad es un valor en un intervalo de puntuaciones para sin clasificar, que, por lo  
10 general, está entre, e incluye, la puntuación límite para el ABC y la puntuación límite para el GCB.

En algunos ejemplos, se mide la expresión de moléculas de ácido nucleico, como el ARNm. En algunos ejemplos, se mide la expresión de moléculas de proteínas. Las medidas de expresión en algunos ejemplos son cuantitativas o semicuantitativas. En algunos ejemplos, la medición de la expresión utiliza una secuenciación, y donde el valor de  
15 expresión para cada gen de firma de DLBCL es un recuento que representa un número (relativo o absoluto) de ácidos nucleicos diana (*por ejemplo*, ARNm)/proteínas presentes en la muestra.

En algún ejemplo, los procedimientos descritos y los clasificadores de DLBCL clasifican menos del 6% de las muestras como sin clasificar, por ejemplo, menos del 5% o del 4 al 6% como sin clasificar. Por el contrario, otros clasificadores  
20 de DLBCL (como los que miden la expresión de estos 12 genes: BCL6, CCND2, ENTPD1, FOXP1, FUT8, IRF4, ITPKB, LMO2, LRMP, MME, MYBL1 y SERPINA9) tiene una mayor tasa de muestras sin clasificar, como al menos el 10, 11 o 12%, por ejemplo, entre el 10 y el 12%. Por lo tanto, el clasificador de DLBCL descrito permite que aproximadamente del 4 al 8% de las muestras que no pudieron clasificarse mediante los procedimientos anteriores, ahora se clasifiquen como ABC o GCB. En algunos ejemplos, los procedimientos descritos subtipan el DLBCL (*por*  
25 *ejemplo*, como GCB o no GCB, por ejemplo, ABC) con una precisión de al menos el 90%, como al menos del 92%.

En algunos ejemplos, los procedimientos miden la expresión usando un ensayo basado en nucleasa. Por ejemplo, los valores de expresión se pueden obtener poniendo (*por ejemplo*, con la colocación en asociación física directa) la muestra en contacto con (1) al menos dos sondas de protección de nucleasa que comprenden una secuencia  
30 flanqueante (NPPF) en condiciones suficientes para que cada NPPF se una específicamente a su molécula de ácido nucleico diana (por ejemplo, el gen de firma de DLBCL), (2) una molécula de ácido nucleico que incluye una secuencia complementaria a la secuencia flanqueante (CFS) en condiciones suficientes para que la secuencia flanqueante se una específicamente a la CFS, y (3) una nucleasa específica para moléculas de ácido nucleico de una sola cadena en condiciones suficientes para eliminar las moléculas de ácido nucleico no unidas. Como se usa en esta invención, las  
35 condiciones suficientes se refieren a cualquier entorno (*por ejemplo*, acuoso, incluyendo u omitiendo  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{Mg}^{++}$ ,  $\text{Zn}^{++}$ , u otros iones específicos, agentes caotrópicos o tensioactivos, o que tengan una concentración particular de tales iones, agentes caotrópicos o tensioactivos, o que tengan un iónico particular fuerza, pH, tipo de tampón o concentración, o temperatura) que permite la actividad deseada, por ejemplo, que permite la unión o hibridación específica entre dos moléculas de ácido nucleico (como una sonda y un ácido nucleico diana) o que permite que una  
40 nucleasa elimine (o digiera) ácidos nucleicos no unidos. Esto da como resultado una muestra digerida que incluye NPPF hibridadas con las moléculas de ácido nucleico diana y las CFS.

Opcionalmente, las moléculas no unidas o de una sola cadena se eliminan de la muestra. Las NPPF hibridadas con la diana y con las CFS se amplifican con uno o más cebadores de amplificación adecuados, generando así amplicones  
45 de NPPF. Al menos una parte de los amplicones de NPPF está secuenciada (*por ejemplo*, para determinar con qué molécula de ácido nucleico diana (*por ejemplo*, un gen de firma de DLBCL) se hibridó). El número de amplicones de NPPF se puede contar, o bien se le puede asignar un valor, determinando así los valores de expresión para cada uno de al menos dos (o todos) los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8; o al menos dos de (o todos) los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC,  
50 REL, STAT3, TNFRSF8 y TYMS en la muestra. Por lo tanto, la NPPF utilizada en el ensayo permite la detección y, en algunos ejemplos, la cuantificación de una molécula de ácido nucleico diana, como un ARNm.

Los procedimientos descritos pueden incluir además administrar una cantidad terapéuticamente efectiva de una terapia adecuada, dependiendo de la clasificación del DLBCL. Por ejemplo, cuando el DLBCL se clasifica como GCB,  
55 es posible administrar al sujeto una cantidad terapéuticamente efectiva de (1) una quimioterapia con ciclofosfamida, doxorubicina, vincristina y prednisona o prednisolona (CHOP), (2) una quimioterapia con rituximab más CHOP (R-CHOP) o (3) un etopósido más quimioterapia R-CHOP (R-EPOCH). En otro ejemplo, una cantidad terapéuticamente efectiva de bendamustina, pixantrona, gemcitabina/oxaliplatino, vincristina liposomal, mAb anti-CD20, mAb anti-CD22,



mAb anti-CD74, mAb anti-CD40, una construcción de mAb anti-CD 19 y CD3 biespecífica de una sola cadena, I-131 tositumomab (radioinmunoterapia con anti-CD20), Inotuzumab ozogamicina (CMC-544) (inmunoconjugado citotóxico dirigido a CD22), 90Y-epratuzumab tetraxetán (mAb anti-CD22 humanizado radiomarcado, Brentuximab vedotina (SGN-35in) (conjugado de mAb anti-CD30 de monometil auristatina E antitubulina), talidomida, lenalidomida, 5 Bortezomib (inhibidor de proteasoma), NPI-0052 (inhibidor de proteasoma), Everolimus (inhibidor de mTOR), Temsirolimus (inhibidor de mTOR), Vorinostat (inhibidor de desacetilasa), Oblimaten sodio (oligonucleótido antisentido Bcl-2), PF-3512676 (antagonista de TLR9), 17-AAG (inhibidor HSP90), Bevacizumab (mAb anti-VEGF), Aflibercept (proteína de fusión VEGF), CAL-101 (inhibidor PI3K), ácido valproico (HDACI), Dinaciclib (inhibidor de CDK1, 2, 5, 9), Fostamatinib (Inhibidor Syk), Dasatinib (inhibidor RTK de BCR-ABL, SRC, c-Kit, PDGF y quinasas receptoras de efrinas), 10 Enzastaurin (inhibidor de la proteína quinasa beta), PCI-32765 (inhibidor de la tirosina quinasa de Bruton), SB1518 (inhibidor JAK2), Sorafenib (inhibidor de TKI de RAF/MEK/ERK/c-kit/Flt3, VEGFR, PDGFR, RETR) o cualquier otra terapia conocida en la técnica (véase, *por ejemplo*, Foon y col., Adv. Hematology, Identificación de artículo No. 302570, 2012, Identificación de material digital: 10.1155/2012/302570) se puede administrar al sujeto cuando el DLBCL se clasifica como ABC o no GCB.

15 También se proporcionan uno o más medios de almacenamiento legibles por ordenador que almacenan instrucciones ejecutables por ordenador para hacer que el sistema informático, cuando se programe de ese modo, realice los procedimientos de clasificación de DLBCL proporcionados en esta invención. También se proporcionan sistemas informáticos adaptados para realizar los procedimientos de clasificación de DLBCL proporcionados en esta invención.

20 Se proporcionan kits útiles con los procedimientos de clasificación de DLBCL descritos. En algunos ejemplos, tales kits tienen un contenedor que incluye al menos dos NPPF diferentes y CFS correspondientes, donde las al menos dos NPPF diferentes son específicas para dos o más de (como todos) los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3 o TNFRSF8; o específicas para dos o más de (como todos) los siguientes: CD47, CD86, 25 ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. En algunos ejemplos, tales kits tienen un contenedor que incluye al menos dos NPP, donde las al menos dos NPP diferentes son específicas para dos o más de (como todos) los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3 o TNFRSF8; o específicas para dos o más de (como todos) los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. Los kits pueden incluir además uno o más 30 de: un recipiente que comprende microesferas que pueden unirse específicamente a amplicones de al menos dos NPPF diferentes; un recipiente que comprende una nucleasa específica para ácidos nucleicos de una sola cadena; un recipiente que comprende un tampón de lisis; un recipiente que comprende un tampón que puede neutralizar la nucleasa, tal como un tampón que puede aumentar el pH por encima de 6; un recipiente que comprende un tampón de lavado; un recipiente que comprende regentes para PCR; un recipiente que comprende etanol; un recipiente que 35 comprende un aceite de desnaturalización; un contenedor que comprende un tampón de ligadura (*por ejemplo*, uno que incluye ligasa); y un recipiente que comprende proteinasa K. En algunos ejemplos, el kit también incluye cebadores específicos para una porción de la NPP o NPPF, como uno que puede adicionar una etiqueta de experimento y/o un adaptador de secuenciación en el extremo 3' o 5' o en ambos extremos de la NPP o NPPF, o CSF. En algunos ejemplos, el kit también incluye un sistema informático que implementa un clasificador para subtipos de DLBCL, como 40 un software o un medio legible por ordenador que recibe valores de expresión para dos o más genes de firma de DLBCL, califica los valores múltiples contra los umbrales correspondientes para cada gen y clasifica una muestra en un marco que indica el subtipo de DLBCL de la muestra.

La FIG. 1 muestra un procedimiento o técnica generalizada (100) para clasificar muestras de DLBCL. Al menos algunos 45 de los actos del procedimiento (100) pueden realizarse mediante un sistema informático, como una herramienta de diagnóstico de propósito especial, un ordenador de escritorio, un ordenador portátil, una tableta u ordenador de pantalla, un teléfono inteligente u otro dispositivo informático móvil. Dicho sistema informático puede incluir un lector que reciba valores que cuantifican los niveles de expresión génica de genes de firma de DLBCL para una muestra (*por ejemplo*, una conexión de red, una interfaz a un dispositivo de almacenamiento, una interfaz de usuario para 50 recibir directamente la entrada del usuario u otra interfaz que reciba valores que cuantifiquen los niveles de expresión génica), un clasificador que califique los valores de expresión génica contra los umbrales para la clasificación y, al menos en parte en los resultados de la puntuación, clasifique la muestra), una memoria de datos (*por ejemplo*, un almacenamiento extraíble o no extraíble, incluyendo discos magnéticos, cintas magnéticas o casetes, un CD-ROM, un DVD o cualquier otro medio que pueda usarse para almacenar información (como niveles medidos de expresión 55 génica de genes de firma de DLBCL, un software para ejecutar el clasificador, valores límites) de forma no transitoria y a la que se pueda acceder dentro del sistema informático) y un módulo de salida (*por ejemplo*, un dispositivo de salida de audio, una pantalla, una pantalla táctil, una impresora, una grabadora de CD u otro dispositivo que proporcione una salida del sistema informático, como una determinación de si la muestra analizada es ABC, GCB o sin clasificar). La arquitectura también puede incluir un módulo de capacitación que calcule umbrales o límites 60 almacenados en la memoria de datos. El clasificador puede incluir un procesador de datos (*por ejemplo*, para multiplicar cada valor de expresión génica a un valor de ponderación particular y proporcionar los valores ponderados a un totalizador de puntuación), un totalizador de puntuación (para sumar los valores ponderados calculados y

proporcionarlos al módulo de probabilidad) y un módulo de probabilidad (*por ejemplo*, para determinar la probabilidad de una clasificación particular utilizando los valores ponderados de suma y para proporcionar la probabilidad o el subtipo de DLBCL a un módulo de salida). Por ejemplo, el módulo de probabilidad puede comparar el o los valores de probabilidad calculados contra los umbrales correspondientes para ABC, GCB y sin clasificar. El clasificador obtiene los umbrales de probabilidad/valores límites de la memoria de datos, que almacena los umbrales para el clasificador de DLBCL. Como resultado de la clasificación, el clasificador puede proporcionar una clasificación resumida como no GCB (*por ejemplo*, ABC), GCB o sin clasificar.

Volviendo a la FIG. 1, para comenzar, el sistema recibe (110) (*por ejemplo*, de un lector o secuenciador, y por consiguiente, los datos pueden ser recibidos desde un estado almacenado) valores múltiples (*por ejemplo*, unidades de luz relativas o recuentos de una diana detectada, lecturas relativas o proporcionales) que cuantifican los niveles de expresión génica (*por ejemplo*, proteína y/o ARNm) de múltiples genes de firma de DLBCL, respectivamente, para una muestra. En algunos ejemplos, se reciben los valores de expresión génica de dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8, and TYMS; así como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8, and TYMS. Los valores de los niveles de expresión génica para los genes de firma de DLBCL se pueden medir para la muestra de prueba usando cualquiera de los enfoques descritos en esta invención o como sepan los expertos en la materia. En algunos ejemplos, los datos sin procesar opcionalmente pueden estar "condicionados", por ejemplo, un valor de fondo puede sustraerse de los datos y/o los datos pueden transformarse logarítmicamente (*por ejemplo*, log<sub>2</sub> transformado). El sistema puede normalizar opcionalmente los valores múltiples antes de calificar usando cualquiera de los enfoques descritos aquí, siempre que la clasificación use el mismo enfoque (o un enfoque similar) a la normalización que se usó durante el entrenamiento para determinar los umbrales.

Para cada gen, su valor de expresión se pondera (120), por ejemplo, multiplicando por un valor de ponderación, que refleja la contribución de la expresión de ese gen a la clasificación de una muestra como ABC, GCB o sin clasificar. Así, por ejemplo, si la ponderación de todos los genes clasificadores fuera 1,0, cada uno de estos genes tendría la misma ponderación en el resultado del clasificador. En la Tabla 5, se proporcionan valores de ponderación ejemplares y su gen correspondiente. El sistema suma los valores ponderados de expresión génica (130). Es decir, se suma el valor ponderado de cada gen analizado, lo que resulta en una suma ponderada. La suma ponderada se usa entonces (*por ejemplo*, en un modelo estadístico) para obtener una puntuación de probabilidad (140), por ejemplo, utilizando la Ecuación 1 en el Ejemplo 1. Un sistema informático que implementa el clasificador se puede utilizar para realizar una o más de estas tareas.

La puntuación de probabilidad se compara con los umbrales correspondientes para la probabilidad pronosticada (150) de que una muestra se clasifique como ABC o GCB o ninguno (*es decir*, sin clasificar). Los umbrales y las reglas utilizados para la puntuación dependen de la implementación. En esta invención, se describen los umbrales de ejemplo, y las reglas de puntuación de los ejemplos se describen a continuación. Alternativamente, el sistema usa otros umbrales y/u otras reglas de puntuación. En un ejemplo, los puntos límites son 0,43 para el punto inferior y 0,57 para el punto superior.

En base a los resultados de la puntuación, al menos en parte, el sistema clasifica (160) la muestra. Las reglas utilizadas para la clasificación dependen de la implementación. A continuación, se describen ejemplos de reglas de clasificación. Alternativamente, el sistema usa otras reglas de clasificación. En un ejemplo, el sistema clasifica una muestra como ABC, si la probabilidad pronosticada es < 0,43; el sistema clasifica una muestra como GCB, si la probabilidad pronosticada es > 0,57; y el sistema clasifica una muestra como sin clasificar, si la probabilidad predicha está dentro, e incluye, de 0,43 a 0,57.

### I. Preparación para obtener datos de expresión génica

La expresión génica es el proceso mediante el cual se transforma la información codificada en el genoma (gen) (*por ejemplo*, mediante procesos de transcripción y traducción) en los productos genéticos correspondientes (*por ejemplo*, ARN y proteínas), que funcionan de forma interrelacionada para dar lugar a un conjunto de características (también conocido como fenotipo). Para los fines de esta descripción, la expresión génica puede medirse mediante cualquier técnica conocida ahora o en el futuro. Comúnmente, la expresión génica se mide detectando los productos de los genes (*por ejemplo*, ARN y/o proteína) expresados en muestras recolectadas de sujetos de interés (como uno con DLBCL). Por lo tanto, se puede medir la expresión de dos o más genes diana, como dos o más genes o marcadores distintivos de DLBCL. En un ejemplo, la expresión génica se mide detectando (*por ejemplo*, secuenciando) una o más sondas que se unen específicamente a un ARNm diana y determinando el número (relativo o absoluto) de sondas que se unen a la diana ARNm. Por lo tanto, el número de sondas detectadas sirve como medida de la expresión génica. Los ejemplos de marcador(es) de DLBCL incluyen dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; así como dos o más de (como todos) los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. Otros ejemplos

específicos de marcador(es) de DLBCL incluyen dos o más (como 2, 3, 4, 5, 6, 7, 8, 9 o 10) de los siguientes: CD47, ENTPD1, FOXP1, FUT8, IL16, NF2, PIM1, STAT3, TNFRSF8 y TYMS (que están asociados con la clasificación ABC); o dos o más (como 2, 3, 4, 5 o 6) de los siguientes: CD86, ITPKB, LRMP, MME, PTPRC y REL (que están asociados con la clasificación GCB). Los sujetos que pueden analizarse con los procedimientos descritos incluyen mamíferos  
 5 humanos y no humanos (*por ejemplo*, sujetos veterinarios como gatos, perros y ratones). En un ejemplo, se sabe o se sospecha que un sujeto tiene un tumor, como un linfoma, como el DLBCL.

#### A. Sujetos y muestras

10 Las muestras adecuadas para su uso en los procedimientos descritos en esta invención incluyen cualquier muestra biológica del sujeto que contenga células (*por ejemplo*, células B) del tumor (*por ejemplo*, un tumor en los ganglios linfáticos, el bazo, el hígado, la médula ósea, el cerebro y/o la médula espinal) para el cual se desea obtener información sobre la expresión de genes o proteínas (como lo que refleja la Tabla 1). Las muestras incluyen las obtenidas de un sujeto, como las muestras clínicas obtenidas de un sujeto (incluidas las muestras de un sujeto humano  
 15 sano o aparentemente sano, o un paciente humano afectado por una afección o enfermedad a diagnosticar o investigar, como el DLBCL, más particularmente, el ABC o el GCB). En un ejemplo, se sabe o se sospecha que el sujeto tiene un DLBCL. Se pueden obtener muestras ejemplares de células o tejidos normales (*por ejemplo*, como muestra de control) y/o de células o tejidos tumorales. En ejemplos particulares, una muestra biológica incluye una muestra tumoral, como una muestra que contiene células de un linfoma. En un ejemplo, una muestra incluye ARN,  
 20 como ARNm. En algunas realizaciones, una muestra es de un sujeto que fue diagnosticado previamente con un DLBCL por histología o un procedimiento clínico (*por ejemplo*, hibridación IHC o *in situ* (ISH)) diferentes a los descritos en esta invención.

Las muestras de uso para determinar la expresión de dos o más genes de firma de DLBCL descrita incluyen cualquier  
 25 muestra biológica que incluya ácido nucleico (como ADN genómico, ADNc, ADN o ARN viral, ARNr, ARNt, ARNm, miARN, oligonucleótidos, fragmentos de ácido nucleico, ácidos nucleicos modificados, ácidos nucleicos sintéticos o similares) o proteínas. En algunos ejemplos, se selecciona un sujeto que tiene un linfoma, como el linfoma no Hodgkin (*por ejemplo*, DLBCL), por ejemplo, para determinar un diagnóstico o pronóstico para el sujeto o para seleccionar una o más terapias.

30 Las muestras ejemplares incluyen, entre otras, células, lisados celulares, frotis de sangre, preparaciones de citocentrifugadora, frotis de citología, fluidos corporales (*por ejemplo*, sangre, suero, saliva, esputo, orina, etc.), biopsias de tejidos (*por ejemplo*, biopsias tumorales), tejido o biopsias de ganglios linfáticos, biopsias líquidas, aspirados con aguja fina, muestras quirúrgicas, médula ósea, muestras de amniocentesis, material de autopsia y/o  
 35 secciones de tejido (*por ejemplo*, secciones de tejido de criostato y/o secciones de tejido incrustadas en parafina). Se apreciará que es posible utilizar cualquier procedimiento para obtener una muestra (como un tejido) de un sujeto, y que la selección del procedimiento utilizado dependerá de varios factores, como el tipo de tejido, la edad del sujeto o los procedimientos disponibles para el médico. Las técnicas estándar para la adquisición de las muestras descritas están disponibles. Véase, por ejemplo, Tubbs y Stoler, Cell and Tissue Based Molecular Pathology, Filadelfia: Churchill  
 40 Livingstone (2009). Por ejemplo, una muestra de un tumor que contiene material celular puede obtenerse mediante escisión quirúrgica de todo o parte del tumor, recogiendo un aspirado con aguja fina del tumor, así como otros procedimientos conocidos en la técnica.

Los procedimientos descritos son sensibles y específicos y permiten la detección de moléculas de ácido nucleico diana  
 45 en una muestra que contiene incluso un número limitado de células. Las muestras que incluyen pequeñas cantidades de células, como menos de 250.000 células (por ejemplo, menos de 100.000, menos de 50.000, menos de 10.000, menos de 1.000, menos de 500, menos de 200, menos de 100 células o menos de 10 células, incluyen, entre otras, muestras FFPE, aspirados con aguja fina (como los de linfa o bazo), biopsias por punción, biopsias por punción, pequeñas poblaciones de (*por ejemplo*, FACS) células clasificadas, pequeñas cantidades de células capturadas con  
 50 láser o macrodisseccionadas, exosomas y otras partículas subcelulares o fluidos corporales (como el líquido cefalorraquídeo). Por ejemplo, un ARN diana puede detectarse en tan solo 1.000 células (como una muestra que incluye 1.000 o más células, como 1.000, 2.000, 3.000, 4.000, 5.000, 6.000, 7.000, 8.000, 9.000, 10.000, 15.000, 20.000, 50.000 o más células). En algunos ejemplos, la expresión de un ARN diana puede detectarse en  
 55 aproximadamente 1.000 a 100.000 células, por ejemplo, aproximadamente 1.000 a 50.000, 1.000 a 15.000, 1.000 a 10.000, 1.000 a 5.000, 3.000 a 50.000, 6.000 a 30.000 o 10.000 a 50.000 células). En algunos ejemplos, la expresión de un ARN diana puede detectarse en aproximadamente 100 a 250.000 células, por ejemplo, aproximadamente 100 a 100.000, 100 a 50.000, 100 a 10.000, 100 a 5.000, 100 a 200 o 100 a 150 células. En otros ejemplos, la expresión de un ARN diana se puede detectar en aproximadamente 1 a 1.000 células (como aproximadamente 1 a 500 células, aproximadamente 1 a 250 células, aproximadamente 1 a 100 células, aproximadamente 1 a 50 células,  
 60 aproximadamente 1 a 25 células o aproximadamente 1 célula).

En ejemplos particulares, las muestras se usan directamente (*por ejemplo*, frescas o congeladas) o pueden

conservarse antes de su uso, por ejemplo, mediante una fijación (*por ejemplo*, mediante la fijación de formalina (como formalina tamponada neutra, formalina de zinc y formalina ácida) o una fijación de etanol) y/o mediante la incrustación en un medio sólido. Los medios de inclusión, por lo general, son inertes, capaces de repeler la humedad y penetrar en el tejido (*por ejemplo*, cera, parafina, celoidina, compuesto OCT™, agar, plásticos y acrílicos). Algunas muestras  
 5 útiles son muestras de tejido fijadas en formalina e incluidas en parafina (FFPE). En ejemplos específicos, una muestra de tejido a analizar se fija o, más particularmente, se fija y se incrusta en cera (parafina). El término desparafinado o descerado se refiere a la eliminación parcial o completa de cualquier tipo de medio de inclusión de una muestra biológica (*por ejemplo*, antes del análisis utilizando los procedimientos descritos). Por ejemplo, las secciones de tejido embebidas en parafina se pueden descerar al pasar a través de solventes orgánicos, tales como tolueno, xileno,  
 10 limoneno u otros solventes adecuados. En otros ejemplos, las secciones de tejido embebido en parafina se utilizan directamente (*por ejemplo*, sin una etapa de descerado).

Los tejidos pueden repararse mediante cualquier proceso adecuado, incluyendo la perfusión o la inmersión en un fijador. Los fijadores pueden clasificarse como agentes de reticulación (como aldehídos, *por ejemplo*, formaldehído, paraformaldehído y glutaraldehído, así como también agentes de reticulación no aldehídos), agentes oxidantes (*por  
 15 ejemplo*, iones metálicos y complejos, como el tetróxido de osmio y el ácido crómico), agentes desnaturizantes de proteínas (*por ejemplo*, ácido acético, metanol y etanol), fijadores de mecanismo desconocido (*por ejemplo*, cloruro mercúrico, acetona y ácido pícrico), reactivos combinados (*por ejemplo*, el Fijador de Carnoy, el metacarn, el líquido de Bouin, el fijador B5, el líquido de Rossman y el líquido de Gendre), microondas y fijadores diversos (*por ejemplo*,  
 20 excluyendo la fijación de volumen y la fijación de vapor). También se pueden incluir aditivos en el fijador, como tampones, detergentes, ácido tánico, fenol, sales metálicas (como cloruro de zinc, sulfato de zinc y sales de litio) y lantano.

Un fijador usado comúnmente en la preparación de muestras de tejido o células es el formaldehído, generalmente en  
 25 forma de una solución de formalina (4% de formaldehído en una solución tampón, denominada formalina tamponada al 10%). En un ejemplo, el fijador es formalina tamponada neutral al 10%.

Algunas muestras, incluidas células o tejidos fijos (*por ejemplo*, FFPE), se pueden preparar suspendiendo la muestra en una solución adecuada (*por ejemplo*, como se describe en los ejemplos a continuación) sin ninguna purificación  
 30 y/o transcripción inversa del ARNm presente en dicha muestra. Todo o parte (*por ejemplo*, un área o volumen previamente medido) de una muestra fijada a una superficie sólida (*por ejemplo*, un portaobjetos de microscopio) se puede raspar directamente en una solución adecuada con una mezcla suave (*por ejemplo*, usando una pipeta) para dispersar y/o romper grandes piezas de la muestra y crear una suspensión que comprende materiales celulares que permanecen en gran medida suspendidos al menos durante un período durante el cual la suspensión se pone en  
 35 contacto con reactivos adecuados. En algunas realizaciones, el ARNm en la muestra puede (pero no necesita) estar parcialmente degradado o es de poca preocupación si se ha producido una degradación incompleta del ARNm en la muestra.

En algunas realizaciones, una muestra es un lisado de células y/o tejido obtenido de un tumor o tejido. El lisado celular  
 40 contiene muchas de las proteínas y ácidos nucleicos contenidos en una célula e incluye, por ejemplo, los biomarcadores que se muestran en la Tabla 1. Los procedimientos para obtener o preparar un lisado celular son bien conocidos en la técnica y se pueden encontrar, por ejemplo, en Ausubel y col. (En Current Protocols in Molecular Biology, John Wiley & Sons, Nueva York, 1998). En algunos ejemplos, las células en la muestra se lisan o se permeabilizan en una solución acuosa (*por ejemplo*, usando un tampón de lisis). La solución acuosa o el tampón de  
 45 lisis pueden incluir detergente (como dodecil sulfato de sodio) y uno o más agentes caotrópicos (como formamida, guanidinio HCl, isotiocianato de guanidinio o urea). La solución también puede contener un tampón (*por ejemplo*, SSC). En algunos ejemplos, el tampón de lisis incluye aproximadamente del 8 al 60% de formamida (v/v) aproximadamente del 0,01 al 0,5% de SDS y aproximadamente 0,5-6X SSC (*por ejemplo*, aproximadamente 3X SSC). El tampón puede incluir opcionalmente ARNt de aproximadamente 0,001 a aproximadamente 2,0 mg/ml o una  
 50 ribonucleasa; ADNasa; proteinasa K; enzimas, *por ejemplo*, colagenasa o lipasa) que degradan proteínas, matrices, carbohidratos, lípidos o una especie de oligonucleótidos, o combinaciones de los mismos. El tampón de lisis también puede incluir un indicador de pH, como el rojo de fenol. Las células se incuban en la solución acuosa durante un período de tiempo suficiente (como aproximadamente 1 a alrededor de 60 minutos, *por ejemplo*, aproximadamente 5 a alrededor de 20 minutos o aproximadamente 10 minutos) y a una temperatura suficiente (como aproximadamente  
 55 de 22 a alrededor de 115°C, *por ejemplo*, aproximadamente de 37 a alrededor de 105°C, aproximadamente de 50 a alrededor de 95°C o aproximadamente de 65 a alrededor de 100°C) para lisar o permeabilizar la célula. En algunos ejemplos, la lisis se realiza a aproximadamente 50, 65 o 95°C, *por ejemplo*, si el ácido nucleico a detectar es ARN. En otros ejemplos, la lisis se realiza a aproximadamente 105°C, *por ejemplo*, si el ácido nucleico a detectar es el ADN. En algunos ejemplos, las condiciones de lisis pueden ser tales que el ADN genómico no sea accesible para las sondas  
 60 mientras que el ARN (*por ejemplo*, ARNm). En algunos ejemplos, la lisis celular cruda se usa directamente sin purificación adicional.

En algunos ejemplos, se aplica una muestra de tejido o célula a un sustrato y se analiza para determinar la presencia de uno o más ácidos nucleicos diana. Un soporte sólido útil en un procedimiento descrito solo necesita llevar la muestra biológica y, opcional, aunque ventajosamente, permitir la detección conveniente de componentes (*por ejemplo*, proteínas y/o secuencias de ácido nucleico) en la muestra. Los soportes ejemplares incluyen un portaobjetos de microscopio (*por ejemplo*, un portaobjetos de microscopio de vidrio o de plástico), un cubreobjetos (*por ejemplo*, un cubreobjetos de vidrio o de plástico), platos de cultivo de tejidos, placas multipocillo, membranas (*por ejemplo*, de nitrocelulosa o fluoruro de polivinilideno (PVDF)) o chips BIACORE™.

Una diana (como el ARN) puede aislarse o extraerse de la muestra, purificándolo así de otros componentes biológicos no diana en una muestra. Sin embargo, dicha etapa es opcional. La purificación se refiere a separar la diana de uno o más componentes extraños que también se encuentran en una muestra. Por ejemplo, antes de la detección basada en PCR de ARNm con cebadores específicos diana emparejados, el ARNm total o soluble (incluido el ARNm diana) a menudo se separa de las proteínas celulares y otros ácidos nucleicos en la muestra. Los componentes que se aíslan, extraen o purifican a partir de una muestra o muestra mixta generalmente se enriquecen en al menos el 50, 60, 75, 90 o 98%, o incluso al menos el 99%, en comparación con la muestra no purificada o no extraída.

En algunos ejemplos, las moléculas de ácido nucleico en la muestra se amplifican, por ejemplo, usando PCR (*por ejemplo*, RT-PCR en tiempo real), antes de su análisis. En un ejemplo, se usa un procedimiento de PCR cuantitativa.

## 20 **B. Muestra de control**

Las muestras de control pueden usarse en paralelo con los procedimientos descritos y pueden incluir cualquier muestra de control adecuada con la cual comparar la expresión de un biomarcador que se muestra en la Tabla 1. En algunas realizaciones, la muestra de control es tejido no tumoral, tal como una pluralidad de muestras de tejido no tumoral. En un ejemplo, el tejido no tumoral es un tejido que se sabe que es benigno, como el tejido linfoide histológicamente normal. En algunos ejemplos, el tejido no tumoral incluye una muestra linfoide o de médula ósea que parece normal; es decir, que muestra la ausencia de indicadores de displasia celular u otra enfermedad conocida (*por ejemplo*, un linfoma, como el DLBCL). En algunos ejemplos, el tejido no tumoral se obtiene del mismo sujeto, como el tejido no tumoral adyacente o incluso distante de una neoplasia maligna linfoide (como un DLBCL). En otros ejemplos, el tejido no tumoral se obtiene de un sujeto de control sano o de varios sujetos de control sanos. Por ejemplo, el tejido no tumoral se puede obtener de una pluralidad de sujetos de control sanos (*por ejemplo*, aquellos que no tienen ningún tipo de cáncer, incluido el cáncer de linfoma (*por ejemplo*, el DLBCL), como muestras que contienen células o tejidos normales (*por ejemplo*, de los ganglios linfáticos, el bazo, el hígado, la médula ósea, el cerebro y/o la médula espinal) de una pluralidad de dichos sujetos. En algunas realizaciones, una o más (*por ejemplo*, una pluralidad de) muestras de control se utilizan para obtener una referencia (*por ejemplo*, un control normal) o intervalos de valores para los niveles de expresión de los biomarcadores que se muestran en la Tabla 1. En algunas realizaciones, un valor de referencia obtenido de muestras de control puede ser una tendencia central de la población (como una media, mediana o promedio), o un intervalo de valores de referencia como  $\pm 0,5$ , 1,0, 1,5 o 2,0 de desviación estándar en torno a una tendencia central de la población.

40

## **C. Opciones analíticas ejemplares**

Algunas realizaciones de procedimientos usan muestras fijas (*por ejemplo*, muestras de tejido FFPE). Las técnicas de fijación pueden variar según el sitio, el país, el investigador, etc. (*Dissecting the Molecular Anatomy of Tissue*, ed. por Emmert-Buck, Gillespie y Chuaqui, Nueva York: Springer-Verlag, 244 páginas (2010)) y pueden afectar la integridad y/o accesibilidad al o los productos génicos a detectar. En algunos de estos procedimientos (*por ejemplo*, que involucran la PCR), la recuperación de ARN (*por ejemplo*, utilizando agentes de reticulación reversibles, fijadores a base de etanol y/o la extracción o purificación de ARN (total o parcialmente) puede ser ventajosa; mientras que, en otros procedimientos representativos (*por ejemplo*, con qNPA o qNPS) la recuperación de ARN es opcional o la recuperación de ARN claramente no es necesaria. De manera similar, el acondicionamiento de tejidos puede usarse para recuperar productos de genes de proteínas de tejidos fijos y, por lo tanto, puede ayudar en la detección de dichos productos de proteínas.

El porcentaje de tumor (*por ejemplo*, DLBCL) en muestras biológicas puede variar; así, en algunas realizaciones descritas, al menos el 5, 10, 25, 50, 75, 80 o 90% del área de muestra (o volumen de muestra) o las células totales en la muestra son tumorales (*por ejemplo*, del DLBCL). En otros ejemplos, las muestras pueden enriquecerse para las células tumorales, *por ejemplo*, mediante la macrodissección de áreas o células de una muestra que son o parecen ser predominantemente tumorales (*por ejemplo*, del DLBCL). Opcionalmente, un patólogo u otro profesional debidamente capacitado puede revisar la muestra (*por ejemplo*, la sección de tejido teñido con H y E) para determinar si hay un tumor suficiente presente en la muestra a analizar y/o marcar el área (*por ejemplo*, el área tumoral más densa) a macrodisseccionar. En ejemplos específicos, la macrodissección de tumor (*por ejemplo*, del DLBCL) evita en la mayor medida posible las áreas necróticas y/o hemorrágicas. Las muestras útiles en algunos procedimientos descritos

60

tendrán menos del 25, 15, 10, 5, 2 o 1% de necrosis por volumen de muestra o área o células totales.

La carga de la muestra puede influir en la cantidad y/o concentración del producto génico (*por ejemplo*, los biomarcadores en la Tabla 1) disponibles para la detección. En realizaciones particulares, al menos 1, 10, 100 ng, 1, 5 10, 100, 500 ug, 1 mg de ARN total, al menos 1, 10, 100 ng, 1, 10, 100, 500 ug, 1 mg de ADN total, o al menos 0,01, 0,1, 1, 10, 100 ng, 1, 10, 100, 500 ug o 1 mg de proteína total se aísla de y/o se presenta en una muestra (como un lisado de muestra). Algunas realizaciones usan muestras de tejido (*por ejemplo*, de los nódulos linfáticos FFPE, del bazo, del hígado, de la médula ósea, del cerebro y/o de la médula espinal) que tienen al menos 3, 5, 8 o 10  $\mu\text{m}$  (*por ejemplo*, aproximadamente de 3 a alrededor de 10  $\mu\text{m}$ ) de espesor y/o al menos 0,15, 0,2, 0,5, 1, 1,5, 2, 5 o 10  $\text{cm}^2$  10 en el área. La concentración de muestra suspendida en el tampón en algunas formas de realización del procedimiento es de al menos 0,006  $\text{cm}^2/\text{ul}$  (*por ejemplo*, 0,15  $\text{cm}^2$  del tejido FFPE por 25 uL de tampón (*por ejemplo*, el tampón de lisis)).

## II. Genes y conjuntos de genes

15 Entre las innovaciones aquí descritas se encuentran genes (también denominados biomarcadores) y conjuntos de genes (también denominados firmas genéticas) útiles para distinguir subtipos de DLBCL. En realizaciones particulares, se describen genes y conjuntos de genes para subtipar el ABC y el GCB de DLBCL (véase la Tabla 1). La expresión de dichos genes y conjuntos de genes es útil para los clasificadores de DLBCL y los algoritmos, y/o para diseñar 20 reactivos específicos de analitos (*por ejemplo*, sondas de ácido nucleico o anticuerpos) para matrices u otras composiciones descritas.

En algunas realizaciones, determinar el nivel de expresión en una muestra biológica (como una biopsia tumoral) incluye detectar dos o más productos génicos (*por ejemplo*, ARN o proteína) mostrados en la Tabla 1, por ejemplo, 25 determinando las cantidades relativas o reales de tales ácidos nucleicos (o proteínas) en la muestra, como se describe en esta invención.

### A. Genes de firma de DLBCL

30 Los genes de firma de DLBCL ejemplares incluyen los que se muestran en la Tabla 1. En algunas realizaciones, la firma del gen DLBCL incluye dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS (como al menos 2, 3, 4, 5, 6, 7, 8, 9 o estos 10); como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS (como al menos 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o estos 16).

35

Tabla 1. Marcadores ejemplares de DLBCL que permiten la clasificación de los subtipos GCB y no GCB

Gen	No. de acceso de GenBank/GI
CD47	NM_001777.3, NM_198793.2, LN680437.1
CD86	NM_175862.4, NG_029928.1, NM_176892.1, NM_001206924.1, NM_001206925.1
ENTPD1 (CD39)	NM_001776.5; NM_001098175.1; NM_001164178.1; NM_001164179.1; NM_001164181.1; NM_001164182.1; NM_001164183.1
FOXP1	NM_001012505.1; NM_032682.5; NM_001244808.1; NM_001244812.1; NM_001244814.1; NM_001244815.1, NM_001244813.1; NM_001244810.1; NM_001244816.1
FUT8	NM_178155.2, NM_004480.4, NM_178156.2, NR_038167.1, NR_038170.1
IL16	NM_004513.5, NM_172217.3, NM_001172128.1, XM_011521520.1, XR_931805.1
ITPKB	NM_002221.3; XM_005273120.1
LRMP	NM_006152.3, NM_001204127.1, NM_001204126.1, XM_006719076.2, XM_011520668.1, XM_011520669.1 XM_005253374.2, XM_011520670.1
MME (CD10)	NM_000902.3, NM_007287.2, NM_007288.2, BC143465.1, NM_007289.2
NF2	NM_000268.3, NM_016418.5, NM_181828.2, NM_181831.2
PIM1	NM_001243186.1, M24779.1, NM_002648.3

Gen	No. de acceso de GenBank/GI
PTPRC	NM_002838.4, NM_080921.3, NR_052021.1, NM_001267798.1
REL	NM_002908.3, NM_001291746.1, DQ314888.1
STAT3	NM_139276.2, NM_003150.3, NM_213662.1, NG_007370.1
TNFRSF8	NM_001243.4, XM_011542442.1, XM_011542444.1, XM_011542443.1, AY498860.1
TYMS	NM_001071.2; NG_028255.1

El grupo de diferenciación 47 (CD47) (*por ejemplo*, OMIM 601028), también conocido como proteína asociada a la integrina, es una proteína transmembrana que se asocia con las integrinas de membrana y también se une a los ligandos trombospondina-1 y la proteína alfa reguladora de la señal. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para CD47 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_001777.3, NM\_198793.2 y LN680437.1 describen secuencias de ácido nucleico CD47 humanas ejemplares, y los números de acceso de GenBank® NP\_001768.1, NP\_942088.1 y CEJ95640.1 describen secuencias de proteínas CD47 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de CD47 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de CD47 proporcionado en esta invención.

El grupo de diferenciación 86 (CD86) (*por ejemplo*, OMIM 601020), también conocido como B7-2, se expresa en las células presentadoras de antígeno y proporciona señales coestimuladoras para la activación y supervivencia de las células T. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para CD86 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_175862.4, NG\_029928.1, NM\_176892.1, NM\_001206924.1 y NM\_001206925.1 describen secuencias de ácido nucleico CD86 humanas ejemplares, y los números de acceso de GenBank® NP\_787058.4, NP\_008820.3, NP\_795711.1, NP\_001193853.1 y NP\_001193854.1 describen secuencias de proteínas CD86 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de CD86 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de CD86 proporcionado en esta invención.

El ectonucleósido trifosfato difosfohidrolasa 1 (ENTPD1) (*por ejemplo*, OMIM 601752), también conocido como CD39, ATPDase y NTPDase-1, es una enzima (EC 3.6.1.5) que hidroliza ATP y ADP a AMP, iniciando la generación de adenosina. El ENTPD1 se expresa ampliamente, incluso en los vasos sanguíneos y las células del sistema inmunitario. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para ENTPD1 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_001776.5; NM\_001098175.1; NM\_001164178.1; NM\_001164179.1; NM\_001164181.1; y NM\_001164182.1; NM\_001164183.1 describen secuencias de ácido nucleico ENTPD 1 humanas ejemplares, y los números de acceso de GenBank® AAH47664.1, NP\_001091645.1, NP001157650.1, NP\_001157653.1, NP\_001767.3, NP\_001157651.1, NP\_001157654.1 y NP\_001157655.1 describen secuencias de proteínas ENTPD1 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de ENTPD 1 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de ENTPD 1 proporcionado en esta invención.

La proteína P1 de la caja de Forkhead (FOXP1) (*por ejemplo*, OMIM 605515), es un factor de transcripción que contiene dominios de unión de ADN y unión de proteína-proteína, que funciona como un represor de la transcripción. El FOXP1 regula una variedad de aspectos importantes del desarrollo, incluyendo el desarrollo de tejidos: pulmones, cerebro, timo y corazón. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para Foxp1 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_001012505.1; NM\_032682.5; NM\_001244808.1; NM\_001244812.1; NM\_001244814.1; NM\_001244815.1, NM\_001244813.1; NM\_001244810.1 y NM\_001244816.1 describen secuencias de ácido nucleico Foxp1 humanas ejemplares, y los números de acceso de GenBank® AAG47632.1, NP\_001231743.1; NP\_001012523.1; NP\_001231737.1, NP\_001231739.1, NP\_001231741.1, NP\_001231742.1 y NP\_001231744.1 describen secuencias de proteínas Foxp1 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de Foxp1 para clasificar un DLBCL, como

aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de Foxp1 proporcionado en esta invención.

5 La fucosiltransferasa 8 (alfa (1,6) fucosiltransferasa) (FUT8) (*por ejemplo*, OMIM 602589), también conocida como GDP-L-Fuc: N-acetil-beta-D-glucosaminida alfa 1,6-fucosiltransferasa; GDP-fucosa-glucoproteína fucosiltransferasa; abd alfa- (1,6) -fucosiltransferasa, es una enzima (EC 2.4.1.68) que cataliza la transferencia de fucosa de GDP-fucosa a glucopéptidos complejos de tipo unido al nitrógeno. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente) y ARN no de codificación. Las secuencias de ácido nucleico y proteína para FUT8 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank®  
 10 NM\_178155.2, NM\_004480.4, NM\_178156.2, NR\_038167.1 y NR\_038170.1 describen secuencias de ácido nucleico FUT8 humanas ejemplares, y los números de acceso de GenBank® AAI42959.1, NP\_835368.1, NP\_004471.4 y NP\_835369.1 describen secuencias de proteínas FUT8 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de FUT8 para clasificar un DLBCL, como aquellas cuya expresión está  
 15 alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de FUT8 proporcionado en esta invención.

La interleucina 16 (IL16) (*por ejemplo*, OMIM 603035) es una citocina que funciona como quimioatrayente para algunas  
 20 células inmunes que expresan CD4. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para IL16 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_004513.5, NM\_172217.3, NM\_001172128.1, XM\_011521520.1 y XR\_931805.1 describen secuencias de ácido nucleico IL16 humanas ejemplares, y los números de acceso de GenBank® AAI36661.1, AAD15990.1, NP\_001165599.1, ACI00236.1 y  
 25 XP\_011519821.1 describen secuencias de proteínas IL16 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de IL16 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de IL16 proporcionado en esta invención.

30 El inositol 1,4,5-trisfosfato 3-quinasa B (ITPKB) (*por ejemplo*, OMIM 147522), también conocido como IP3 3-quinasa B, IP3K B y proteína 37 inductora de proliferación, fosforila el inositol 1,4,5-trifosfato a Ins(1,3,4,5)P4. Las secuencias de ácido nucleico y proteína para ITPKB están públicamente disponibles. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Por ejemplo, los números de acceso de  
 35 GenBank® NM\_002221.3 y XM\_005273120.1 describen secuencias de ácido nucleico ITPKB humanas ejemplares, y los números de acceso de GenBank® NP\_002212.3 y AAH15009.1 describen secuencias de proteínas ITPKB humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de ITPKB para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier  
 40 número de acceso de GenBank® de ITPKB proporcionado en esta invención.

La proteína de membrana con restricción linfóide (LRMP) (*por ejemplo*, OMIM 602003), también conocida como JAW1, se expresa en líneas y tejidos celulares linfoides. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para LRMP  
 45 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_006152.3, NM\_001204127.1, NM\_001204126.1, XM\_006719076.2, XM\_011520668.1, XM\_011520669.1, XM\_005253374.2 y XM\_011520670.1 describen secuencias de ácido nucleico LRMP humanas ejemplares, y los números de acceso de GenBank® NP\_006143.2, NP\_001191056.1, XP\_011518971.1, XP\_011518972.1, Q12912.3, and NP\_001191055.1 describen secuencias de proteínas LRMP humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes  
 50 de tales secuencias de LRMP para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de LRMP proporcionado en esta invención.

La metaloendopeptidasa de membrana (MME) (*por ejemplo*, OMIM 120520), también conocida como CD10, atriopeptidasa, antígeno de leucemia linfocítica aguda común, encefalinasa, neprilisina y endopeptidasa neutra, es una endopeptidasa neutra (EC 3.4.24.11) que corta los péptidos en el lado amino de los residuos hidrofóbicos e inactiva varias hormonas peptídicas, incluyendo el glucagón, las encefalinas, la sustancia P, la neurotensina, la oxitocina y la bradiquinina. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para MME están públicamente  
 60 disponibles. Por ejemplo, los números de acceso de GenBank® NM\_000902.3, NM\_007287.2, NM\_007288.2, BC143465.1 y NM\_007289.2 describen secuencias de ácido nucleico MME humanas ejemplares, y los números de acceso de GenBank® AAI43466.1, AAI01659.1, NP\_000893.2, NP\_009218.2, NP\_009219.2, y NP\_009220.2



describen secuencias de proteínas MME humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de MEE para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de MME proporcionado en esta invención.

5

La neurofibromina 2 (NF2) (*por ejemplo*, OMIM 607379), también conocida como merlina, es una proteína del citoesqueleto que tiene propiedades supresoras de tumores. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para NF2 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_000268.3, NM\_016418.5, NM\_181828.2 y NM\_181831.2 describen secuencias de ácido nucleico NF2 humanas ejemplares, y los números de acceso de GenBank® NP\_000259.1, NP\_057502.2, NP\_861966.1 y NP\_861969.1 describen secuencias de proteínas NF2 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de NF2 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de NF2 proporcionado en esta invención.

La proto-oncogen serina/treonina-proteína quinasa Pim-1 (PIM1) (*por ejemplo*, OMIM 164960) es un proto-oncogen involucrado en la señalización de citoquinas. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para PIM1 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_001243186.1, M24779.1 y NM\_002648.3 describen secuencias de ácido nucleico PIM1 humanas ejemplares, y los números de acceso de GenBank® NP\_001230115.1, NP\_002639.1 y AAH20224.1 describen secuencias de proteínas PIM1 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de PIM1 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de PIM1 proporcionado en esta invención.

La proteína tirosina fosfatasa, tipo receptor, C (PTPRC) (*por ejemplo*, OMIM 151460), también conocida como CD45, es una enzima (EC 3.1.3.48) expresada en células hematopoyéticas que regula la señalización del receptor de antígeno de células T y B. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Este gen contiene 34 exones y tres exones de las transcripciones primarias se empalman alternativamente para generar hasta ocho ARNm maduros diferentes y, después de la traducción, ocho productos proteicos diferentes. Estos tres exones generan las isoformas RA, RB y RC. Las secuencias de ácido nucleico y proteína para PTPRC están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_002838.4, NM\_080921.3, NR\_052021.1 y NM\_001267798.1 describen secuencias de ácido nucleico PTPRC humanas ejemplares, y los números de acceso de GenBank® NP\_002829.3, NP\_563578.2, P08575.2 y NP\_001254727.1 describen secuencias de proteínas PTPRC humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de PTPRC para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de PTPRC proporcionado en esta invención.

El homólogo de oncogén viral de reticuloendoteliosis viral aviar v-rel (REL) (*por ejemplo*, OMIM 164910), es un factor de transcripción que contiene un dominio de homología Rel y desempeña un papel en la supervivencia y proliferación de las células B. El gen REL se amplifica o muta en DLBCL. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para REL están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_002908.3, NM\_001291746.1 y DQ314888.1 describen secuencias de ácido nucleico REL humanas ejemplares, y los números de acceso de GenBank® NP\_002899.1, NP\_001278675.1 y AAI43886.1 describen secuencias de proteínas REL humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de REL para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de REL proporcionado en esta invención.

El transductor de señal y activador de transcripción (STAT3) (*por ejemplo*, OMIM 102582), es un activador de la transcripción, que puede tener una función oncogénica o supresora de tumores, dependiendo de los antecedentes genéticos del tumor. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para STAT3 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_139276.2, NM\_003150.3, NM\_213662.1 y NG\_007370.1 describen secuencias de ácido nucleico STAT3 humanas ejemplares, y los números de acceso de GenBank® NP\_644805.1, NP\_003141.2 y NP\_998827.1 describen secuencias de proteínas STAT3 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de STAT3 para

clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de STAT3 proporcionado en esta invención.

- 5 La superfamilia de receptores del factor de necrosis tumoral, miembro 8 (TNFRSF8) (*por ejemplo*, OMIM 153243), también conocida como CD30, se expresa mediante las células T y B activadas. Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para TNFRSF8 están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_001243.4, XM\_011542442.1, XM\_011542444.1, XM\_011542443.1 y AY498860.1 describen secuencias de ácido nucleico TNFRSF8 humanas ejemplares, y los números de acceso de GenBank® NP001234.3, XP\_011540746.1, y NP\_001268359.2 describen secuencias de proteínas TNFRSF8 humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de TNFRSF8 para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de TNFRSF8  
15 proporcionado en esta invención.

- La timidilato sintetasa (TYMS) (*por ejemplo*, OMIM 188350) es una enzima (EC 2.1.1.45) que cataliza la conversión de monofosfato de desoxuridina (dUMP) en monofosfato de desoxitimidina (dTMP). Para este gen, se han identificado variantes de transcripción (*por ejemplo*, transcripciones empalmadas alternativamente). Las secuencias de ácido nucleico y proteína para TYMS están públicamente disponibles. Por ejemplo, los números de acceso de GenBank® NM\_001071.2 y NG\_028255.1 describen secuencias de ácido nucleico TYMS humanas ejemplares, y los números de acceso de GenBank® NP001062.1, EAX01716.1 y EAX01717.1 describen secuencias de proteínas TYMS humanas ejemplares. Un experto en la materia apreciará que se pueden usar variantes de tales secuencias de TYMS para clasificar un DLBCL, como aquellas cuya expresión está alterada en DLBCL (por ejemplo, en ABC y/o GCB), como una secuencia que tiene al menos el 90, 95, 96, 97, 98 o el 99% de identidad de secuencia para cualquier número de acceso de GenBank® de TYMS proporcionado en esta invención.  
25

Describimos el siguiente subtipado útil de DLBCL, que incluye, sin limitación, determinar o medir la expresión de:

- 30 a. uno o más (*por ejemplo*, al menos o fijo en dos, tres, cuatro, cinco, seis, siete, ocho, nueve o 10) de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;
- b. todos los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;
- 35 c. uno o más (*por ejemplo*, al menos o fijo en dos, tres, cuatro, cinco, seis, siete, ocho, nueve, 10, 11, 12, 13, 14, 15 o 16) de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;
- d. uno o más (*por ejemplo*, al menos o fijo en dos, tres, cuatro, cinco, seis, siete, ocho, nueve o 10) de los siguientes:  
40 CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;
- e. uno o más (*por ejemplo*, al menos o fijo en dos, tres, cuatro, cinco o seis) de los siguientes: CD86, ITPKB, LRMP, MME, PTPRC y REL;  
45
- f. cualquier conjunto de genes que comprenda o consista en cualquiera, ya sea una o dos (en la medida no duplicativa) de las siguientes combinaciones de dos genes: {CD47,CD86} {CD47,IL16} {CD47,NF2} {CD47,PIM1} {CD47,PTPRC} {CD47,REL} {CD47,STAT3} {CD47,TNFRSF8} {CD47,TYMS} {CD86,IL16} {CD86,NF2} {CD86,PIM1} {CD86,PTPRC} {CD86,REL} {CD86,STAT3} {CD86,TNFRSF8} {CD86,TYMS} {IL16,NF2} {IL16,PIM1} {IL16,PTPRC} {IL16,REL} {IL16,STAT3} {IL16,TNFRSF8} {IL16,TYMS} {NF2,PIM1} {NF2,PTPRC} {NF2,REL} {NF2,STAT3} {NF2,TNFRSF8} {NF2,TYMS} {PIM1,PTPRC} {PIM1,REL} {PIM1,STAT3} {PIM1,TNFRSF8} {PIM1,TYMS} {PTPRC,REL} {PTPRC,STAT3} {PTPRC,TNFRSF8} {PTPRC,TYMS} {REL,STAT3} {REL,TNFRSF8} {REL,TYMS} {STAT3,TNFRSF8} {STAT3,TYMS} {TNFRSF8,TYMS};  
50
- g. cualquier conjunto de genes que comprenda o consista en cualquiera, ya sea una o dos (en la medida no duplicativa) de las siguientes combinaciones de tres genes: {CD47,CD86,IL16} {CD47,CD86,NF2} {CD47,CD86,PIM1} {CD47,CD86,PTPRC} {CD47,CD86,REL} {CD47,CD86,STAT3} {CD47,CD86,TNFRSF8} {CD47,CD86,TYMS} {CD47,IL16,NF2} {CD47,IL16,PIM1} {CD47,IL16,PTPRC} {CD47,IL16,REL} {CD47,IL16,STAT3} {CD47,IL16,TNFRSF8} {CD47,IL16,TYMS} {CD47,NF2,PIM1} {CD47,NF2,PTPRC} {CD47,NF2,REL} {CD47,NF2,STAT3} {CD47,NF2,TNFRSF8} {CD47,NF2,TYMS} {CD47,PIM1,PTPRC} {CD47,PIM1,REL} {CD47,PIM1,STAT3} {CD47,PIM1,TNFRSF8} {CD47,PIM1,TYMS} {CD47,PTPRC,REL} {CD47,PTPRC,STAT3} {CD47,PTPRC,TNFRSF8} {CD47,PTPRC,TYMS} {CD47,REL,STAT3} {CD47,REL,TNFRSF8} {CD47,REL,TYMS};  
60

- {CD47,STAT3,TNFRSF8} {CD47,STAT3,TYMS} {CD47,TNFRSF8,TYMS} {CD86,IL16,NF2} {CD86,IL16,PIM1}  
 {CD86,IL16,PTPRC} {CD86,IL16,REL} {CD86,IL16,STAT3} {CD86,IL16,TNFRSF8} {CD86,IL16,TYMS}  
 {CD86,NF2,PIM1} {CD86,NF2,PTPRC} {CD86,NF2,REL} {CD86,NF2,STAT3} {CD86,NF2,TNFRSF8}  
 {CD86,NF2,TYMS} {CD86,PIM1,PTPRC} {CD86,PIM1,REL} {CD86,PIM1,STAT3} {CD86,PIM1,TNFRSF8}
- 5 {CD86,PIM1,TYMS} {CD86,PTPRC,REL} {CD86,PTPRC,STAT3} {CD86,PTPRC,TNFRSF8} {CD86,PTPRC,TYMS}  
 {CD86,REL,STAT3} {CD86,REL,TNFRSF8} {CD86,REL,TYMS} {CD86,STAT3,TNFRSF8} {CD86,STAT3,TYMS}  
 {CD86,TNFRSF8,TYMS} {IL16,NF2,PIM1} {IL16,NF2,PTPRC} {IL16,NF2,REL} {IL16,NF2,STAT3}  
 {IL16,NF2,TNFRSF8} {IL16,NF2,TYMS} {IL16,PIM1,PTPRC} {IL16,PIM1,REL} {IL16,PIM1,STAT3}  
 {IL16,PIM1,TNFRSF8} {IL16,PIM1,TYMS} {IL16,PTPRC,REL} {IL16,PTPRC,STAT3} {IL16,PTPRC,TNFRSF8}
- 10 {IL16,PTPRC,TYMS} {IL16,REL,STAT3} {IL16,REL,TNFRSF8} {IL16,REL,TYMS} {IL16,STAT3,TNFRSF8}  
 {IL16,STAT3,TYMS} {IL16,TNFRSF8,TYMS} {NF2,PIM1,PTPRC} {NF2,PIM1,REL} {NF2,PIM1,STAT3}  
 {NF2,PIM1,TNFRSF8} {NF2,PIM1,TYMS} {NF2,PTPRC,REL} {NF2,PTPRC,STAT3} {NF2,PTPRC,TNFRSF8}  
 {NF2,PTPRC,TYMS} {NF2,REL,STAT3} {NF2,REL,TNFRSF8} {NF2,REL,TYMS} {NF2,STAT3,TNFRSF8}  
 {NF2,STAT3,TYMS} {NF2,TNFRSF8,TYMS} {PIM1,PTPRC,REL} {PIM1,PTPRC,STAT3} {PIM1,PTPRC,TNFRSF8}
- 15 {PIM1,PTPRC,TYMS} {PIM1,REL,STAT3} {PIM1,REL,TNFRSF8} {PIM1,REL,TYMS} {PIM1,STAT3,TNFRSF8}  
 {PIM1,STAT3,TYMS} {PIM1,TNFRSF8,TYMS} {PTPRC,REL,STAT3} {PTPRC,REL,TNFRSF8} {PTPRC,REL,TYMS}  
 {PTPRC,STAT3,TNFRSF8} {PTPRC,STAT3,TYMS} {PTPRC,TNFRSF8,TYMS} {REL,STAT3,TNFRSF8}  
 {REL,STAT3,TYMS} {REL,TNFRSF8,TYMS} {STAT3,TNFRSF8,TYMS}
- 20 h. cualquier conjunto de genes que comprenda o consista en cualquiera, ya sea una o dos (en la medida no duplicativa)  
 de las siguientes combinaciones de cuatro genes: {CD47,CD86,IL16,NF2} {CD47,CD86,IL16,PIM1}  
 {CD47,CD86,IL16,PTPRC} {CD47,CD86,IL16,REL} {CD47,CD86,IL16,STAT3} {CD47,CD86,IL16,TNFRSF8}  
 {CD47,CD86,IL16,TYMS} {CD47,CD86,NF2,PIM1} {CD47,CD86,NF2,PTPRC} {CD47,CD86,NF2,REL}  
 {CD47,CD86,NF2,STAT3} {CD47,CD86,NF2,TNFRSF8} {CD47,CD86,NF2,TYMS} {CD47,CD86,PIM1,PTPRC}
- 25 {CD47,CD86,PIM1,REL} {CD47,CD86,PIM1,STAT3} {CD47,CD86,PIM1,TNFRSF8} {CD47,CD86,PIM1,TYMS}  
 {CD47,CD86,PTPRC,REL} {CD47,CD86,PTPRC,STAT3} {CD47,CD86,PTPRC,TNFRSF8}  
 {CD47,CD86,PTPRC,TYMS} {CD47,CD86,REL,STAT3} {CD47,CD86,REL,TNFRSF8} {CD47,CD86,REL,TYMS}  
 {CD47,CD86,STAT3,TNFRSF8} {CD47,CD86,STAT3,TYMS} {CD47,CD86,TNFRSF8,TYMS} {CD47,IL16,NF2,PIM1}  
 {CD47,IL16,NF2,PTPRC} {CD47,IL16,NF2,REL} {CD47,IL16,NF2,STAT3} {CD47,IL16,NF2,TNFRSF8}
- 30 {CD47,IL16,NF2,TYMS} {CD47,IL16,PIM1,PTPRC} {CD47,IL16,PIM1,REL} {CD47,IL16,PIM1,STAT3}  
 {CD47,IL16,PIM1,TNFRSF8} {CD47,IL16,PIM1,TYMS} {CD47,IL16,PTPRC,REL} {CD47,IL16,PTPRC,STAT3}  
 {CD47,IL16,PTPRC,TNFRSF8} {CD47,IL16,PTPRC,TYMS} {CD47,IL16,REL,STAT3} {CD47,IL16,REL,TNFRSF8}  
 {CD47,IL16,REL,TYMS} {CD47,IL16,STAT3,TNFRSF8} {CD47,IL16,STAT3,TYMS} {CD47,IL16,TNFRSF8,TYMS}  
 {CD47,NF2,PIM1,PTPRC} {CD47,NF2,PIM1,REL} {CD47,NF2,PIM1,STAT3} {CD47,NF2,PIM1,TNFRSF8}
- 35 {CD47,NF2,PIM1,TYMS} {CD47,NF2,PTPRC,REL} {CD47,NF2,PTPRC,STAT3} {CD47,NF2,PTPRC,TNFRSF8}  
 {CD47,NF2,PTPRC,TYMS} {CD47,NF2,REL,STAT3} {CD47,NF2,REL,TNFRSF8} {CD47,NF2,REL,TYMS}  
 {CD47,NF2,STAT3,TNFRSF8} {CD47,NF2,STAT3,TYMS} {CD47,NF2,TNFRSF8,TYMS} {CD47,PIM1,PTPRC,REL}  
 {CD47,PIM1,PTPRC,STAT3} {CD47,PIM1,PTPRC,TNFRSF8} {CD47,PIM1,PTPRC,TYMS} {CD47,PIM1,REL,STAT3}  
 {CD47,PIM1,REL,TNFRSF8} {CD47,PIM1,REL,TYMS} {CD47,PIM1,STAT3,TNFRSF8} {CD47,PIM1,STAT3,TYMS}
- 40 {CD47,PIM1,TNFRSF8,TYMS} {CD47,PTPRC,REL,STAT3} {CD47,PTPRC,REL,TNFRSF8}  
 {CD47,PTPRC,REL,TYMS} {CD47,PTPRC,STAT3,TNFRSF8} {CD47,PTPRC,STAT3,TYMS}  
 {CD47,PTPRC,TNFRSF8,TYMS} {CD47,REL,STAT3,TNFRSF8} {CD47,REL,STAT3,TYMS}  
 {CD47,REL,TNFRSF8,TYMS} {CD47,STAT3,TNFRSF8,TYMS} {CD86,IL16,NF2,PIM1} {CD86,IL16,NF2,PTPRC}
- 45 {CD86,IL16,NF2,REL} {CD86,IL16,NF2,STAT3} {CD86,IL16,NF2,TNFRSF8} {CD86,IL16,NF2,TYMS}  
 {CD86,IL16,PIM1,PTPRC} {CD86,IL16,PIM1,REL} {CD86,IL16,PIM1,STAT3} {CD86,IL16,PIM1,TNFRSF8}  
 {CD86,IL16,PIM1,TYMS} {CD86,IL16,PTPRC,REL} {CD86,IL16,PTPRC,STAT3} {CD86,IL16,PTPRC,TNFRSF8}  
 {CD86,IL16,PTPRC,TYMS} {CD86,IL16,REL,STAT3} {CD86,IL16,REL,TNFRSF8} {CD86,IL16,REL,TYMS}  
 {CD86,IL16,STAT3,TNFRSF8} {CD86,IL16,STAT3,TYMS} {CD86,IL16,TNFRSF8,TYMS} {CD86,NF2,PIM1,PTPRC}
- 50 {CD86,NF2,PIM1,REL} {CD86,NF2,PIM1,STAT3} {CD86,NF2,PIM1,TNFRSF8} {CD86,NF2,PIM1,TYMS}  
 {CD86,NF2,PTPRC,REL} {CD86,NF2,PTPRC,STAT3} {CD86,NF2,PTPRC,TNFRSF8} {CD86,NF2,PTPRC,TYMS}  
 {CD86,NF2,REL,STAT3} {CD86,NF2,REL,TNFRSF8} {CD86,NF2,REL,TYMS} {CD86,NF2,STAT3,TNFRSF8}  
 {CD86,NF2,STAT3,TYMS} {CD86,NF2,TNFRSF8,TYMS} {CD86,PIM1,PTPRC,REL} {CD86,PIM1,PTPRC,STAT3}  
 {CD86,PIM1,PTPRC,TNFRSF8} {CD86,PIM1,PTPRC,TYMS} {CD86,PIM1,REL,STAT3} {CD86,PIM1,REL,TNFRSF8}
- 55 {CD86,PIM1,REL,TYMS} {CD86,PIM1,STAT3,TNFRSF8} {CD86,PIM1,STAT3,TYMS} {CD86,PIM1,TNFRSF8,TYMS}  
 {CD86,PTPRC,REL,STAT3} {CD86,PTPRC,REL,TNFRSF8} {CD86,PTPRC,REL,TYMS}  
 {CD86,PTPRC,STAT3,TNFRSF8} {CD86,PTPRC,STAT3,TYMS} {CD86,PTPRC,TNFRSF8,TYMS}  
 {CD86,REL,STAT3,TNFRSF8} {CD86,REL,STAT3,TYMS} {CD86,REL,TNFRSF8,TYMS}
- 60 {CD86,STAT3,TNFRSF8,TYMS} {IL16,NF2,PIM1,PTPRC} {IL16,NF2,PIM1,REL} {IL16,NF2,PIM1,STAT3}  
 {IL16,NF2,PIM1,TNFRSF8} {IL16,NF2,PIM1,TYMS} {IL16,NF2,PTPRC,REL} {IL16,NF2,PTPRC,STAT3}  
 {IL16,NF2,PTPRC,TNFRSF8} {IL16,NF2,PTPRC,TYMS} {IL16,NF2,REL,STAT3} {IL16,NF2,REL,TNFRSF8}  
 {IL16,NF2,REL,TYMS} {IL16,NF2,STAT3,TNFRSF8} {IL16,NF2,STAT3,TYMS} {IL16,NF2,TNFRSF8,TYMS}  
 {IL16,PIM1,PTPRC,REL} {IL16,PIM1,PTPRC,STAT3} {IL16,PIM1,PTPRC,TNFRSF8} {IL16,PIM1,PTPRC,TYMS}

{IL16,PIM1,REL,STAT3} {IL16,PIM1,REL,TNFRSF8} {IL16,PIM1,REL,TYMS} {IL16,PIM1,STAT3,TNFRSF8}  
 {IL16,PIM1,STAT3,TYMS} {IL16,PIM1,TNFRSF8,TYMS} {IL16,PTPRC,REL,STAT3} {IL16,PTPRC,REL,TNFRSF8}  
 {IL16,PTPRC,REL,TYMS} {IL16,PTPRC,STAT3,TNFRSF8} {IL16,PTPRC,STAT3,TYMS}  
 {IL16,PTPRC,TNFRSF8,TYMS} {IL16,REL,STAT3,TNFRSF8} {IL16,REL,STAT3,TYMS} {IL16,REL,TNFRSF8,TYMS}  
 5 {IL16,STAT3,TNFRSF8,TYMS} {NF2,PIM1,PTPRC,REL} {NF2,PIM1,PTPRC,STAT3} {NF2,PIM1,PTPRC,TNFRSF8}  
 {NF2,PIM1,PTPRC,TYMS} {NF2,PIM1,REL,STAT3} {NF2,PIM1,REL,TNFRSF8} {NF2,PIM1,REL,TYMS}  
 {NF2,PIM1,STAT3,TNFRSF8} {NF2,PIM1,STAT3,TYMS} {NF2,PIM1,TNFRSF8,TYMS} {NF2,PTPRC,REL,STAT3}  
 {NF2,PTPRC,REL,TNFRSF8} {NF2,PTPRC,REL,TYMS} {NF2,PTPRC,STAT3,TNFRSF8}  
 {NF2,PTPRC,STAT3,TYMS} {NF2,PTPRC,TNFRSF8,TYMS} {NF2,REL,STAT3,TNFRSF8} {NF2,REL,STAT3,TYMS}  
 10 {NF2,REL,TNFRSF8,TYMS} {NF2,STAT3,TNFRSF8,TYMS} {PIM1,PTPRC,REL,STAT3}  
 {PIM1,PTPRC,REL,TNFRSF8} {PIM1,PTPRC,REL,TYMS} {PIM1,PTPRC,STAT3,TNFRSF8}  
 {PIM1,PTPRC,STAT3,TYMS} {PIM1,PTPRC,TNFRSF8,TYMS} {PIM1,REL,STAT3,TNFRSF8}  
 {PIM1,REL,STAT3,TYMS} {PIM1,REL,TNFRSF8,TYMS} {PIM1,STAT3,TNFRSF8,TYMS}  
 {PTPRC,REL,STAT3,TNFRSF8} {PTPRC,REL,STAT3,TYMS} {PTPRC,REL,TNFRSF8,TYMS}  
 15 {PTPRC,STAT3,TNFRSF8,TYMS} {REL,STAT3,TNFRSF8,TYMS};

i. cualquier conjunto de genes que comprenda o consista en cualquiera, ya sea una o dos (en la medida no duplicativa) de las siguientes combinaciones de dos genes: {CD47,CD86} {CD47,ENTPD1} {CD47,FOXP1} {CD47,FUT8}  
 {CD47,IL16} {CD47,ITPKB} {CD47,LRMP} {CD47,MME} {CD47,NF2} {CD47,PIM1} {CD47,PTPRC} {CD47,REL}  
 20 {CD47,STAT3} {CD47,TNFRSF8} {CD47,TYMS} {CD86,ENTPD1} {CD86,FOXP1} {CD86,FUT8} {CD86,IL16}  
 {CD86,ITPKB} {CD86,LRMP} {CD86,MME} {CD86,NF2} {CD86,PIM1} {CD86,PTPRC} {CD86,REL} {CD86,STAT3}  
 {CD86,TNFRSF8} {CD86,TYMS} {ENTPD1,FOXP1} {ENTPD1,FUT8} {ENTPD1,IL16} {ENTPD1,ITPKB}  
 {ENTPD1,LRMP} {ENTPD1,MME} {ENTPD1,NF2} {ENTPD1,PIM1} {ENTPD1,PTPRC} {ENTPD1,REL}  
 {ENTPD1,STAT3} {ENTPD1,TNFRSF8} {ENTPD1,TYMS} {FOXP1,FUT8} {FOXP1,IL16} {FOXP1,ITPKB}  
 25 {FOXP1,LRMP} {FOXP1,MME} {FOXP1,NF2} {FOXP1,PIM1} {FOXP1,PTPRC} {FOXP1,REL} {FOXP1,STAT3}  
 {FOXP1,TNFRSF8} {FOXP1,TYMS} {FUT8,IL16} {FUT8,ITPKB} {FUT8,LRMP} {FUT8,MME} {FUT8,NF2}  
 {FUT8,PIM1} {FUT8,PTPRC} {FUT8,REL} {FUT8,STAT3} {FUT8,TNFRSF8} {FUT8,TYMS} {IL16,ITPKB} {IL16,LRMP}  
 {IL16,MME} {IL16,NF2} {IL16,PIM1} {IL16,PTPRC} {IL16,REL} {IL16,STAT3} {IL16,TNFRSF8} {IL16,TYMS}  
 {ITPKB,LRMP} {ITPKB,MME} {ITPKB,NF2} {ITPKB,PIM1} {ITPKB,PTPRC} {ITPKB,REL} {ITPKB,STAT3}  
 30 {ITPKB,TNFRSF8} {ITPKB,TYMS} {LRMP,MME} {LRMP,NF2} {LRMP,PIM1} {LRMP,PTPRC} {LRMP,REL}  
 {LRMP,STAT3} {LRMP,TNFRSF8} {LRMP,TYMS} {MME,NF2} {MME,PIM1} {MME,PTPRC} {MME,REL}  
 {MME,STAT3} {MME,TNFRSF8} {MME,TYMS} {NF2,PIM1} {NF2,PTPRC} {NF2,REL} {NF2,STAT3} {NF2,TNFRSF8}  
 {NF2,TYMS} {PIM1,PTPRC} {PIM1,REL} {PIM1,STAT3} {PIM1,TNFRSF8} {PIM1,TYMS} {PTPRC,REL}  
 {PTPRC,STAT3} {PTPRC,TNFRSF8} {PTPRC,TYMS} {REL,STAT3} {REL,TNFRSF8} {REL,TYMS}  
 35 {STAT3,TNFRSF8} {STAT3,TYMS} {TNFRSF8,TYMS}; y

j. cualquier conjunto de genes que comprenda o consista en cualquiera, ya sea una o dos (en la medida no duplicativa) de las siguientes combinaciones de tres genes: {CD47,CD86,ENTPD1} {CD47,CD86,FOXP1} {CD47,CD86,FUT8}  
 {CD47,CD86,IL16} {CD47,CD86,ITPKB} {CD47,CD86,LRMP} {CD47,CD86,MME} {CD47,CD86,NF2}  
 40 {CD47,CD86,PIM1} {CD47,CD86,PTPRC} {CD47,CD86,REL} {CD47,CD86,STAT3} {CD47,CD86,TNFRSF8}  
 {CD47,CD86,TYMS} {CD47,ENTPD1,FOXP1} {CD47,ENTPD1,FUT8} {CD47,ENTPD1,IL16} {CD47,ENTPD1,ITPKB}  
 {CD47,ENTPD1,LRMP} {CD47,ENTPD1,MME} {CD47,ENTPD1,NF2} {CD47,ENTPD1,PIM1}  
 {CD47,ENTPD1,PTPRC} {CD47,ENTPD1,REL} {CD47,ENTPD1,STAT3} {CD47,ENTPD1,TNFRSF8}  
 {CD47,ENTPD1,TYMS} {CD47,FOXP1,FUT8} {CD47,FOXP1,IL16} {CD47,FOXP1,ITPKB} {CD47,FOXP1,LRMP}  
 45 {CD47,FOXP1,MME} {CD47,FOXP1,NF2} {CD47,FOXP1,PIM1} {CD47,FOXP1,PTPRC} {CD47,FOXP1,REL}  
 {CD47,FOXP1,STAT3} {CD47,FOXP1,TNFRSF8} {CD47,FOXP1,TYMS} {CD47,FUT8,IL16} {CD47,FUT8,ITPKB}  
 {CD47,FUT8,LRMP} {CD47,FUT8,MME} {CD47,FUT8,NF2} {CD47,FUT8,PIM1} {CD47,FUT8,PTPRC}  
 {CD47,FUT8,REL} {CD47,FUT8,STAT3} {CD47,FUT8,TNFRSF8} {CD47,FUT8,TYMS} {CD47,IL16,ITPKB}  
 {CD47,IL16,LRMP} {CD47,IL16,MME} {CD47,IL16,NF2} {CD47,IL16,PIM1} {CD47,IL16,PTPRC} {CD47,IL16,REL}  
 50 {CD47,IL16,STAT3} {CD47,IL16,TNFRSF8} {CD47,IL16,TYMS} {CD47,ITPKB,LRMP} {CD47,ITPKB,MME}  
 {CD47,ITPKB,NF2} {CD47,ITPKB,PIM1} {CD47,ITPKB,PTPRC} {CD47,ITPKB,REL} {CD47,ITPKB,STAT3}  
 {CD47,ITPKB,TNFRSF8} {CD47,ITPKB,TYMS} {CD47,LRMP,MME} {CD47,LRMP,NF2} {CD47,LRMP,PIM1}  
 {CD47,LRMP,PTPRC} {CD47,LRMP,REL} {CD47,LRMP,STAT3} {CD47,LRMP,TNFRSF8} {CD47,LRMP,TYMS}  
 {CD47,MME,NF2} {CD47,MME,PIM1} {CD47,MME,PTPRC} {CD47,MME,REL} {CD47,MME,STAT3}  
 55 {CD47,MME,TNFRSF8} {CD47,MME,TYMS} {CD47,NF2,PIM1} {CD47,NF2,PTPRC} {CD47,NF2,REL}  
 {CD47,NF2,STAT3} {CD47,NF2,TNFRSF8} {CD47,NF2,TYMS} {CD47,PIM1,PTPRC} {CD47,PIM1,REL}  
 {CD47,PIM1,STAT3} {CD47,PIM1,TNFRSF8} {CD47,PIM1,TYMS} {CD47,PTPRC,REL} {CD47,PTPRC,STAT3}  
 {CD47,PTPRC,TNFRSF8} {CD47,PTPRC,TYMS} {CD47,REL,STAT3} {CD47,REL,TNFRSF8} {CD47,REL,TYMS}  
 {CD47,STAT3,TNFRSF8} {CD47,STAT3,TYMS} {CD47,TNFRSF8,TYMS} {CD86,ENTPD1,FOXP1}  
 60 {CD86,ENTPD1,FUT8} {CD86,ENTPD1,IL16} {CD86,ENTPD1,ITPKB} {CD86,ENTPD1,LRMP} {CD86,ENTPD1,MME}  
 {CD86,ENTPD1,NF2} {CD86,ENTPD1,PIM1} {CD86,ENTPD1,PTPRC} {CD86,ENTPD1,REL}  
 {CD86,ENTPD1,STAT3} {CD86,ENTPD1,TNFRSF8} {CD86,ENTPD1,TYMS} {CD86,FOXP1,FUT8}

{CD86,FOXP1,IL16} {CD86,FOXP1,ITPKB} {CD86,FOXP1,LRMP} {CD86,FOXP1,MME} {CD86,FOXP1,NF2}  
 {CD86,FOXP1,PIM1} {CD86,FOXP1,PTPRC} {CD86,FOXP1,REL} {CD86,FOXP1,STAT3} {CD86,FOXP1,TNFRSF8}  
 {CD86,FOXP1,TYMS} {CD86,FUT8,IL16} {CD86,FUT8,ITPKB} {CD86,FUT8,LRMP} {CD86,FUT8,MME}  
 {CD86,FUT8,NF2} {CD86,FUT8,PIM1} {CD86,FUT8,PTPRC} {CD86,FUT8,REL} {CD86,FUT8,STAT3}  
 5 {CD86,FUT8,TNFRSF8} {CD86,FUT8,TYMS} {CD86,IL16,ITPKB} {CD86,IL16,LRMP} {CD86,IL16,MME}  
 {CD86,IL16,NF2} {CD86,IL16,PIM1} {CD86,IL16,PTPRC} {CD86,IL16,REL} {CD86,IL16,STAT3}  
 {CD86,IL16,TNFRSF8} {CD86,IL16,TYMS} {CD86,ITPKB,LRMP} {CD86,ITPKB,MME} {CD86,ITPKB,NF2}  
 {CD86,ITPKB,PIM1} {CD86,ITPKB,PTPRC} {CD86,ITPKB,REL} {CD86,ITPKB,STAT3} {CD86,ITPKB,TNFRSF8}  
 {CD86,ITPKB,TYMS} {CD86,LRMP,MME} {CD86,LRMP,NF2} {CD86,LRMP,PIM1} {CD86,LRMP,PTPRC}  
 10 {CD86,LRMP,REL} {CD86,LRMP,STAT3} {CD86,LRMP,TNFRSF8} {CD86,LRMP,TYMS} {CD86,MME,NF2}  
 {CD86,MME,PIM1} {CD86,MME,PTPRC} {CD86,MME,REL} {CD86,MME,STAT3} {CD86,MME,TNFRSF8}  
 {CD86,MME,TYMS} {CD86,NF2,PIM1} {CD86,NF2,PTPRC} {CD86,NF2,REL} {CD86,NF2,STAT3}  
 {CD86,NF2,TNFRSF8} {CD86,NF2,TYMS} {CD86,PIM1,PTPRC} {CD86,PIM1,REL} {CD86,PIM1,STAT3}  
 {CD86,PIM1,TNFRSF8} {CD86,PIM1,TYMS} {CD86,PTPRC,REL} {CD86,PTPRC,STAT3} {CD86,PTPRC,TNFRSF8}  
 15 {CD86,PTPRC,TYMS} {CD86,REL,STAT3} {CD86,REL,TNFRSF8} {CD86,REL,TYMS} {CD86,STAT3,TNFRSF8}  
 {CD86,STAT3,TYMS} {CD86,TNFRSF8,TYMS} {ENTPD1,FOXP1,FUT8} {ENTPD1,FOXP1,IL16}  
 {ENTPD1,FOXP1,ITPKB} {ENTPD1,FOXP1,LRMP} {ENTPD1,FOXP1,MME} {ENTPD1,FOXP1,NF2}  
 {ENTPD1,FOXP1,PIM1} {ENTPD1,FOXP1,PTPRC} {ENTPD1,FOXP1,REL} {ENTPD1,FOXP1,STAT3}  
 {ENTPD1,FOXP1,TNFRSF8} {ENTPD1,FOXP1,TYMS} {ENTPD1,FUT8,IL16} {ENTPD1,FUT8,ITPKB}  
 20 {ENTPD1,FUT8,LRMP} {ENTPD1,FUT8,MME} {ENTPD1,FUT8,NF2} {ENTPD1,FUT8,PIM1} {ENTPD1,FUT8,PTPRC}  
 {ENTPD1,FUT8,REL} {ENTPD1,FUT8,STAT3} {ENTPD1,FUT8,TNFRSF8} {ENTPD1,FUT8,TYMS}  
 {ENTPD1,IL16,ITPKB} {ENTPD1,IL16,LRMP} {ENTPD1,IL16,MME} {ENTPD1,IL16,NF2} {ENTPD1,IL16,PIM1}  
 {ENTPD1,IL16,PTPRC} {ENTPD1,IL16,REL} {ENTPD1,IL16,STAT3} {ENTPD1,IL16,TNFRSF8}  
 {ENTPD1,IL16,TYMS} {ENTPD1,ITPKB,LRMP} {ENTPD1,ITPKB,MME} {ENTPD1,ITPKB,NF2}  
 25 {ENTPD1,ITPKB,PIM1} {ENTPD1,ITPKB,PTPRC} {ENTPD1,ITPKB,REL} {ENTPD1,ITPKB,STAT3}  
 {ENTPD1,ITPKB,TNFRSF8} {ENTPD1,ITPKB,TYMS} {ENTPD1,LRMP,MME} {ENTPD1,LRMP,NF2}  
 {ENTPD1,LRMP,PIM1} {ENTPD1,LRMP,PTPRC} {ENTPD1,LRMP,REL} {ENTPD1,LRMP,STAT3}  
 {ENTPD1,LRMP,TNFRSF8} {ENTPD1,LRMP,TYMS} {ENTPD1,MME,NF2} {ENTPD1,MME,PIM1}  
 {ENTPD1,MME,PTPRC} {ENTPD1,MME,REL} {ENTPD1,MME,STAT3} {ENTPD1,MME,TNFRSF8}  
 30 {ENTPD1,MME,TYMS} {ENTPD1,NF2,PIM1} {ENTPD1,NF2,PTPRC} {ENTPD1,NF2,REL} {ENTPD1,NF2,STAT3}  
 {ENTPD1,NF2,TNFRSF8} {ENTPD1,NF2,TYMS} {ENTPD1,PIM1,PTPRC} {ENTPD1,PIM1,REL}  
 {ENTPD1,PIM1,STAT3} {ENTPD1,PIM1,TNFRSF8} {ENTPD1,PIM1,TYMS} {ENTPD1,PTPRC,REL}  
 {ENTPD1,PTPRC,STAT3} {ENTPD1,PTPRC,TNFRSF8} {ENTPD1,PTPRC,TYMS} {ENTPD1,REL,STAT3}  
 {ENTPD1,REL,TNFRSF8} {ENTPD1,REL,TYMS} {ENTPD1,STAT3,TNFRSF8} {ENTPD1,STAT3,TYMS}  
 35 {ENTPD1,TNFRSF8,TYMS} {FOXP1,FUT8,IL16} {FOXP1,FUT8,ITPKB} {FOXP1,FUT8,LRMP} {FOXP1,FUT8,MME}  
 {FOXP1,FUT8,NF2} {FOXP1,FUT8,PIM1} {FOXP1,FUT8,PTPRC} {FOXP1,FUT8,REL} {FOXP1,FUT8,STAT3}  
 {FOXP1,FUT8,TNFRSF8} {FOXP1,FUT8,TYMS} {FOXP1,IL16,ITPKB} {FOXP1,IL16,LRMP} {FOXP1,IL16,MME}  
 {FOXP1,IL16,NF2} {FOXP1,IL16,PIM1} {FOXP1,IL16,PTPRC} {FOXP1,IL16,REL} {FOXP1,IL16,STAT3}  
 {FOXP1,IL16,TNFRSF8} {FOXP1,IL16,TYMS} {FOXP1,ITPKB,LRMP} {FOXP1,ITPKB,MME} {FOXP1,ITPKB,NF2}  
 40 {FOXP1,ITPKB,PIM1} {FOXP1,ITPKB,PTPRC} {FOXP1,ITPKB,REL} {FOXP1,ITPKB,STAT3}  
 {FOXP1,ITPKB,TNFRSF8} {FOXP1,ITPKB,TYMS} {FOXP1,LRMP,MME} {FOXP1,LRMP,NF2} {FOXP1,LRMP,PIM1}  
 {FOXP1,LRMP,PTPRC} {FOXP1,LRMP,REL} {FOXP1,LRMP,STAT3} {FOXP1,LRMP,TNFRSF8}  
 {FOXP1,LRMP,TYMS} {FOXP1,MME,NF2} {FOXP1,MME,PIM1} {FOXP1,MME,PTPRC} {FOXP1,MME,REL}  
 {FOXP1,MME,STAT3} {FOXP1,MME,TNFRSF8} {FOXP1,MME,TYMS} {FOXP1,NF2,PIM1} {FOXP1,NF2,PTPRC}  
 45 {FOXP1,NF2,REL} {FOXP1,NF2,STAT3} {FOXP1,NF2,TNFRSF8} {FOXP1,NF2,TYMS} {FOXP1,PIM1,PTPRC}  
 {FOXP1,PIM1,REL} {FOXP1,PIM1,STAT3} {FOXP1,PIM1,TNFRSF8} {FOXP1,PIM1,TYMS} {FOXP1,PTPRC,REL}  
 {FOXP1,PTPRC,STAT3} {FOXP1,PTPRC,TNFRSF8} {FOXP1,PTPRC,TYMS} {FOXP1,REL,STAT3}  
 {FOXP1,REL,TNFRSF8} {FOXP1,REL,TYMS} {FOXP1,STAT3,TNFRSF8} {FOXP1,STAT3,TYMS}  
 {FOXP1,TNFRSF8,TYMS} {FUT8,IL16,ITPKB} {FUT8,IL16,LRMP} {FUT8,IL16,MME} {FUT8,IL16,NF2}  
 50 {FUT8,IL16,PIM1} {FUT8,IL16,PTPRC} {FUT8,IL16,REL} {FUT8,IL16,STAT3} {FUT8,IL16,TNFRSF8}  
 {FUT8,IL16,TYMS} {FUT8,ITPKB,LRMP} {FUT8,ITPKB,MME} {FUT8,ITPKB,NF2} {FUT8,ITPKB,PIM1}  
 {FUT8,ITPKB,PTPRC} {FUT8,ITPKB,REL} {FUT8,ITPKB,STAT3} {FUT8,ITPKB,TNFRSF8} {FUT8,ITPKB,TYMS}  
 {FUT8,LRMP,MME} {FUT8,LRMP,NF2} {FUT8,LRMP,PIM1} {FUT8,LRMP,PTPRC} {FUT8,LRMP,REL}  
 {FUT8,LRMP,STAT3} {FUT8,LRMP,TNFRSF8} {FUT8,LRMP,TYMS} {FUT8,MME,NF2} {FUT8,MME,PIM1}  
 55 {FUT8,MME,PTPRC} {FUT8,MME,REL} {FUT8,MME,STAT3} {FUT8,MME,TNFRSF8} {FUT8,MME,TYMS}  
 {FUT8,NF2,PIM1} {FUT8,NF2,PTPRC} {FUT8,NF2,REL} {FUT8,NF2,STAT3} {FUT8,NF2,TNFRSF8}  
 {FUT8,NF2,TYMS} {FUT8,PIM1,PTPRC} {FUT8,PIM1,REL} {FUT8,PIM1,STAT3} {FUT8,PIM1,TNFRSF8}  
 {FUT8,PIM1,TYMS} {FUT8,PTPRC,REL} {FUT8,PTPRC,STAT3} {FUT8,PTPRC,TNFRSF8} {FUT8,PTPRC,TYMS}  
 {FUT8,REL,STAT3} {FUT8,REL,TNFRSF8} {FUT8,REL,TYMS} {FUT8,STAT3,TNFRSF8} {FUT8,STAT3,TYMS}  
 60 {FUT8,TNFRSF8,TYMS} {IL16,ITPKB,LRMP} {IL16,ITPKB,MME} {IL16,ITPKB,NF2} {IL16,ITPKB,PIM1}  
 {IL16,ITPKB,PTPRC} {IL16,ITPKB,REL} {IL16,ITPKB,STAT3} {IL16,ITPKB,TNFRSF8} {IL16,ITPKB,TYMS}  
 {IL16,LRMP,MME} {IL16,LRMP,NF2} {IL16,LRMP,PIM1} {IL16,LRMP,PTPRC} {IL16,LRMP,REL} {IL16,LRMP,STAT3}

{IL16,LRMP,TNFRSF8} {IL16,LRMP,TYMS} {IL16,MME,NF2} {IL16,MME,PIM1} {IL16,MME,PTPRC} {IL16,MME,REL}  
 {IL16,MME,STAT3} {IL16,MME,TNFRSF8} {IL16,MME,TYMS} {IL16,NF2,PIM1} {IL16,NF2,PTPRC} {IL16,NF2,REL}  
 {IL16,NF2,STAT3} {IL16,NF2,TNFRSF8} {IL16,NF2,TYMS} {IL16,PIM1,PTPRC} {IL16,PIM1,REL} {IL16,PIM1,STAT3}  
 {IL16,PIM1,TNFRSF8} {IL16,PIM1,TYMS} {IL16,PTPRC,REL} {IL16,PTPRC,STAT3} {IL16,PTPRC,TNFRSF8}  
 5 {IL16,PTPRC,TYMS} {IL16,REL,STAT3} {IL16,REL,TNFRSF8} {IL16,REL,TYMS} {IL16,STAT3,TNFRSF8}  
 {IL16,STAT3,TYMS} {IL16,TNFRSF8,TYMS} {ITPKB,LRMP,MME} {ITPKB,LRMP,NF2} {ITPKB,LRMP,PIM1}  
 {ITPKB,LRMP,PTPRC} {ITPKB,LRMP,REL} {ITPKB,LRMP,STAT3} {ITPKB,LRMP,TNFRSF8} {ITPKB,LRMP,TYMS}  
 {ITPKB,MME,NF2} {ITPKB,MME,PIM1} {ITPKB,MME,PTPRC} {ITPKB,MME,REL} {ITPKB,MME,STAT3}  
 {ITPKB,MME,TNFRSF8} {ITPKB,MME,TYMS} {ITPKB,NF2,PIM1} {ITPKB,NF2,PTPRC} {ITPKB,NF2,REL}  
 10 {ITPKB,NF2,STAT3} {ITPKB,NF2,TNFRSF8} {ITPKB,NF2,TYMS} {ITPKB,PIM1,PTPRC} {ITPKB,PIM1,REL}  
 {ITPKB,PIM1,STAT3} {ITPKB,PIM1,TNFRSF8} {ITPKB,PIM1,TYMS} {ITPKB,PTPRC,REL} {ITPKB,PTPRC,STAT3}  
 {ITPKB,PTPRC,TNFRSF8} {ITPKB,PTPRC,TYMS} {ITPKB,REL,STAT3} {ITPKB,REL,TNFRSF8} {ITPKB,REL,TYMS}  
 {ITPKB,STAT3,TNFRSF8} {ITPKB,STAT3,TYMS} {ITPKB,TNFRSF8,TYMS} {LRMP,MME,NF2} {LRMP,MME,PIM1}  
 {LRMP,MME,PTPRC} {LRMP,MME,REL} {LRMP,MME,STAT3} {LRMP,MME,TNFRSF8} {LRMP,MME,TYMS}  
 15 {LRMP,NF2,PIM1} {LRMP,NF2,PTPRC} {LRMP,NF2,REL} {LRMP,NF2,STAT3} {LRMP,NF2,TNFRSF8}  
 {LRMP,NF2,TYMS} {LRMP,PIM1,PTPRC} {LRMP,PIM1,REL} {LRMP,PIM1,STAT3} {LRMP,PIM1,TNFRSF8}  
 {LRMP,PIM1,TYMS} {LRMP,PTPRC,REL} {LRMP,PTPRC,STAT3} {LRMP,PTPRC,TNFRSF8} {LRMP,PTPRC,TYMS}  
 {LRMP,REL,STAT3} {LRMP,REL,TNFRSF8} {LRMP,REL,TYMS} {LRMP,STAT3,TNFRSF8} {LRMP,STAT3,TYMS}  
 {LRMP,TNFRSF8,TYMS} {MME,NF2,PIM1} {MME,NF2,PTPRC} {MME,NF2,REL} {MME,NF2,STAT3}  
 20 {MME,NF2,TNFRSF8} {MME,NF2,TYMS} {MME,PIM1,PTPRC} {MME,PIM1,REL} {MME,PIM1,STAT3}  
 {MME,PIM1,TNFRSF8} {MME,PIM1,TYMS} {MME,PTPRC,REL} {MME,PTPRC,STAT3} {MME,PTPRC,TNFRSF8}  
 {MME,PTPRC,TYMS} {MME,REL,STAT3} {MME,REL,TNFRSF8} {MME,REL,TYMS} {MME,STAT3,TNFRSF8}  
 {MME,STAT3,TYMS} {MME,TNFRSF8,TYMS} {NF2,PIM1,PTPRC} {NF2,PIM1,REL} {NF2,PIM1,STAT3}  
 {NF2,PIM1,TNFRSF8} {NF2,PIM1,TYMS} {NF2,PTPRC,REL} {NF2,PTPRC,STAT3} {NF2,PTPRC,TNFRSF8}  
 25 {NF2,PTPRC,TYMS} {NF2,REL,STAT3} {NF2,REL,TNFRSF8} {NF2,REL,TYMS} {NF2,STAT3,TNFRSF8}  
 {NF2,STAT3,TYMS} {NF2,TNFRSF8,TYMS} {PIM1,PTPRC,REL} {PIM1,PTPRC,STAT3} {PIM1,PTPRC,TNFRSF8}  
 {PIM1,PTPRC,TYMS} {PIM1,REL,STAT3} {PIM1,REL,TNFRSF8} {PIM1,REL,TYMS} {PIM1,STAT3,TNFRSF8}  
 {PIM1,STAT3,TYMS} {PIM1,TNFRSF8,TYMS} {PTPRC,REL,STAT3} {PTPRC,REL,TNFRSF8} {PTPRC,REL,TYMS}  
 {PTPRC,STAT3,TNFRSF8} {PTPRC,STAT3,TYMS} {PTPRC,TNFRSF8,TYMS} {REL,STAT3,TNFRSF8}  
 30 {REL,STAT3,TYMS} {REL,TNFRSF8,TYMS} {STAT3,TNFRSF8,TYMS}.

### III. Obtención de información de expresión génica

La expresión génica es el proceso que convierte la información heredada (*por ejemplo*, genes; regiones de ADN) en  
 35 el genoma de un organismo a productos funcionales específicos (a veces llamados productos genéticos). Los  
 productos genéticos incluyen ARN (ácido ribonucleico) y proteínas. La expresión génica en una muestra biológica  
 particular puede medirse detectando la expresión de productos genéticos, incluyendo el ARN (como ARNm, ARNt, ARNr  
 y/o miARN) o una proteína, en dicha muestra. En algunos ejemplos, la expresión génica se mide secuenciando una o  
 más sondas que se unen a un ácido nucleico diana (*por ejemplo*, ARNm).

40

Hay una variedad de técnicas disponibles (o que pueden volverse disponibles) para medir la expresión génica en una  
 muestra de interés. Sin embargo, la descripción no se limita a procedimientos particulares para obtener, medir o  
 detectar la expresión génica. Muchas de estas técnicas involucran detectar los productos de los genes (*por ejemplo*,  
 ácidos nucleicos (como ARN) y/o proteínas) expresados en tales muestras. También puede ser (o volverse posible)  
 45 detectar directamente la actividad de un gen o del ADN cromosómico (*por ejemplo*, la tasa de transcripción)  
 independientemente de medir sus productos genéticos resultantes y tales técnicas también son útiles en los  
 procedimientos descritos en esta invención.

La expresión génica se puede utilizar para determinar un subtipo de DLBCL. Por ejemplo, la expresión génica de dos  
 50 o más genes (como dos o más genes de firma de DLBCL) se utiliza para determinar si un tumor de DLBCL es del  
 subtipo GCB o no GCB (*por ejemplo*, ABC) o sin clasificar. En algunas realizaciones, la firma de DLBCL incluye (1)  
 dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8, and TYMS, o (2) dos o  
 más de (como todos) los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1,  
 PTPRC, REL, STAT3, TNFRSF8 y TYMS. Otros ejemplos específicos de marcador(es) de DLBCL incluyen dos o más  
 55 (como 2, 3, 4, 5, 6, 7, 8, 9 o 10) de los siguientes: CD47, ENTPD1, FOXP1, FUT8, IL16, NF2, PIM1, STAT3, TNFRSF8  
 y TYMS (que están asociados con la clasificación ABC); o dos o más (como 2, 3, 4, 5 o 6) de los siguientes: CD86,  
 ITPKB, LRMP, MME, PTPRC y REL (que están asociados con la clasificación GCB).

#### A. Detección de productos genéticos de ácido nucleico

60

Los productos genéticos de ácido nucleico son, como su nombre indica, productos de expresión génica que son ácidos  
 nucleicos. Los ejemplos de ácidos nucleicos incluyen ADN o ARN, como ADNc, ARN codificante de proteínas (*por*

*ejemplo*, ARNm) o ARN no codificante (*por ejemplo*, ARN largo no codificante (Inc)). El emparejamiento de bases entre cadenas complementarias de ARN o ADN (*es decir*, la hibridación de ácidos nucleicos) forma toda o parte de la base de una gran clase representativa de técnicas para detectar productos de genes de ácidos nucleicos. Otras técnicas de detección representativas involucran la secuenciación de ácidos nucleicos, la cual puede involucrar, o no, etapas de hibridación y/o etapas bioinformáticas (*por ejemplo*, para asociar información de secuencias de ácidos nucleicos a su gen correspondiente). Estos y otros procedimientos de detección de ácidos nucleicos son conocidos en la técnica y, aunque se describen técnicas representativas en esta invención, esta descripción no se limita a procedimientos particulares de detección de ácidos nucleicos.

## 10 **1. Aislamiento opcional del ácido nucleico**

En algunos ejemplos, los ácidos nucleicos se aíslan o extraen de la muestra de prueba antes de poner en contacto dichos ácidos nucleicos en la muestra con una sonda o cebador complementario de ácido nucleico y/o detectar dichos ácidos nucleicos en la muestra. Los ácidos nucleicos (como el ARN (*por ejemplo*, ARNm o ARNnc) o ADN) pueden aislarse de la muestra según cualquiera de una serie de procedimientos. Los procedimientos representativos de aislamiento y purificación de ácidos nucleicos se describen en detalle en el Capítulo 3 de *Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes*, Parte I. Theory and Nucleic Acid Preparation, P. Tijssen, ed. Elsevier, Nueva York (1993) y el Capítulo 3 de *Laboratory Techniques in Biochemistry and Molecular Biology: Hybridization With Nucleic Acid Probes*, Parte I. Theory and Nucleic Acid Preparation, P. Tijssen, ed. Elsevier, Nueva York (1993). Los procedimientos representativos para la extracción de ARN (*por ejemplo*, ARNm o ARNnc), de manera similar, son bien conocidos en la técnica y se describen en libros de texto estándar de biología molecular, incluyendo Ausubel y col., *Current Protocols of Molecular Biology*, John Wiley e hijos (1997).

Después del aislamiento o extracción de ácidos nucleicos (*por ejemplo*, ARN (como ARNm o ARNnc) o ADN) de una muestra, se puede realizar cualquiera de una serie de otras etapas opcionales para preparar dichos ácidos nucleicos para la detección, incluyendo la medición de la concentración del ácido nucleico aislado, la reparación (o recuperación) de ARN degradado o dañado, la transcripción inversa de ARN y/o la amplificación de ARN o ADN.

En otros ejemplos, una muestra (*por ejemplo*, una muestra de tejido FFPE) se suspende en un tampón (*por ejemplo*, un tampón de lisis) y los ácidos nucleicos (como ARN o ADN) presentes en la muestra suspendida no se aíslan ni se extraen (*por ejemplo*, purificados total o parcialmente) de dicha muestra suspendida y se ponen en contacto en dicha suspensión con una o más sondas de ácido nucleico complementarias (*por ejemplo*, las NPP o NPPF); eliminando así la necesidad de aislamiento o extracción de los ácidos nucleicos (*por ejemplo*, el ARN) de la muestra. Este procedimiento es particularmente ventajoso cuando los ácidos nucleicos (como ARN o ADN) presentes en la muestra suspendida están reticulados o fijados a estructuras celulares y no son fácilmente aislables o extraíbles. Las sondas relativamente cortas (*por ejemplo*, las sondas de detección de menos de 100 pares de bases, tales como de 75-25 pares de bases o 50-25 pares de bases) para la detección son útiles en algunas realizaciones de procedimientos sin extracción. En esta invención, se describen los procedimientos específicos (*por ejemplo*, qNPA y qNPS) para detectar ácidos nucleicos (*por ejemplo*, ARN) en una muestra sin extracción previa de dichos ácidos nucleicos.

40

## **2. Hibridación de ácidos nucleicos**

En algunos ejemplos, determinar el nivel de expresión de dos o más biomarcadores de DLBCL descritos (como dos o más de los de la Tabla 1) en los procedimientos proporcionados en esta invención puede incluir poner en contacto la muestra con una pluralidad de sondas de ácido nucleico (como NPP o NPPF + CFS) o cebadores de amplificación emparejados, donde cada sonda o cebador emparejado en la pluralidad es específico y complementario a un biomarcador en la Tabla 1, en condiciones que permiten que la pluralidad de sondas de ácido nucleico o cebadores emparejados se hibride con su(s) biomarcador(es) complementario(s) de la Tabla 1. En un ejemplo, el procedimiento también puede incluir, después de poner en contacto la muestra con la pluralidad de sondas de ácido nucleico (como NPP o NPPF + CFS), poner en contacto la muestra con una nucleasa que digiere moléculas de ácido nucleico de una sola cadena. En otros ejemplos, cada uno de los al menos dos biomarcadores en la Tabla 2 se pone en contacto con "sondas en mosaico" que consisten en múltiples sondas (*por ejemplo*, al menos 2, 5 o 10 sondas, como 2, 3, 4, 5, 6, 7, 8, 9 o 10 sondas; *por ejemplo*, NPP o NPPF) específicas para cada biomarcador, cuyo diseño puede ser útil, por ejemplo, para aumentar la señal obtenida de dicho producto génico o para detectar múltiples variantes del mismo producto génico. Por ejemplo, cada sonda en el conjunto de sondas en mosaico puede unirse a una región diferente de una molécula de ácido nucleico diana (o superponerse ligeramente).

En algunos ejemplos, los ácidos nucleicos se detectan mediante la hibridación de ácidos nucleicos. La hibridación de ácido nucleico involucra proporcionar una sonda y un ácido nucleico diana (*por ejemplo*, al menos dos de los de la Tabla 1) en condiciones en las que la sonda y su diana complementaria pueden formar dúplex híbridos estables a través del emparejamiento de bases complementarias. En algunos ejemplos, los ácidos nucleicos que no forman dúplex híbridos se eliminan (*por ejemplo*, lavados, digeridos por nucleasa o eliminados físicamente) dejando que se

60

detecten los ácidos nucleicos hibridados, tal como mediante la detección de un marcador detectable unido (directa o indirectamente). En ejemplos específicos, los ácidos nucleicos que no forman dúplex híbridos, como cualquier sonda en exceso que no se hibrida con su diana respectiva, y las regiones de la secuencia diana que no son complementarias a las sondas, se pueden digerir mediante la adición de nucleasa, dejando solo los dúplex híbridos de la secuencia  
5 diana de la sonda complementaria.

En general, se reconoce que los ácidos nucleicos se desnaturalizan aumentando la temperatura y/o disminuyendo la concentración de sal del tampón que contiene los ácidos nucleicos. En condiciones de baja rigurosidad, se formarán dúplex híbridos (*por ejemplo*, de baja temperatura y/o alta salinidad; *por ejemplo*, ADN:ADN, ARN:ARN o ARN:ADN)  
10 incluso cuando las secuencias recocidas no sean perfectamente complementarias. Por lo tanto, la especificidad de la hibridación se reduce a una menor rigurosidad. Por el contrario, a mayor rigurosidad (*por ejemplo*, una temperatura más alta o salinidad más baja) la hibridación exitosa requiere menos discordancias. Un experto en la materia apreciará que las condiciones de hibridación pueden diseñarse para proporcionar diferentes grados de rigurosidad. La fuerza de la hibridación se puede aumentar sin disminuir la rigurosidad de la hibridación y, por consiguiente, la especificidad de  
15 la hibridación se puede mantener en un tampón de alta rigurosidad, mediante la inclusión de bases no naturales en las sondas, por ejemplo, de ácidos nucleicos bloqueados o peptídicos.

Los cambios en la expresión de un ácido nucleico y/o la presencia de ácido nucleico detectado por estos procedimientos, por ejemplo, pueden incluir aumentos o disminuciones en el nivel (cantidad) o actividad funcional de  
20 tales ácidos nucleicos, su expresión o traducción en proteínas, o en su localización o estabilidad. Un aumento o una disminución, por ejemplo, en relación con un biomarcador de normalización, puede ser, por ejemplo, un cambio de al menos 1, 2 o 5 veces, como aproximadamente 1, 2, 3, 4 o 5 veces (en aumento o disminución) en la expresión y/o la presencia de un ácido nucleico particular, tal como un ácido nucleico correspondiente al biomarcador que se muestra en la Tabla 1.

En un ejemplo, la expresión génica se mide utilizando una metodología multiplexada. En tales procedimientos, una pluralidad de mediciones (*por ejemplo*, las mediciones de expresión génica) se pueden hacer en una sola muestra. Varias tecnologías han evolucionado, las cuales permiten el monitoreo de grandes cantidades de genes en una sola muestra (*por ejemplo*, micromatrices tradicionales, una PCR multiplexada, el análisis en serie de la expresión génica  
30 (SAGE; *por ejemplo*, la Patente de los EE.UU. No. 5.866.330), la amplificación de sonda dependiente de la ligadura múltiple (MLPA), la secuenciación de alto rendimiento, la tecnología basada en microesferas marcadas (*por ejemplo*, las Patentes de los EE.UU. No. 5.736.330 y 6.449.562), la tecnología de código de barras molecular digital (*por ejemplo*, la Patente de los EE.UU. No. 7.473.767).

Las matrices son un conjunto de herramientas para la detección múltiple de la expresión génica. Una matriz es una disposición sistemática de elementos (*por ejemplo*, reactivos de captura de analitos (como sondas de oligonucleótidos específicos de diana, aptámeros o anticuerpos)) donde un conjunto de valores (*por ejemplo*, los valores de expresión génica) pueden asociarse con una clave de identificación. Los elementos agrupados pueden identificarse sistemáticamente en una sola superficie (*por ejemplo*, por mapeo espacial o por etiquetado diferencial), utilizando  
40 superficies identificables por separado (*por ejemplo*, canales de flujo o microesferas) o mediante una combinación de los mismos.

Otros procedimientos útiles involucran una metodología de alto rendimiento, con la cual se pueden consultar varias muestras al mismo tiempo. También se contemplan realizaciones multiplexadas de alto rendimiento (que miden  
45 simultáneamente la expresión de una pluralidad de genes en una pluralidad de muestras). Los ejemplos de procedimientos y sistemas de ensayo que pueden usarse para detectar los biomarcadores descritos son técnicas de ensayo de alto rendimiento descritas en las publicaciones de patentes internacionales No. WO 2003/002750 y WO 2008/121927, WO 1999/032663, WO 2000/079008, WO/2000/037684 y WO 2000/037683 y las Patentes de los EE.UU. No. 6.232.066, 6.458.533, 6.238.869 y 7.659.063.

En algunos procedimientos de matriz, las secuencias de ácido nucleico de interés (como los oligonucleótidos) que están diseñadas para capturar (directa o indirectamente) dos o más productos de los genes que se muestran en la Tabla 1 se colocan en placas o se agrupan en un sustrato de microchip. Por ejemplo, la matriz puede incluir oligonucleótidos que tengan suficientes complementarios para al menos dos de los genes mostrados en la Tabla 2  
55 (como al menos 3, 4, 5, 6, 7, 8, 9, 10 o los 16 genes mostrados en la Tabla 1). En otros ejemplos, la matriz puede incluir oligonucleótidos complementarios a una porción de una sonda de protección de nucleasa (NPP) que tiene una porción que es complementaria a un producto de al menos dos de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8; como al menos 3, 4, 5, 6, 7, 8, 9, 10 o los 16  
60 genes mostrados en la Tabla 1.

Las secuencias ordenadas se hibridan después con ácidos nucleicos, como ADNc o ARN (*por ejemplo*, ARNm, ARNm<sub>i</sub>)



y/o ARNInc), de la muestra de prueba (*por ejemplo*, DLBCL muestra obtenida de un sujeto, cuya caracterización como ABC o GCB es deseada). En un ejemplo, los ácidos nucleicos de la muestra de prueba (que pueden aislarse) están marcados, de modo que se puede determinar su hibridación con el oligonucleótido complementario específico en la matriz. Alternativamente, los ácidos nucleicos de la muestra de prueba no están marcados, y la hibridación entre los  
 5 oligonucleótidos en la matriz y el ácido nucleico diana se detecta usando un ensayo sándwich, por ejemplo, usando oligonucleótidos adicionales complementarios a la diana marcada.

En un ejemplo, los ácidos nucleicos hibridados se detectan al detectar uno o más marcadores unidos a los ácidos nucleicos de muestra o unidos a una sonda de ácido nucleico que se hibrida directa o indirectamente a los ácidos  
 10 nucleicos diana. Los marcadores se pueden incorporar mediante cualquiera de una serie de procedimientos. En un ejemplo, el marcador se incorpora simultáneamente durante la etapa de amplificación en la preparación de los ácidos nucleicos de la muestra. Así, por ejemplo, la reacción en cadena de la polimerasa (PCR) con cebadores marcados o nucleótidos marcados proporcionará un producto de amplificación marcado. En una realización, la amplificación de la transcripción usando un nucleótido marcado (tal como UTP y/o CTP marcado con fluoresceína) incorpora un marcador  
 15 en los ácidos nucleicos transcritos.

Los marcadores detectables adecuados para usar en cualquiera de los procedimientos descritos (por ejemplo, para la detección o secuenciación de ácido nucleico o proteína) incluyen cualquier compuesto o composición que se conjuga directa o indirectamente a otra molécula (como una molécula de ácido nucleico o un nucleótido) para facilitar la  
 20 detección de esa molécula, por ejemplo, que puede detectarse por medios espectroscópicos, fotoquímicos, bioquímicos, inmunquímicos, eléctricos, ópticos o químicos. Los ejemplos no limitantes de marcadores incluyen residuos fluorescentes y fluorogénicos, residuos cromogénicos, haptenos, etiquetas de afinidad, enzimas e isótopos radiactivos. Los marcadores útiles incluyen biotina para la tinción con conjugado de estreptavidina marcado, microesferas magnéticas (por ejemplo, DYNABEADS™), colorantes fluorescentes (por ejemplo, fluoresceína, rojo de  
 25 Texas, rodamina, proteína fluorescente verde y similares), radiomarcadores (por ejemplo, <sup>3</sup>H, <sup>125</sup>I, <sup>35</sup>S, <sup>14</sup>C o <sup>32</sup>P), enzimas (por ejemplo, peroxidasa de rábano picante, fosfatasa alcalina y otras comúnmente utilizadas en un ELISA), y marcadores colorimétricos como oro coloidal o vidrio coloreado, o plástico (por ejemplo, poliestireno, polipropileno, látex, etc.). Las patentes que enseñan el uso de dichos marcadores incluyen la Patente de los EE.UU. No. 3.817.837; la Patente de los EE.UU. No. 3.850.752; la Patente de los EE.UU. No. 3.939.350; la Patente de los EE.UU. No.  
 30 3.996.345; la Patente de los EE.UU. No. 4.277.437; la Patente de los EE.UU. No. 4.275.149; y la Patente de los EE.UU. No. 4.366.241.

Los medios para detectar tales marcadores también son bien conocidos. Así, por ejemplo, los radiomarcadores pueden detectarse usando una película fotográfica o contadores de centelleo, los marcadores fluorescentes pueden detectarse  
 35 usando un fotodetector para detectar la luz emitida. Los marcadores enzimáticos se detectan generalmente proporcionando a la enzima un sustrato y detectando el producto de reacción producido por la acción de la enzima sobre el sustrato, y los marcadores colorimétricos se detectan simplemente visualizando el marcador coloreado.

El marcador puede adicionarse al ácido o ácidos nucleicos diana (muestra) antes o después de la hibridación. Los  
 40 marcadores denominados como "directos" son marcadores detectables que se unen directamente o se incorporan al ácido nucleico diana (muestra) antes de la hibridación. En contraste, los marcadores denominados como "indirectos" se unen al dúplex híbrido después de la hibridación. A menudo, el marcador indirecto se une a un residuo de unión que se ha unido al ácido nucleico diana antes de la hibridación. Así, por ejemplo, el ácido nucleico diana puede biotinilarse antes de la hibridación. Después de la hibridación, un fluoróforo conjugado con avidina se unirá a los dúplex  
 45 híbridos que contienen biotina proporcionando un marcador que se detecta fácilmente (véase, *Laboratory Techniques in Biochemistry and Molecular Biology*, Vol. 24: *Hybridization With Nucleic Acid Probes*, P. Tijssen, ed. Elsevier, Nueva York, 1993).

La hibridación *in situ* (ISH), como la hibridación cromogénica *in situ* (CISH) y la hibridación *in situ* con plata, (SISH) es  
 50 un procedimiento ejemplar para detectar y comparar la expresión de genes de interés (como al menos dos de los de la Tabla 1). La ISH es un tipo de hibridación que utiliza un ácido nucleico complementario para localizar una o más secuencias específicas de ácido nucleico en una porción o sección de tejido (*in situ*), o, si el tejido es lo suficientemente pequeño, en todo el tejido (ISH de montaje completo). La ISH de ARN puede usarse para analizar patrones de expresión en un tejido, como la expresión de los biomarcadores de la Tabla 1. La ISH de ADN (como la CISH y la  
 55 SISH) se puede utilizar para detectar ácidos nucleicos a nivel genómico.

Las células o tejidos de muestra se tratan para aumentar su permeabilidad a fin de permitir que una sonda, como una sonda específica para un biomarcador de la Tabla 1, ingrese a las células. La sonda se adiciona a las células tratadas, se deja hibridar a la temperatura pertinente y el exceso de sonda se lava. Se puede marcar una sonda complementaria  
 60 con un marcador detectable, como una etiqueta radioactiva, fluorescente o antigénica, de modo tal que se pueda determinar la ubicación y la cantidad de la sonda en el tejido, por ejemplo, utilizando una autorradiografía, una microscopía de fluorescencia o un inmunoensayo.

La PCR *in situ* es la amplificación basada en PCR de las secuencias de ácido nucleico diana antes de la ISH. Para la detección de ARN, se introduce una etapa de transcripción inversa intracelular para generar ADN complementario a partir de plantillas de ARN antes de la PCR *in situ*. Esto permite la detección de secuencias de ARN de baja copia.

5 Antes de la PCR *in situ*, las muestras de células o tejidos se fijan y se permeabilizan para preservar la morfología y permitir el acceso de los reactivos de PCR a las secuencias intracelulares a amplificar. La amplificación por PCR de las secuencias diana se realiza después en células intactas en suspensión o directamente en preparaciones de citocentrífugadora o secciones de tejido en portaobjetos de vidrio. En el primer enfoque, las células fijas suspendidas en la mezcla de reacción de PCR se ciclan térmicamente usando ciclos térmicos convencionales. Después de la PCR,  
10 las células se citocentrifugan en portaobjetos de vidrio con visualización de productos de PCR intracelular por ISH o inmunohistoquímica. La PCR *in situ* en portaobjetos de vidrio se realiza superponiendo las muestras con la mezcla de PCR bajo un cubreobjetos que después se sella para evitar la evaporación de la mezcla de reacción. El ciclo térmico se logra colocando los portaobjetos de vidrio directamente sobre el bloque de calentamiento de un ciclador térmico convencional o especialmente diseñado o utilizando hornos de ciclo térmico.

15 La detección de productos de PCR intracelular generalmente se logra mediante una de dos técnicas diferentes, PCR indirecta *in situ* por ISH con sondas específicas de productos de PCR, o PCR directa *in situ* sin ISH a través de la detección directa de nucleótidos marcados (como digoxigenina-11-dUTP, fluoresceína-dUTP, 3H-CTP o biotina-16-dUTP), que se han incorporado a los productos de PCR durante el ciclo térmico.

20

### **3. Amplificación de ácidos nucleicos**

En algunos ejemplos, las moléculas de ácido nucleico diana (como los productos génicos de ácido nucleico (*por ejemplo*, ARNm o ARNInc)) se amplifican como un medio para su detección (o como una etapa de su detección). En  
25 algunos ejemplos, los niveles de expresión de ácidos nucleicos se determinan durante la amplificación, por ejemplo, usando RT-PCR en tiempo real. Por lo tanto, uno o más de los biomarcadores de DLBCL en la Tabla 1 se pueden detectar (por ejemplo, cuantitativamente) en una muestra utilizando los procedimientos de amplificación y los cebadores adecuados.

30 Los ejemplos de los procedimientos de amplificación *in vitro* que pueden usarse incluyen, entre otros, una PCR cuantitativa en tiempo real, una RT-PCR cuantitativa en tiempo real, la amplificación del desplazamiento de cadena; la amplificación isotérmica sin transcripción; la reparación de la amplificación de la reacción en cadena; la amplificación de la reacción en cadena de la ligasa; la amplificación de la reacción de cadena de ligasa de llenado de huecos; la detección de ligasa acoplada y la PCR; y la amplificación sin transcripción de ARN NASBA™. En un ejemplo, se usa  
35 un procedimiento de amplificación basado en la ligadura.

### **4. Secuenciación de ARN**

La secuenciación de ARN puede usarse para obtener la información de expresión génica multiplexada y, en algunas  
40 realizaciones, de alto rendimiento, tal como información con respecto a la expresión de los marcadores en la Tabla 1. Se conocen procedimientos de secuenciación de ARN (*por ejemplo*, véase Chu y Corey, Nuc. Acid Therapeutics, 22:271 (2012)). Las técnicas de secuenciación de transcriptoma completo y secuenciación de ARN dirigida están disponibles y son útiles en los procedimientos descritos. Los procedimientos representativos para el análisis de expresión génica basado en secuenciación incluyen el análisis en serie de expresión génica (SAGE), el análisis de  
45 expresión génica mediante secuenciación de firma masivamente paralela (MPSS), la secuenciación de escopeta de transcriptoma completo (también conocida como WTSS o ARN-Seq) o la secuenciación de protección de nucleasa (también conocida como qNPS o NPSeq; véase la publicación del Tratado de cooperación en materia de patentes (PCT) No. WO 2012/151111; analizado con más detalles a continuación).

### **5. Ensayo cuantitativo de protección de nucleasa (qNPA)**

En algunos ejemplos, las moléculas de ácido nucleico en la muestra cuya expresión se va a medir se detectan en la muestra utilizando un ensayo cuantitativo de protección de nucleasa, por ejemplo, como se describe en las  
55 Publicaciones de patentes internacionales No. WO 99/032663; WO 00/037683; WO 00/037684; WO 00/079008; WO 03/002750 ; WO 08/121927; y WO 2012/151111 y las Patentes de los EE.UU. No. 6.238.869; 6.458.533; 7.659.063 y 8.741.564. Véanse también, Martel y col., Assay and Drug Development Technologies. 2002, 1 (1-1):61-71; Martel y col., Progress in Biomedical Optics and Imaging, 2002, 3:35-43; Martel y col., Gene Cloning and Expression Technologies, Q. Lu y M. Weiner, Eds., Eaton Publishing, Natick (2002); Seligmann Pharmacogenomics, 2003, 3:36-43; Martel y col., "Array Formats" in "Microarray Technologies and Applications," U.R. Muller y D. Nicolau, Eds,  
60 Springer-Verlag, Heidelberg (2005); Sawada y col., Toxicology in Vitro, 20:1506-1513, 2006; Bakir, y col., Bioorg. & Med. Chem Lett, 17:3473-3479, 2007; Kris y col., Plant Physiol. 144:1256-1266, 2007; Roberts y col., Laboratory Investigation, 87:979-997, 2007; Rimsza y col., Blood, 15 de octubre de 2008, 112 (8):3425-3433; Pechhold y col.,

Nature Biotechnology, 27:1038-1042, 2009.

Usando procedimientos qNPA, una sonda de protección de nucleasa (NPP) puede hibridarse con la secuencia diana (como una secuencia en la Tabla 1), que es seguida por la incubación de la muestra con una nucleasa que digiere moléculas de ácido nucleico de una sola cadena. Por lo tanto, si se detecta la NPP, (*por ejemplo*, no es digerida por la nucleasa), entonces la diana de la sonda, por ejemplo, un ácido nucleico diana que se muestra en la Tabla 1, está presente en la muestra, y esta presencia puede cuantificarse. Las NPP pueden diseñarse para dianas individuales y adicionarse a un ensayo como un cóctel para la identificación (*por ejemplo*, en una matriz (denominada qNPA) o mediante secuenciación (denominada qNPS)). Por lo tanto, se pueden medir múltiples dianas de genes dentro del mismo ensayo y/o matriz (*por ejemplo*, mediante el uso de múltiples centrales nucleares).

La NPP es una molécula de ácido nucleico (como un ADN o ARN) que presenta complementariedad suficiente con un ácido nucleico diana (*por ejemplo*, un ARNm) que le permite hibridarse específicamente con el ácido nucleico diana. La NPP protege la molécula de ácido nucleico diana complementaria de la escisión por una nucleasa, como una nucleasa específica para ácidos nucleicos de una sola cadena. En algunos ejemplos, la NPP tiene al menos 35 nucleótidos (como 40 a 80 o 50 a 150 nucleótidos) y comparte al menos el 90, 95 o 100% de complementariedad con la molécula de ácido nucleico diana. La NPP puede incluir bases no naturales. En algunos ejemplos, los procedimientos descritos se usan para detectar (o secuenciar) varias moléculas de ácido nucleico diana diferentes en una muestra (como las que se enumeran en la Tabla 1) usando una pluralidad de NPP (como NPPF, véase más abajo), donde cada NPP (o NPPF) se une específicamente a una molécula de ácido nucleico diana particular.

En un ejemplo específico, la NPP incluye además una o más secuencias flanqueantes en el extremo 5' y/o el extremo 3', y se denomina NPPF (véase el documento de los EE.UU. 8.741.564). Las NPPF incluyen una secuencia que es complementaria a la totalidad o una parte de la molécula de ácido nucleico diana, lo que permite la unión o hibridación específica entre la molécula de ácido nucleico diana y la NPPF. Por ejemplo, la región de la NPPF que es complementaria a una región de la molécula de ácido nucleico diana se une o hibrida a esa región de la molécula de ácido nucleico diana con alta especificidad (y en algunos ejemplos también puede unirse a una región de un enlazador bifuncional). La porción de la NPPF que es complementaria a una región de la molécula de ácido nucleico diana puede tener al menos 6 nucleótidos de longitud, como al menos 10, 25 o 60 nucleótidos, como de 6 a 60 nucleótidos de longitud. Las NPPF incluyen además una o más secuencias flanqueantes en el extremo 5' y/o el extremo 3' de la NPPF. Así, dicha una o más secuencias flanqueantes están ubicadas como 5', 3', o ambas, respecto de la secuencia complementaria a la molécula de ácido nucleico diana. Si la NPPF incluye una secuencia flanqueante tanto en el extremo 5' como en el extremo 3', en algunos ejemplos, la secuencia de cada NPPF es diferente y no se complementarían. La secuencia o secuencias flanqueantes incluyen varios nucleótidos contiguos que tienen una secuencia (como una secuencia de al menos 6, 12 o 25 nucleótidos, como de 12 a 50 nucleótidos) que no se encuentra en una molécula de ácido nucleico presente en la muestra, y proporcionan una secuencia de hibridación y/o amplificación universal. Esta secuencia de hibridación y/o amplificación universal, cuando tiene una secuencia complementaria a al menos una parte de un cebador de amplificación, permite la multiplexación, ya que los mismos cebadores de amplificación pueden usarse para amplificar NPPF específicas para diferentes moléculas de ácido nucleico diana. También proporciona una secuencia de hibridación universal para todas las NPPF, que se puede usar para adicionar un marcador detectable para la NPPF o para capturar y concentrar la NPPF. Por ejemplo, si la misma secuencia flanqueante está presente en NPPF específicas para diferentes moléculas de ácido nucleico diana, se puede usar el mismo cebador para amplificar cualquier NPPF que tenga la misma secuencia flanqueante, incluso si la NPPF se dirige a una molécula de ácido nucleico diferente. Por ejemplo, la secuencia flanqueante se puede usar para capturar las NPPF, por ejemplo, en una superficie. La secuencia flanqueante puede contener una secuencia variable, como una secuencia que es específica para cada NPPF específica y puede usarse para capturar esa NPPF en una superficie o para otros fines, como para identificar la NPPF. En algunos ejemplos, la NPPF tiene al menos 35 nucleótidos, como de 40 a 80 o de 50 a 150 nucleótidos. En algunos ejemplos, la NPPF incluye dos secuencias flanqueantes: una en el extremo 5' y la otra en el extremo 3'. En algunos ejemplos, la secuencia flanqueante en el extremo 5' difiere de la secuencia flanqueante en el extremo 3'. Además, si la NPPF incluye dos secuencias flanqueantes, idealmente las dos secuencias flanqueantes tienen una temperatura de fusión ( $T_m$ ) similar, como una  $T_m$  de +/- 5°C.

En algunos ejemplos, la NPP o NPPF se hibrida específicamente y/o es complementaria a toda o parte de la secuencia diana. Por lo tanto, las NPP y NPPF pueden incluir al menos 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 o más nucleótidos consecutivos complementarios a una molécula de ácido nucleico diana. En algunos ejemplos, las NPP y NPPF no tienen más de 500 nucleótidos, como no más de 400, 300, 250, 200, 100, 50 o incluso no más de 25 nucleótidos consecutivos complementarios a una molécula de ácido nucleico diana (como de 10 a 500, de 10 a 400, de 10 a 250, de 10 a 200, de 10 a 100, de 25 a 75, de 10 a 60, de 40 a 80, de 100 a 200 o de 10 a 50 nucleótidos consecutivos complementarios a una molécula de ácido nucleico diana. En ejemplos particulares, las NPP y NPPF tienen al menos 10 nucleótidos de longitud, como al menos 10 nucleótidos contiguos complementarios a una secuencia de ácido nucleico diana o una porción de la misma, como las secuencias diana descritas en esta invención (por

- ejemplo, las de la Tabla 1). Las longitudes particulares de NPP y NPPF que se pueden usar para practicar los procedimientos de la presente descripción incluyen sondas que tienen al menos 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75 o más nucleótidos contiguos complementarios a una molécula de ácido nucleico diana o una porción de la misma. En algunos ejemplos, las NPP y NPPF comprenden (o las NPP consisten en) una secuencia de ácido nucleico con al menos un 95% de identidad de secuencia (como al menos un 95, 96, 97, 98, 99 o incluso un 100% de identidad de secuencia) al respecto del complemento inverso de una secuencia diana. En algunos ejemplos, las NPP y NPPF pueden incluir 1, 2, 3 o más discordancias en comparación con el complemento inverso de la molécula de ácido nucleico diana.
- En algunos ejemplos, las NPP y NPPF pueden incluir una o más (como 1, 2, 3 o más) bases sintéticas o bases alternativas (como la inosina). En otros ejemplos, las NPP pueden incluir uno o más nucleótidos modificados o análogos de ácidos nucleicos, como uno o más ácidos nucleicos bloqueados (véase, *por ejemplo*, la Patente de los EE.UU. No. 6.794.499) o uno o más ácidos nucleicos peptídicos. Los nucleótidos modificados, los no naturales, los sintéticos o alternativos se pueden usar en una NPP o NPPF en una o más posiciones (como 1, 2, 3, 4, 5 o más posiciones). Las NPP y NPPF también pueden degenerarse en una o más posiciones (como 1, 2, 3, 4, 5 o más posiciones), por ejemplo, una NPP o NPPF que incluye una mezcla de nucleótidos (como 2, 3, o 4 nucleótidos) en una posición especificada en la NPP o NPPF.
- En algunos ejemplos, una muestra de un sujeto (como una muestra que incluye ácidos nucleicos, por ejemplo, ARN) puede ponerse en contacto con una pluralidad de NPP o NPPF, donde al menos algunas de las NPP y NPPF se unen específicamente a al menos dos moléculas de ácido nucleico diana diferentes (por ejemplo, las que se muestran en la Tabla 1). En algunos ejemplos, las NPP o NPPF se unen específicamente a al menos dos de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS (como 1, 2, 3, 4, 5, 6, 7, 8, 9 o todos estos 10); como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS (como 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o todos estos 16). En algunos ejemplos, una muestra de un sujeto (como una muestra que incluye ácidos nucleicos, como los ARN) también se pone en contacto con una o más NPP o NPPF específicas para uno o más genes de mantenimiento/normalización (*por ejemplo*, una o más NPP o NPPF específicas para ANT).
- En algunos ejemplos, la pluralidad de NPP o NPPF incluye más de una (como 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o 16 o más) NPP o NPPF diferentes, cada una específica para un solo ARNm diana o una región única de un ARNm diana (por ejemplo, véase la parte específica de la diana de las secuencias de NPPF en la Tabla 2). La pluralidad de NPP o NPPF se incuba con la muestra en condiciones suficientes para que las NPP o NPPF se hibriden específicamente con su respectivo ARNm diana. Si se utilizan NPPF, el procedimiento incluye poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la o las secuencias flanqueantes (CFS) en condiciones suficientes para que la o las secuencias flanqueantes de la NPPF se unan específicamente a la CFS; La muestra se pone en contacto con una nucleasa específica para ácidos nucleicos de una sola cadena (por ejemplo, nucleasa S1), y se detecta (o secuencia) la presencia de cada NPP o NPPF. En algunos ejemplos, las NPP o NPPF se amplifican antes de su detección (o secuenciación). Los ARNm se identifican como presentes en la muestra y/o se cuantifican cuando se detectan sus respectivas NPP o NPPF.

#### a. Tratamiento de la muestra

- En algunos ejemplos, las muestras (*por ejemplo*, las células o el tejido) del sujeto primero se lisan o se permeabilizan en una solución acuosa (por ejemplo, usando un tampón de lisis). Las células se incuban en la solución acuosa durante un período de tiempo suficiente (como aproximadamente 1 a alrededor de 60 minutos, por ejemplo, aproximadamente 5 a alrededor de 20 minutos o aproximadamente 10 minutos) y a una temperatura suficiente (como aproximadamente de 22 a alrededor de 115°C, por ejemplo, aproximadamente de 37 a alrededor de 105°C o aproximadamente de 90°C a alrededor de 110°C) para lisar o permeabilizar la célula. En algunos ejemplos, la lisis se realiza a aproximadamente 95°C. En algunos ejemplos, la etapa de lisis incluye incubar la muestra a aproximadamente 95°C durante aproximadamente 5 a 15 minutos para desnaturalizar el ARN en la muestra, pero no el ADN genómico. En otros ejemplos, la etapa de lisis incluye incubar la muestra a aproximadamente 105°C durante aproximadamente 5 a 15 minutos para desnaturalizar los ácidos nucleicos en la muestra. En un ejemplo, la Proteínasa K se incluye con el tampón de lisis.

- En algunos ejemplos, una o más NPP o NPPF (como las que incluyen o tienen una secuencia mostrada en las SEQ ID NO: 1-16) se puede adicionar a una muestra complementaria a una o más dianas, en una concentración que varía de aproximadamente de 10 pM a aproximadamente 10 nM (como aproximadamente de 30 pM a 5 nM, aproximadamente de 100 pM a alrededor de 1 nM), en un tampón como uno que contiene NaCl, KCl, H<sub>2</sub>PO<sub>4</sub>, EDTA, 0,05% Triton X-100, o combinaciones de los mismos (por ejemplo, 6X SSPE-T: NaCl 0,9 M, 60 mM NaH<sub>2</sub>PO<sub>4</sub>, EDTA 6 mM y Triton X-100 al 0,05%) o un tampón de lisis. En un ejemplo, cada NPP o NPPF se adiciona a la muestra en

una concentración final de al menos 10 pM, tal como al menos 20, 30, 50, 100, 150, 200, 500 pM, al menos 1 nM o 10 nM. En un ejemplo, cada NPP o NPPF se adiciona a la muestra en una concentración final de aproximadamente 30 pM. En otro ejemplo, cada NPP o NPPF se adiciona a la muestra en una concentración final de aproximadamente 167 pM. En incluso otro ejemplo, cada NPP o NPPF se adiciona a la muestra en una concentración final de aproximadamente 1 nM. En un ejemplo, si se usan NPPF, cada CFS se adiciona a la muestra en una concentración final de aproximadamente al menos 6 veces la cantidad de sonda, como al menos 10 o 20 veces la cantidad de sonda (por ejemplo, de 6 a 20 veces la cantidad de sonda). En un ejemplo, cada CFS se adiciona en al menos 1, 5, 10, 50, 100 o 200 nM, por ejemplo, de 1 a 100, de 5 a 100 o de 5 a 50 nM. Por ejemplo, si hay seis sondas, cada una a 166 pM, cada CFS se puede adicionar de 5 a 50 nM. En tales ejemplos, con las NPP o NPPF no digeridas por una nucleasa, como S1, si la NPP o NPPF se hibrida (forma un dúplex con) una secuencia complementaria, como una secuencia diana.

#### b. Condiciones ejemplares de hibridación

15 Un experto en la materia puede identificar condiciones suficientes para que una NPP o NPPF (+ CFS) se hibride específicamente con su diana presente en la muestra de prueba. Por ejemplo, un experto en la materia puede determinar experimentalmente las características (tales como longitud, composición de base y grado de complementariedad) que permitirán que un ácido nucleico (*por ejemplo*, una sonda de fusión) se hibride con otro ácido nucleico (*por ejemplo*, un ácido nucleico diana en cualquiera de las Tablas 2-8) en condiciones de rigurosidad seleccionada, mientras se minimiza la hibridación no específica a otras sustancias o moléculas. Generalmente, la secuencia de ácido nucleico de una NPP tendrá suficiente complementariedad con la secuencia diana correspondiente para permitirle hibridarse bajo condiciones de hibridación rigurosas seleccionadas, por ejemplo, hibridación a aproximadamente 37°C o más (tal como aproximadamente 37, 42, 50, 55, 60, 65, 70, 75°C o más). Entre los parámetros de reacción de hibridación que se pueden variar están la concentración de sal, el tampón, el pH, la temperatura, el tiempo de incubación, la cantidad y el tipo de desnaturizante, como la formamida.

Las moléculas de ácido nucleico en la muestra se desnaturalizan (por ejemplo, a aproximadamente 95 a alrededor de 105°C durante aproximadamente 5-15 minutos) y se hibridan con la pluralidad de NPP o NPPF (+CFS) durante entre aproximadamente 10 y alrededor de 72 horas (por ejemplo, entre aproximadamente 10 minutos y alrededor de 24 horas, como al menos entre aproximadamente 1 y 20 horas, o aproximadamente de 6 a 16 horas) a una temperatura que varía de aproximadamente 4 a alrededor de 70°C (por ejemplo, aproximadamente de 37 a alrededor de 65°C, aproximadamente de 42 a alrededor de 60°C o aproximadamente de 50 a alrededor de 60°C). En algunos ejemplos, la pluralidad de NPP o NPPF (+CFS) se incuba con la muestra a una temperatura de al menos aproximadamente 37, 40, 45, 50, 55, 60, 65 o 70°C. En un ejemplo, la pluralidad de NPP o NPPF (+CFS) se incuba con la muestra a aproximadamente 37°C. En otro ejemplo, la pluralidad de NPP o NPPF (+CFS) se incuba con la muestra a aproximadamente 42°C. En otro ejemplo, la pluralidad de NPP o NPPF (+CFS) se incuba con la muestra a aproximadamente 50°C. Estas temperaturas de hibridación son ejemplares, y un experto en la materia puede seleccionar la temperatura de hibridación adecuada dependiendo de factores tales como la longitud y la composición de nucleótidos de las NPP o NPPF (+CFS).

En algunos ejemplos, los procedimientos no incluyen la purificación de ácido nucleico (por ejemplo, la purificación de ácido nucleico no se realiza antes de poner en contacto la muestra con las NPP o NPPF (+CFS) y/o la purificación de ácido nucleico no se realiza después de contactar la muestra con las NPP o NPPF (+CFS)). En algunos ejemplos, los procedimientos no incluyen la amplificación de ácido nucleico (por ejemplo, la amplificación de ácido nucleico no se realiza antes de poner en contacto la muestra con las NPP o NPPF (+CFS) y/o la amplificación de ácido nucleico no se realiza después de contactar la muestra con las NPP o NPPF (+CFS)). En algunos ejemplos, no se requiere preprocesamiento de la muestra, excepto la lisis celular. En algunos ejemplos, la lisis celular y el contacto de la muestra con la pluralidad de NPP o NPPF (+CFS) se producen secuencialmente. En otros ejemplos, la lisis celular y el contacto de la muestra con la pluralidad de NPP o NPPF (+CFS) se producen simultáneamente, en algunos ejemplos no limitantes sin ninguna etapa intermedia.

#### c. Tratamiento con nucleasa

Después de la hibridación de una o más NPP o NPPF (+CFS) y ácidos nucleicos en la muestra, la muestra se somete a un procedimiento de protección de nucleasa. Las NPP o NPPF que se han hibridado con un ácido nucleico diana no son hidrolizadas por la nucleasa y pueden detectarse posteriormente.

El tratamiento con una o más nucleasas destruirá todas las moléculas de ácido nucleico ss (incluidos el ARN y el ADN en la muestra que no está hibridada (por lo tanto, no está protegida por) con las NPP o NPPF, las NPP o NPPF que no están hibridadas con el ácido nucleico diana y ( cuando se usan) las CFS no hibridadas con una NPPF), pero no destruirá las moléculas de ácido nucleico ds, como las NPPF que se han hibridado con la CFS y una molécula de ácido nucleico diana presente en la muestra y los NPP que se han hibridado con una molécula de ácido nucleico diana

presente en la muestra. Por ejemplo, si la muestra incluye un extracto celular o lisado, los ácidos nucleicos no deseados, como el ADN genómico no diana, el ARNt, el ARNr, el ARNm, el ARNm<sub>i</sub> y las porciones de la(s) molécula(s) de ácido nucleico diana que no se hibridan con secuencias de NPP o NPPF complementarias (como los excedentes), que en el caso de las dianas de ácido nucleico de ARNm o ADN constituirán la mayoría de la secuencia diana nucleica, pueden destruirse sustancialmente en esta etapa. Esto deja una cantidad estequiométrica de dúplex NPP/ácido nucleico diana o dúplex NPPF/CFS/ácido nucleico diana. Si la molécula diana se reticula al tejido que se produce por la fijación, las NPP o NPPF se hibridan con la molécula diana reticulada sin la necesidad de revertir la reticulación, o liberar el ácido nucleico diana del tejido al que está reticulado

- 10 Se puede usar cualquiera de una variedad de nucleasas, incluyendo el ARNsa pancreático, la nucleasa de frijol mungo, la nucleasa S1, el ARNse A, la ribonucleasa T1, la exonucleasa III, la exonucleasa VII, el ARNse CLB, el ARNse PhyM, el ARNse U2 o similares, dependiendo de naturaleza de los complejos hibridados y de los ácidos nucleicos indeseables presentes en la muestra. En un ejemplo particular, la nucleasa es específica para ácidos nucleicos de una sola cadena, por ejemplo, la nucleasa S1. Una ventaja de usar una nucleasa específica para ácidos nucleicos de una sola cadena en algunas formas de realización del procedimiento descritas en esta invención es eliminar tales moléculas de una sola cadena ("adhesivas") de las etapas de reacción posteriores en las que pueden conducir a un fondo no deseable o reactividad cruzada. La nucleasa S1 está disponible comercialmente de, por ejemplo, Promega, Madison, WI (catálogo No. M5761); Life Technologies/Invitrogen, Carlsbad, CA (catálogo No. 18001-016); Fermentas, Glen Burnie, MD (catálogo No. EN0321) y otros. Las condiciones de reacción para estas enzimas son bien conocidas en la técnica y pueden optimizarse empíricamente.

En algunos ejemplos, la nucleasa S1 se diluye en un tampón (como uno que contiene acetato de sodio NaCl, KCl, ZnSO<sub>4</sub>, KATHON, o combinaciones de los mismos) se adiciona a la mezcla hibridada de sonda/muestra y se incuba a aproximadamente de 37 a alrededor de 60°C (tal como aproximadamente a 50°C) durante 10 a 120 minutos (por ejemplo, de 10 a 30 minutos, de 30 a 60 minutos, de 60 a 90 minutos o 120 minutos) para digerir el ácido nucleico no hibridado de la muestra y las NPP o NPPF no hibridadas.

Las muestras pueden tratarse opcionalmente para eliminar de otro modo el material no hibridado y/o para inactivar o eliminar enzimas residuales (*por ejemplo*, por calentamiento, extracción con fenol, precipitación, filtración en columna, adición de proteinasa k, adición de un inhibidor de nucleasa, quelación de cationes divalentes requeridos por la nucleasa para la actividad o combinaciones de los mismos). En algunos ejemplos, las muestras se tratan opcionalmente para disociar el ácido nucleico diana y las CFS de su NPPF complementaria o el ácido nucleico diana de su NPP complementaria (*por ejemplo*, utilizando una hidrólisis de base y calor). En algunos ejemplos, después de la hibridación y el tratamiento con nucleasa, una molécula de ácido nucleico diana hibridada con la NPP o NPPF puede degradarse, *por ejemplo*, disociando el dúplex con la NPP o NPPF en la base y después destruyendo el ácido nucleico mediante nucleasas o tratamientos químicos/físicos, como la hidrólisis de la base a temperatura elevada, dejando la NPP o NPPF en proporción directa a cuánto se había hibridado con el ácido nucleico diana. Alternativamente, la muestra puede tratarse para dejar la porción hibridada (una sola cadena) del ácido nucleico diana, o el dúplex formado por el ácido nucleico diana hibridado y la sonda, para ser analizados adicionalmente.

En algunos ejemplos después de la incubación con una nucleasa, se adiciona una base (como NaOH o KOH) para aumentar el pH a aproximadamente de 9 a 12 y la muestra se calienta (por ejemplo, a 95°C durante 10 minutos). Esto disocia los dímeros, dejando la NPP o NPPF en un estado de una sola cadena y, en el caso del ARN, hidroliza las moléculas diana de ARN. Esta etapa también puede neutralizar o desactivar la nucleasa, por ejemplo, elevando el pH por encima de aproximadamente 6.

En algunos ejemplos, la muestra se trata para ajustar el pH a aproximadamente de 7 a alrededor de 8, por ejemplo, mediante la adición de ácido (por ejemplo, HCl). En algunos ejemplos, el pH se eleva a aproximadamente de 7 a alrededor de 8 en un tampón Tris. Elevar el pH puede evitar la depuración del ADN y también evita que muchas nucleasas específicas de ss (*por ejemplo*, S1) funcionen por completo.

En algunos ejemplos, la muestra se purifica o se separa para eliminar el ácido nucleico no deseado u otras moléculas, por ejemplo, mediante la purificación en gel u otro procedimiento de separación.

#### 55 d. Amplificación opcional

En algunos ejemplos, las NPP o NPPF se amplifican antes de su detección. Por ejemplo, las NPP, las NPPF o las moléculas de ácido nucleico diana resultantes que se han separado de la NPP o NPPF, se pueden amplificar, por ejemplo, utilizando procedimientos de rutina, como la PCR, u otras formas de amplificación enzimática o procedimientos de amplificación basados en la ligadura.

Los ejemplos de los procedimientos de amplificación *in vitro* que pueden usarse incluyen, entre otros, una PCR

- cuantitativa en tiempo real, la amplificación del desplazamiento de cadena; la amplificación isotérmica sin transcripción; la reparación de la amplificación de la reacción en cadena; la amplificación de la reacción en cadena de la ligasa; la amplificación de la reacción de cadena de ligasa de llenado de huecos; la detección de ligasa acoplada y la PCR; y la amplificación sin transcripción de ARN NASBA™. En un ejemplo, se usa un procedimiento de amplificación basado en la ligadura, donde los cebadores son específicos de la NPPF y se juntan para que puedan unirse y fundirse, y después los cebadores nuevos se ligan durante una serie de ciclos. La ligadura puede ser enzimática o no enzimática. Si las secuencias flanqueantes de NPPF se usan para la hibridación de los cebadores, la amplificación puede ser universal.
- 10 Durante la amplificación, se puede incorporar una etiqueta de experimento y/o un adaptador de secuenciación como, por ejemplo, parte de las construcciones de cebador y extensión, por ejemplo, en el extremo 3' o 5' o en ambos extremos de la NPP o NPPF. En algún ejemplo, tales etiquetas, adaptadores se adicionan antes de la amplificación. La amplificación también se puede utilizar para introducir un marcador detectable en los amplicones de NPP o NPPF generados (por ejemplo, si la NPP o NPPF originalmente no estaba marcada o si se desea un marcaje adicional) u
- 15 otra molécula que permita la detección o la extinción. Por ejemplo, el cebador de amplificación puede incluir un marcador detectable, haptán o desactivador que se incorpora a la NPP o NPPF durante la amplificación. Dicho marcador, haptán o desactivador se puede introducir en cualquier extremo del amplicón de la NPP o NPPF (o en ambos extremos), o en cualquier punto intermedio.
- 20 En algunos ejemplos, los amplicones de NPP o NPPF resultantes se limpian antes de la detección o secuenciación. Por ejemplo, la mezcla de reacción de amplificación se puede limpiar usando procedimientos bien conocidos en la técnica (*por ejemplo*, la purificación en gel, la captura y liberación de biotina/avidina y la electroforesis capilar). En un ejemplo, los amplicones de NPP o NPPF se biotinilan (o incluyen otro haptano) y se capturan en una superficie o microesfera recubierta con avidina o anti-haptán, se lavan y después se liberan para su detección o secuenciación.
- 25 Los productos amplificados también se pueden limpiar después de la última etapa de amplificación, mientras que todavía son bicatenarios, mediante un procedimiento que utiliza una nucleasa que hidroliza oligonucleótidos de una sola cadena (como la Exonucleasa I), que a su vez puede inactivarse antes de continuar a la siguiente etapa, como la hibridación con una superficie.
- 30 e. Detección de la NPP o NPPF utilizando una etiqueta**
- Se puede detectar la presencia de NPP o NPPF (o la diana restante, complejo diana:NPP o complejo diana:NPPF:CFS), los amplicones de NPP o los amplicones de NPPF. Se puede usar cualquier procedimiento adecuado para detectar las sondas o amplicones de NPP o NPPF (o la diana restante, complejo diana:NPP o complejo
- 35 diana:NPPF:CFS). En algunos ejemplos, las NPP o NPPF (o sus amplicones) incluyen un marcador detectable y la detección de la presencia de las NPP o NPPF (o sus amplicones) incluye la detección del marcador detectable. En algunos ejemplos, las NPP o NPPF (o sus amplicones) están marcadas con el mismo marcador detectable. En otros ejemplos, las NPP o NPPF (o sus amplicones) están marcadas con diferentes marcadores detectables (como un marcador diferente para cada diana). En otros ejemplos, las NPP o NPPF (o sus amplicones) se detectan
- 40 indirectamente, por ejemplo, por hibridación con un ácido nucleico marcado.
- En un ejemplo específico, no limitativo, las NPP o NPPF (o sus amplicones) incluyen un marcador de biotina. En este ejemplo, las NPP o NPPF (o sus amplicones) pueden detectarse incubando (como en un soporte, *por ejemplo*, una matriz o una microesfera, que contiene las NPP o NPPF (o sus amplicones)) con avidina-HRP o estreptavidina-HRP,
- 45 o un conjugado con otra enzima adecuada como la fosfatasa alcalina, y después poniendo el soporte en contacto con quimioluminiscencia cromogénica o un sustrato generador de fluorescencia. En un ejemplo no limitante, el sustrato es TMA-3 (Lumigen, Southfield, MI). Hay sustratos quimioluminiscentes adicionales disponibles comercialmente, como LumiGlo® (KPL, Gaithersburg, MD), SuperSignal® (Pierce, Rockford, IL) y FCL™ (Amersham/GE Healthcare, Piscataway, NJ). La señal producida por el sustrato se detecta, por ejemplo, utilizando un generador de imágenes de micromatrices (como un escáner de imágenes OMIX, OMIX HD, CAPELLA o SUPERCAPELLA, HTG Molecular
- 50 Diagnostics, Tucson, Arizona), o visualmente, como en un dispositivo de flujo lateral. Se puede utilizar la luminiscencia basada en europio, así como también la electroluminiscencia o la dispersión de la luz, o la luminiscencia eléctrica (por ejemplo, la conductividad o resistencia).
- 55 En otro ejemplo, las NPP o NPPF (o sus amplicones) incluyen un marcador fluorescente, como Cy-3 o Cy-5. Las NPP o NPPF (o sus amplicones) se pueden detectar utilizando un generador de imágenes de micromatrices estándar (como un generador de imágenes Typhoon™ (GE Life Sciences, Piscataway, NJ), un escáner de micromatrices GenePix® (Molecular Devices, Sunnyvale, CA), un escáner GeneChip® (Affymetrix, Santa Clara, CA), procedimientos de citometría de flujo o procedimientos de microscopía fluorescente. Una persona con conocimiento ordinario en la técnica
- 60 puede seleccionar procedimientos de detección y reactivos adecuados para estos u otros marcadores detectables.

**f. Detección de NPP o NPPF utilizando moléculas de captura**

En algunos ejemplos, después del tratamiento de hibridación y nucleasa (y amplificación opcional), la muestra se pone en contacto con una superficie que incluye múltiples regiones espacialmente discretas, cada una de las cuales incluye una molécula de captura o se contacta con una pluralidad de superficies, cada una de las cuales incluye una molécula de captura. Por ejemplo, la superficie puede ser una población de microesferas, donde las subpoblaciones de microesferas incluyen al menos una molécula de captura. Por ejemplo, una primera subpoblación podría incluir al menos una molécula de captura, mientras que una segunda subpoblación podría incluir al menos una molécula de captura que tenga una secuencia diferente a la primera y así sucesivamente. En algunos ejemplos, la molécula de captura incluye al menos un anclaje asociado con un enlazador bifuncional (también denominado "enlazador de programación"). Alternativamente, la molécula de captura incluye una sonda de captura de ácido nucleico, que tiene una secuencia que es complementaria al menos a una porción de NPP o NPPF (o sus amplicones), como complementaria a toda o una porción de una región flanqueante de una NPPF o su amplicón.

En un ejemplo donde la molécula de captura incluye al menos un anclaje asociado con un enlazador bifuncional, el anclaje y el enlazador bifuncional están asociados por la hibridación, el recocido, el enlace covalente u otra unión. El conector bifuncional incluye una primera porción que se une específicamente (por ejemplo, es complementaria a) el anclaje y una segunda porción que se une específicamente (por ejemplo, es complementaria a) una de la pluralidad de NPP o NPPF (o su amplicón).

En algunos ejemplos, los procedimientos descritos incluyen un anclaje en una superficie (por ejemplo, en una matriz), que está asociada con un conector bifuncional que se utiliza para capturar las NPP o NPPF (o sus amplicones) después de la etapa de amplificación o nucleasa. En algunos ejemplos, un anclaje es un oligonucleótido de aproximadamente 8 a 150 nucleótidos de longitud (por ejemplo, de aproximadamente 8 a 100, 15 a 100, 20 a 80, 25 a 75 o 25 a 50, como de aproximadamente 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140 o 150 nucleótidos). En un ejemplo no limitante, el anclaje tiene aproximadamente 25 nucleótidos de longitud. En algunos ejemplos, el anclaje incluye una primera porción que se une específicamente a la primera porción del conector bifuncional y una segunda porción que actúa como un separador entre la superficie y la primera porción del anclaje. En algunos ejemplos, la segunda porción del anclaje tiene aproximadamente de 6 a 60 átomos de carbono o nucleótidos de longitud (como aproximadamente 6, 12, 24, 30, 36, 42, 48, 54 o 60 átomos de carbono o nucleótidos). En otros ejemplos, la segunda porción del anclaje tiene aproximadamente de 5 a 100 átomos de carbono o nucleótidos de longitud (como aproximadamente 10 a 50, 15 a 40, 20 a 30 o alrededor de 25 átomos de carbono o nucleótidos).

La composición de base para los anclajes para los procedimientos descritos es tal que la estabilidad termodinámica del anclaje y el emparejamiento del enlazador bifuncional es alta. En algunos ejemplos, el porcentaje de composición base para los anclajes es aproximadamente del 30 al 40% G, del 30 al 40% C, del 10 al 20% A y del 10 al 20% T. En algunos ejemplos, la frecuencia de vecino más cercano en los anclajes minimiza los vecinos más cercanos de G-G o C-C para reducir las reacciones secundarias mediadas a través de la formación del cuarteto G. En otros ejemplos, se pueden incorporar bases no naturales, o ácidos nucleicos peptídicos, en el anclaje o en el conector bifuncional para modificar sus propiedades.

Los procedimientos para diseñar y sintetizar anclas de uso en los procedimientos descritos se describen, por ejemplo, en la Publicación del PCT No. WO 98/24098. En algunos ejemplos, se desea un conjunto de anclajes que sean sustancialmente diferentes entre sí. Un algoritmo ejemplar para obtener un conjunto de anclajes diferentes es el siguiente:

1) Se define el tamaño del conjunto. En algunas realizaciones, 16, 24, 36, 48, 49, 64, 81, 96 y 100 constituyen tamaños útiles.

2) Se define la estructura de secuencia general del conjunto de anclaje. La longitud y la composición base como se describió anteriormente se usan para definir tales parámetros. En general, el número de bases G y bases C se mantienen iguales, tal como sucede en el caso del número de bases A y bases T. Esta igualdad optimiza la diversidad configuracional de los conjuntos finales. Por lo tanto, tales conjuntos se describirán mediante la ecuación  $G_n C_n A_m T_m$ .

3) Para una estructura de conjunto definida por m y n, se emplea un generador de números aleatorios, a fin de producir un conjunto de isómeros de secuencia aleatoria.

4) Se selecciona un miembro del conjunto de secuencias aleatorias para utilizarlo como elemento #1 del conjunto.

5) Se define la similitud máxima permitida entre los miembros del conjunto. La similitud se define en términos de comparación de base local por pares. Por ejemplo, cuando dos cadenas de oligómero de idéntica longitud n están alineadas de modo que los extremos 5' y 3' están en registro, la falta de discordancias se refiere a la situación donde en todas las posiciones 1-n, las bases en las dos cadenas son idénticas. La discordancia completa se refiere a la



situación donde, en todas las posiciones 1-n, las bases en las dos cadenas son diferentes. Por ejemplo, una similitud máxima útil podría ser 10 o más discordancias dentro de un conjunto de 16, sondas de captura de 16mer.

6) Se selecciona un segundo miembro del conjunto de secuencias aleatorias y se determina su similitud con el elemento #1. Si el elemento #2 posee menos de la similitud máxima permitida con el elemento #1, se mantendrá en el conjunto. Si el elemento #2 posee una similitud mayor que la máxima permitida, se descarta y se elige una nueva secuencia para la comparación. Este proceso se repite hasta que se haya determinado un segundo elemento.

7) De manera secuencial, se eligen miembros adicionales del conjunto de secuencias aleatorias que satisfacen las restricciones de disimilitud con respecto a todos los elementos previamente seleccionados.

En otros ejemplos en los que la molécula de captura incluye al menos una sonda de captura de ácido nucleico, que tiene una secuencia que es complementaria a al menos una porción de una NPP o NPPF (o su amplicón), tal como complementaria a toda o una porción de una región flanqueante de un amplicón NPPF. Por ejemplo, la sonda de captura de ácido nucleico puede incluir una región que es complementaria a la NPP o NPPF (o su amplicón), y puede incluir una región que no lo es (como una región que permite la unión de la sonda a una superficie). La sonda de captura de ácido nucleico se puede unir directamente a una superficie. Por ejemplo, la sonda de captura de ácido nucleico puede incluir una amina para la unión covalente a una superficie. En algunos ejemplos, una sonda de captura de ácido nucleico es un oligonucleótido de al menos 8 nucleótidos de longitud, como de al menos 10, 15, 20, 30, 50 o 100 nucleótidos de longitud (por ejemplo, de aproximadamente 8 a 100, 15 a 100, 20 a 80, 25 a 75 o 25 a 50, como de aproximadamente 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140 o 150 nucleótidos). Un experto en la materia apreciará que la región de la sonda de captura de ácido nucleico complementaria a una región de NPP o NPPF (o su amplicón) no necesita ser 100% complementaria, siempre que pueda producirse hibridación entre la sonda de captura de ácido nucleico y la NPP o NPPF adecuada (o su amplicón). En algunos ejemplos, la región de la sonda de captura de ácido nucleico complementaria a una región de la NPP o NPPF (o su amplicón) es de al menos 8 nucleótidos de longitud, como de al menos 8, 10, 15, 20, 30, 50 o 100 nucleótidos de longitud (por ejemplo, de aproximadamente 8 a 100, 15 a 100, 20 a 80, 25 a 75 o 25 a 50, como de aproximadamente 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140 o 150 nucleótidos de longitud).

En algunos ejemplos, la muestra que contiene NPP o NPP (o sus amplicones) se desnaturaliza antes de entrar en contacto con la superficie de la matriz (por ejemplo, calentando la muestra a 95°C durante 5 minutos y enfriándola rápidamente en hielo). En algunos ejemplos, la muestra que contiene NPP o NPPF (o sus amplicones) se ajusta antes de entrar en contacto con la superficie (por ejemplo, para ajustar la concentración de sal o formamida). La muestra que contiene NPP o NPPF (o sus amplicones) se incuba con la superficie (por ejemplo, una matriz o microesferas) durante un período de tiempo suficiente para que las NPP o NPPF (o sus amplicones) se unan específicamente (por ejemplo, se hibriden) con la molécula de captura. En algunos ejemplos, la incubación de la muestra con la superficie a aproximadamente de 37 a alrededor de 65°C (por ejemplo, aproximadamente de 45 a alrededor de 60°C o aproximadamente de 50 a alrededor de 60°C, como a 50°C) durante al menos 1 hora (por ejemplo, de 1 a 8 horas, de 1 a 36 horas, de 12 a 24 horas, de 16 a 24 horas o durante la noche) para permitir que las NPP o NPPF (o sus amplicones) se hibriden con la molécula de captura. El tiempo de captura puede acortarse, por ejemplo, si se usan dispositivos microfluídicos o macrofluídicos, dispositivos de flujo lateral, o al reducir la difusión y usar mezcla o flujo activo.

Algunas de las superficies (o sustratos) que se pueden usar en los procedimientos descritos se encuentran inmediatamente disponibles de manos de proveedores comerciales. En algunas realizaciones, la superficie es una placa de microtitulación de 96, 384 o 1536 pocillos, tal como las placas modificadas que vende Corning Costar. En otras realizaciones, un sustrato incluye una o más microesferas (como una población de microesferas que se pueden diferenciar por tamaño o color, por ejemplo, por citometría de flujo). Alternativamente, se puede formar una superficie que comprende pocillos que, a su vez, comprenden hendiduras u "hoyuelos" micromaquinando una sustancia como aluminio o acero para preparar un molde, después microinyectando plástico o un material similar en el molde para formar una estructura. Alternativamente, se puede ensamblar una estructura compuesta de vidrio, plástico, cerámica o similar. El separador puede ser, por ejemplo, una pieza de un material, *por ejemplo*, silicona, con agujeros espaciados en todas partes, de modo que cada agujero formará las paredes de un pocillo de prueba cuando se unan las tres piezas. El subdivisor puede ser, por ejemplo, una pieza delgada de un material, por ejemplo, silicona, con forma de pantalla o malla fina. El divisor en la superficie que separa las diferentes reacciones también puede ser una superficie recubierta a la que las soluciones no se adherirán, o una nanoestructura, o simplemente gotas individuales, capilares, canales o ubicaciones microfluídicas. En algunos ejemplos, la base es una pieza plana de material (por ejemplo, vidrio o plástico), por ejemplo, con la forma de la porción inferior de una microplaca típica utilizada para un ensayo bioquímico. La superficie superior de la base puede ser plana o puede formarse con hendiduras que se alinearán con la forma del subdivisor para proporcionar subdivisiones o pocillos completos dentro de cada pocillo de muestra. Las tres piezas se pueden unir mediante procedimientos estándares, por ejemplo, los procedimientos

utilizados en el ensamblaje de obleas de silicio.

Es posible emplear una amplia variedad de formatos de matriz para la disposición de los anclajes según la presente descripción. Un formato adecuado incluye un patrón bidimensional de células discretas (como 4096 cuadrados en una matriz de 64 por 64). Como aprecian los expertos en la materia, otros formatos de matriz que incluyen, entre otros, ranuras (rectangulares) y matrices circulares son igualmente adecuados para su uso (véase la Patente de los EE.UU. No. 5.981.185). En algunos ejemplos, la matriz es una placa de múltiples pocillos.

En un ejemplo, los anclajes de ácido nucleico preformados (por ejemplo, anclajes de oligonucleótidos) o sondas de captura de ácido nucleico que tienen una secuencia complementaria a al menos una porción de una NPP o NPPF (o su amplicón), se pueden ubicar sobre o dentro de la superficie de una prueba región por cualquiera de una variedad de técnicas convencionales, que incluyen la fijación química fotolitográfica o de serigrafía, la disposición por tecnología de chorro de tinta, capilar, pantalla o chip de canal de fluido, el diseño electroquímico utilizando conjuntos de electrodos, el contacto con un alfiler o pluma o la desnaturalización seguida de horneado o irradiando con UV sobre los filtros (véase, *por ejemplo*, Rava y col. (1996), Patente de los EE.UU. No. 5.545.531; Fodor y col. (1996), Patente de los EE.UU. No. 5.510.270; Zanzucchi y col. (1997), Patente de los EE.UU. No. 5,643,738; Brennan (1995). Patente de los EE.UU. No. 5.474.796; PCT WO 92/10092; PCT WO 90/15070). Los anclajes o sondas de oligonucleótidos pueden colocarse sobre la superficie de una región de prueba o pueden, por ejemplo, en el caso de una almohadilla de gel de poliacrilamida, incrustarse dentro de la superficie de tal manera que parte del anclaje o sonda sobresalga del gel estructura en porciones acuosas dentro del gel y la superficie del gel y esté disponible para interactuar con un conector, una NPP o una NPPF (o sus amplicones). Esto es cierto para superficies permeables y superficies parcialmente permeables, como una superficie donde la primera porción, como el área de la superficie en contacto con las soluciones que contienen enlaces bifuncionales, la NPP o NPPF es permeable, pero una segunda porción, como con cierta distancia a la superficie, no es permeable. En una realización, los anclajes o sondas de oligonucleótidos preformados se derivatizan en el extremo 5' con un grupo amino libre; disuelto a una concentración rutinariamente determinada empíricamente (*por ejemplo*, aproximadamente 1  $\mu$ M) en un tampón tal como tampón fosfato 50 mM, pH 8,5 y EDTA 1 mM; y distribuido con un dispensador de nanojet Pixus (Cartesian Technologies) en gotas de aproximadamente 10,4 nanolitros en ubicaciones específicas dentro de un pocillo de prueba cuya superficie superior es la de una placa de unión de ADN fresca y seca (Corning Costar). Dependiendo de la tasa relativa de unión y evaporación de oligonucleótidos, puede ser necesario controlar la humedad en los pocillos durante la preparación. En otra realización, los anclajes o sondas de oligonucleótidos se pueden sintetizar directamente en la superficie de una región de prueba, usando procedimientos convencionales tales como, por ejemplo, la desprotección activada mediante la luz de cadenas de oligonucleótidos en crecimiento (por ejemplo, junto con el uso de una "máscara" que dirige el sitio) o mediante la dispensación con patrón de gotitas de nanolitro de un compuesto desactivante utilizando un dispensador de nanojet. Por ejemplo, puede efectuarse la desprotección de todos los oligonucleótidos en crecimiento que van a recibir un solo nucleótido, y el nucleótido después se adiciona a lo largo de la superficie. En otra realización, los anclajes o sondas de oligonucleótidos se unen a la superficie a través de los extremos 3' de los oligonucleótidos, usando una metodología convencional.

#### 40 g. Detección de las NPP o NPPF utilizando procedimientos alternativos

En algunos ejemplos, después de la hibridación, el tratamiento con nucleasas y la amplificación opcional, las NPP o NPPF (o sus amplicones) se detectan utilizando procedimientos alternativos, como plataformas de alto rendimiento. En algunos ejemplos, se detectan NPP o NPPF (o sus amplicones) utilizando la electroforesis en gel, la cromatografía, la espectrometría de masas, la secuenciación, el análisis convencional de micromatrices, la detección durante la amplificación o la captura híbrida. En algunas realizaciones, las NPP o NPPF (o sus amplicones) no incluyen un marcador detectable y se utilizan procedimientos de detección indirectos. Dichos procedimientos son conocidos por un experto en la técnica e incluyen, entre otros, los descritos en esta invención.

En un ejemplo, se detectan NPP o NPPF (o sus amplicones) utilizando un ensayo basado en microesferas, tal como una matriz de microesferas. Un ejemplo de un ensayo basado en microesferas utiliza microesferas X-MAP® (Luminex, Austin, TX), como un ensayo QBEAD. En algunos ejemplos, las NPP o NPPF (o sus amplicones) se capturan en microesferas X-MAP® u otras microesferas por hibridación con un oligonucleótido asociado con las microesferas (por ejemplo, alrededor de 1 hora a aproximadamente 50°C). El marcador detectable incluida en las NPP o NPPF (o sus amplicones) se puede detectar, por ejemplo, mediante una citometría de flujo (por ejemplo, utilizando un Luminex 200, Flexmap 3D u otro instrumento adecuado).

En otro ejemplo, las NPP o NPPF (o sus amplicones) se detectan utilizando una micromatriz estándar. Un ejemplo de dicha matriz es una micromatriz de Nimblegen (Nimblegen, Madison, WI). En algunos ejemplos, las NPP o NPPF (o sus amplicones) se hibridan con una matriz que incluye oligonucleótidos que se unen específicamente a las NPP o NPPF (o sus amplicones). Se puede detectar el marcador detectable incluido en las NPP o NPPF (o sus amplicones).

En algunos ejemplos, las NPP o NPPF (o sus amplicones) se detectan utilizando un ensayo de "código de barras". Un ejemplo de tal ensayo es el Sistema de Análisis nCounter® (Nanostring Technologies, Seattle, WA). En algunos ejemplos, las NPP o NPPF (o sus amplicones) se hibridan con una sonda que incluye una o más etiquetas codificadas por color (un "código de barras"). La detección de las etiquetas codificadas por colores proporciona la identificación de las NPP o NPPF (o sus amplicones). Véase, *por ejemplo*, el documento WO 07/0761282; el documento WO 07/076129; y el documento WO 07/139

#### **h Secuencia de amplicones (qNPS)**

- 10 En algunos ejemplos, las NPP o NPPF resultantes (o sus amplicones) se secuencian, por ejemplo, secuenciando la NPP o NPPF completa (o sus amplicones), o una porción de la misma (como una cantidad suficiente para permitir la identificación de la molécula ácido nucleico diana). La descripción no se limita a un procedimiento de secuenciación particular. En algunos ejemplos, múltiples y diferentes NPP o NPPF (o sus amplicones) se secuencian en una sola reacción. En un ejemplo, se puede secuenciar una etiqueta de experimento de las NPP o NPPF (o sus amplicones), lo cual puede utilizarse para corresponder a una secuencia diana particular. Por lo tanto, si el extremo 3' de su amplicón NPPF tiene una secuencia en los nucleótidos de terminales 2 a 25 (como los nucleótidos de terminal 2 a 5 o 2 a 7, por ejemplo, los nucleótidos de terminal 2, 3, 4, 5, 6, 7, 8, 9 o 10) que representan una secuencia única para cada diana medida, entonces esto es todo su amplicón NPPF que necesita secuenciarse para identificar la diana, y contando el número de etiquetas de experimento secuenciadas, es posible determinar la cantidad de cada diana en la muestra.
- 20 En un ejemplo, las NPP o NPPF resultantes (o sus amplicones), como una compuesta de ADN, se secuencian usando el procedimiento de terminación de cadena. Los fragmentos resultantes se separan por tamaño, por ejemplo, mediante la electroforesis en un gel de poliacrilamida en placa, o en un tubo de vidrio estrecho (capilar) lleno de un polímero viscoso. Un procedimiento alternativo es la secuenciación del terminador de tinte. En otro ejemplo, se usa la pirosecuenciación, como los procedimientos comercializados por Biotage (para la secuenciación de bajo rendimiento) y 454 Life Sciences (para la secuenciación de alto rendimiento). En otro ejemplo, las NPP o NPPF (o sus amplicones) se secuencian usando una PCR puente (*por ejemplo*, Illumina®) (*por ejemplo*, HiSeq) o Ion Torrent®, 454®, Helicos®, PacBio®, Solid® (Applied Vioasystems) o cualquier otro sistema comercial de secuenciación. Los adaptadores de secuenciación (como las colas poli-A o poli T presentes en las NPP o NPPF (o sus amplicones), por ejemplo, introducidos usando PCR) se utilizan para la captura.

#### **6. Secuenciación cuantitativa directa de protección de nucleasa (qNPS)**

- 35 En algunos ejemplos, las moléculas de ácido nucleico en la muestra cuya expresión se va a medir se detectan en la muestra utilizando una alternativa del ensayo cuantitativo de protección de nucleasa descrito anteriormente, por ejemplo, como se describe en la Solicitud de Patente de los EE.UU. No. PCT/US2017/017512 (FIG. 9A). Este procedimiento también utiliza las NPPF y CFS, pero en lugar de detectar o secuenciar el sustituto de NPPF, este procedimiento permite la secuenciación directa de una molécula de ácido nucleico diana.
- 40 Los procedimientos incluyen poner en contacto la muestra con al menos una NPPF en condiciones suficientes para que la NPPF se una específicamente a, o se hibride con, la molécula de ácido nucleico diana. Si la NPPF incluye una secuencia flanqueante tanto en el extremo 5' como en el extremo 3', en algunos ejemplos, la secuencia de cada NPPF es diferente y no se complementarían entre sí. En un ejemplo, la NPPF (por ejemplo, en una o ambas secuencias flanqueantes y/o la secuencia de la región complementaria a toda o una porción de la molécula de ácido nucleico diana) incluye al menos un dUTP, como al menos dos, tres, cuatro o cinco dUTP. En un ejemplo, todas las "Ts" en una NPPF se reemplazan por "U". La presencia de tales bases permite que la NPPF de una sola cadena se degrade o se destruya con uracilo ADN glicosilasa (UDG) en la etapa posterior (*por ejemplo*, después de la desnaturalización de la NPPF de la diana ligada, pero antes de la secuenciación, por ejemplo, antes o durante la amplificación de una diana ligada).
- 50 Las secuencias flanqueantes son complementarias para las CFS. En algunos ejemplos, al menos una secuencia flanqueante incluye al menos un dUTP. En algunos ejemplos, la NPPF incluye dos secuencias flanqueantes, cada una con al menos un dUTP. En algunos ejemplos, la NPPF incluye dos secuencias flanqueantes, pero solo una presenta al menos un dUTP. En algunos ejemplos, la NPPF incluye una única secuencia flanqueante que presenta al menos un dUTP. En algunos ejemplos, la ubicación del dUTP está cerca de la secuencia complementaria a una región de la molécula de ácido nucleico diana, como a 1, 2, 3, 4, o 5 bases lejos de (*por ejemplo*, dentro de 1, 2, 3, 4, o 5 bases de) la secuencia complementaria a una región de la molécula de ácido nucleico diana. En algunos ejemplos, la ubicación del dUTP está al menos a dos bases (como al menos 3, 4 o 5 bases) lejos de la secuencia complementaria a una región de la molécula de ácido nucleico diana.
- 60 Los procedimientos incluyen además poner en contacto la muestra con al menos una molécula de ácido nucleico que tiene complementariedad con la secuencia flanqueante (CFS) en condiciones suficientes para que la CFS se una o

hibride específicamente con la secuencia flanqueante de la NPPF. Por ejemplo, si la NPPF presenta una secuencia flanqueante 5', la muestra se pone en contacto con una molécula de ácido nucleico que presenta una complementariedad de secuencia a la secuencia flanqueante 5' (5CFS) y un fosfato del extremo 5', en condiciones suficientes para que la secuencia flanqueante 5' se hibride específicamente con la 5CFS; De manera similar, si la NPPF presenta una secuencia flanqueante 3', la muestra se pone en contacto con una molécula de ácido nucleico que presenta una complementariedad de secuencia a la secuencia flanqueante 3' (3CFS), en condiciones suficientes para que la secuencia flanqueante 3' se hibride específicamente con la 3CFS; En un ejemplo, al menos una de las 3CFS y 5CFS incluye un residuo de captura que permite la captura, separación, recuperación o aislamiento de una secuencia diana. Un experto en la materia apreciará que, en lugar de usar una sola CFS para proteger una secuencia flanqueante, se pueden usar múltiples CFS para proteger una secuencia flanqueante (por ejemplo, se pueden usar múltiples 5CFS para proteger una secuencia flanqueante 5'). En algunos ejemplos, la molécula de ácido nucleico diana es el ADN, y la 5CFS y la 3CFS son ADN, o la 5CFS es ADN y la 3CFS es ARN. En algunos ejemplos, la molécula de ácido nucleico diana es ARN, y la 5CFS es ADN y la 3CFS es ARN o la 5CFS es ARN y la 3CFS es ARN. En algunos ejemplos, la molécula de ácido nucleico diana es un ARN o ADN, y la 5CFS y/o la 3CFS es un oligo híbrido de ARN-ADN, por ejemplo, donde la o las bases 5' de las 5CFS y/o la o las bases 3' de las 3CFS son ARN, y el resto de las 5CFS y 3CFS son ADN.

Esto da como resultado la generación de moléculas de NPPF que se han unido a una molécula de ácido nucleico diana (o una porción de la misma), así como a la o las CFS, generando así una molécula bicatenaria que incluye bases de la NPPF involucradas en la hibridación con una base complementaria en la diana y la CFS. La o las CFS se hibridan y, por lo tanto, protegen su secuencia flanqueante correspondiente de la digestión con la nucleasa en las etapas posteriores. En algunos ejemplos, cada CFS es la longitud exacta de su secuencia flanqueante correspondiente. En algunos ejemplos, la CFS es completamente complementaria a su secuencia flanqueante correspondiente. Sin embargo, un experto en la materia apreciará que el extremo 3' de una 5CFS que protege una secuencia flanqueante del extremo 5' o el extremo 5' de una 3CFS que protege la secuencia flanqueante del extremo 3' puede tener una diferencia, como un nucleótido en cada una de estas posiciones.

Después de permitir que la molécula de ácido nucleico diana y las CFS se unan a las NPPF, el procedimiento incluye además poner en contacto la muestra con una nucleasa específica para moléculas de ácido nucleico de una sola cadena (ss) o regiones ss de una molécula de ácido nucleico, tales como la nucleasa S1, en condiciones suficientes para eliminar las bases de ácido nucleico que no se hibridan con una base complementaria. Así, por ejemplo, las NPPF que no se han unido a la molécula de ácido nucleico diana o CFS, así como a las moléculas de ácido nucleico diana no unido, a otras moléculas de ácido nucleico ss en la muestra y a las CFS no unidos, se degradan. Esto genera una muestra digerida que incluye NPPF intactas presentes como aductos bicatenarios hibridados con la 5CFS, la 3CFS, o ambas, y una porción del ácido nucleico diana. En algunos ejemplos, por ejemplo, si la NPPF está compuesta de ADN, la nucleasa puede incluir una exonucleasa, una endonucleasa o una combinación de las mismas.

Posteriormente, las CFS pueden ligarse al ácido nucleico diana, por ejemplo, el fosfato 5' de la 5CFS se liga al extremo 3' de la molécula de ácido nucleico diana, y el extremo 3' de la 3CFS se liga al extremo 5' de la molécula de ácido nucleico diana, generando así una molécula de ácido nucleico diana ligada (denominada en esta invención como diana ligada).

La molécula de ácido nucleico diana ligada de NPPF bicatenaria se separa en moléculas de ácido nucleico ss (por ejemplo, calentando o aumentando el pH de la muestra), generando así una mezcla de NPPF ss y moléculas de ácido nucleico ss diana ligadas. Las moléculas de ácido nucleico diana ligadas a ss se recuperan opcionalmente, por ejemplo, capturándolas, por ejemplo, por hibridación o mediante el uso de un residuo de captura (*por ejemplo*, en la 5CFS o 3CFS). En algunos ejemplos, la mezcla de las NPPF ss y las moléculas de ácido nucleico diana ligadas a ss se incuban con uracilo ADN glicosilasa (UDG) en condiciones suficientes para degradar o destruir las NPPF ss que tienen al menos un dUTP.

Los procedimientos pueden incluir una o más etapas que permiten secuenciar la captura o recuperación de la diana, por ejemplo, mediante el uso de una NPPF, una 5CFS o una 3CFS. Por ejemplo, como se muestra en las FIG. 9B a 9D, dicha captura puede ocurrir durante la hibridación, después de la digestión de nucleasas o después de la hibridación, pero antes de la ligadura. Las FIG. 9A y 9C muestran un ejemplo donde la captura se realiza después de la digestión con nucleasa. Además, se pueden incluir etapas de captura adicionales, por ejemplo, después de la ligadura. En un ejemplo, la NPPF, la 5CFS o la 3CFS está unida a un sustrato sólido (como una placa de múltiples pocillos), lo que permite capturar moléculas cuando se hibridan con la NPPF, la 5CFS o la 3CFS anclada. En otro ejemplo, se puede usar un residuo de captura (como una partícula o marcador unido a la diana, la NPPF, la 5CFS o la 3CFS) para permitir la separación física de moléculas unidas o hibridadas a la molécula que contiene el residuo de captura de los otros componentes de la muestra. Dependiendo del procedimiento de captura, dichos procedimientos pueden incluir la centrifugación para recolectar residuos de captura sólidos, la captura de microesferas magnéticas, la unión o hibridación a un soporte sólido, la filtración, etc. Además, dichas etapas de captura pueden incluir lavar las

moléculas de ácido nucleico capturadas (por lo tanto, permitiendo que las moléculas no capturadas, tales como NPPF no hibridadas y CFS no hibridadas, se separen o eliminen). Dichos procedimientos pueden incluir la liberación de las moléculas de ácido nucleico capturadas de un soporte sólido, como un pocillo de una placa de múltiples pocillos o un tubo o recipiente de reacción única, o las moléculas de ácido nucleico capturadas pueden permanecer fijadas al soporte sólido o dentro de él. En algunos ejemplos, esta etapa incluye la centrifugación, por ejemplo, para capturar la molécula de ácido nucleico diana en el fondo de un recipiente o tubo, y lavar para permitir la eliminación de agentes no deseados o innecesarios. En algunos ejemplos, esta etapa incluye la captura de microesferas magnéticas de la molécula de ácido nucleico diana y el lavado para permitir la eliminación de agentes no deseados o innecesarios. En algunos ejemplos, esta etapa incluye la filtración para capturar la molécula de ácido nucleico diana y el lavado para permitir la eliminación de agentes no deseados o innecesarios. Esta etapa puede incluir el lavado de la molécula de ácido nucleico diana, que en algunos ejemplos elimina los agentes que ya no son necesarios (*por ejemplo*, la nucleasa) y/o permite que la molécula de ácido nucleico diana se suspenda en otra solución, lo que puede facilitar una o más etapas siguientes en una realización de procedimiento particular.

15 En algunos ejemplos, se realizan múltiples etapas del procedimiento (como dos, tres o cuatro de las etapas mostradas en la FIG. 9A) en el mismo recipiente o recipiente. Por ejemplo, los procedimientos descritos permiten que el o los componentes deseados permanezcan en el mismo recipiente durante varias etapas, mientras que los componentes no deseados o innecesarios se eliminan (*por ejemplo*, usando repetidas etapas de captura y lavado). Por ejemplo, la muestra a analizar se puede lisar en un recipiente. Las NPPF y CFS adecuadas se adicionan a ese recipiente, la nucleasa se adiciona al mismo recipiente y la molécula de ácido nucleico ds se captura en el recipiente (*por ejemplo*, con las microesferas magnéticas siendo extraídas del fondo del recipiente). La molécula de ácido nucleico ds capturada se lava para eliminar los agentes no deseados y los tampones o reactivos deseados se adicionan al recipiente (*por ejemplo*, un tampón de ligasa). Las moléculas de ácido nucleico diana ligadas a ss pueden capturarse en el recipiente y lavarse.

25 Las moléculas de ácido nucleico diana ligadas se amplifican opcionalmente, por ejemplo, usando la amplificación mediante PCR. Tales procedimientos pueden usarse para adicionar una etiqueta de experimento y/o un adaptador de secuencia a la diana ligada, y/o para aumentar el número de copias de la diana ligada. Al menos una porción de la molécula de ácido nucleico diana ligada a ss (o su amplicón) se secuencia, determinando así la secuencia de la al menos una molécula de ácido nucleico diana en la muestra. En algunos ejemplos, las moléculas de ácido nucleico diana son ARN, y el procedimiento incluye la transcripción inversa en ADN antes de la secuenciación.

Las dianas ligadas pueden amplificarse usando uno o más cebadores de amplificación, generando así amplicones diana ligados. Al menos un cebador de amplificación incluye una región que es complementaria a una CFS ligada a la diana. En algunos ejemplos, la diana se liga a una 5CFS en su extremo 5' y a una 3CFS en su extremo 3', y se usan dos cebadores de amplificación, donde un cebador de amplificación tiene una región que es complementaria a una región de la 5CFS y el otro cebador de amplificación tiene una región que es complementaria a una región de la 3CFS. En algunos ejemplos, la diana se liga a una 5CFS o una 3CFS en su extremo 5' y 3', respectivamente, y se usa un cebador de amplificación (*por ejemplo*, usando una amplificación rápida o extremos de ADNc), donde el cebador de amplificación tiene una región que es complementaria a una región de la 5CFS o la 3CFS. Uno o ambos cebadores de amplificación pueden incluir una secuencia que permita la unión de una etiqueta de experimento y/o un adaptador de secuenciación al amplicón diana ligado durante la amplificación, y uno o ambos cebadores pueden marcarse para permitir que se marque el amplicón NPPF. En algunos ejemplos, se adicionan tanto una etiqueta de experimento como un adaptador de secuenciación, *por ejemplo*, en los extremos opuestos del amplicón diana ligado. Por ejemplo, el uso de tales cebadores puede generar una etiqueta de experimento y/o un adaptador de secuencia que se extiende desde el extremo 5' o el extremo 3' del amplicón diana ligado, o desde el extremo 3' y el extremo 5' hasta aumentar el grado de multiplexación posible. La etiqueta del experimento puede incluir una secuencia de ácido nucleico única que permite la identificación de una muestra, sujeto o secuencia de ácido nucleico diana. El adaptador de secuenciación puede incluir una secuencia de ácido nucleico que permite la captura del amplicón diana ligado resultante en una plataforma de secuenciación. En algunos ejemplos, los cebadores se eliminan de la mezcla antes de la secuenciación.

La diana ligada o los amplicones diana ligados (o una porción de los mismos) se secuencian, *por ejemplo*, usando los procedimientos descritos anteriormente. Se puede usar cualquier procedimiento para secuenciar la diana ligada o los amplicones diana ligados, y la descripción no se limita a los procedimientos de secuenciación particulares. En algunos ejemplos, el procedimiento de secuenciación utilizado es la secuenciación de terminación de cadena, la secuenciación de terminación de tinte, la secuenciación de pirosecuenciación, la secuenciación de nanoporos o la secuenciación masiva paralela (también llamada secuenciación de próxima generación (o NGS)), que se ejemplifica mediante la máquina de genoma personal ThermoFisher Ion Torrent™ (PGM™), los secuenciadores NGS de la marca Illumina (*por ejemplo*, MiSeq™, HiSeq™) (o como se deriva de la secuenciación Solexa™) y la secuenciación 454 disponible de Roche Life Sciences. En algunos ejemplos, se utiliza la secuenciación de moléculas individuales. En algunos ejemplos, el procedimiento también incluye comparar la secuencia diana obtenida con una base de datos de secuencia, *por ejemplo*, para determinar si una mutación diana está presente o ausente. En algunos ejemplos, el

procedimiento incluye determinar el número (*por ejemplo*, el conteo) de cada secuencia diana obtenida.

La FIG. 9A es un diagrama esquemático que muestra una visión general de una realización para usar la NPPF 202 para secuenciar una molécula de ácido nucleico para los procedimientos descritos. Como se muestra en la etapa 1, una muestra (como una que se sabe o se sospecha que contiene un ácido nucleico diana, 200 que se ha tratado con una interrupción de la muestra o un tampón de lisis (*por ejemplo*, lisada o tratada de otro modo para hacer que los ácidos nucleicos sean accesibles) se pone en contacto o se incuba con una pluralidad de sondas de protección de nucleasa que tienen una o más secuencias flanqueantes (NPPF) 202 (mostradas aquí con una secuencia flanqueante 5' y 3', 204 y 206, respectivamente), que incluye al menos una NPPF que se une específicamente a un primer ácido nucleico diana 200 (tal como un ADN o ARN diana). La reacción también puede incluir otras NPPF que se unen específicamente a un segundo ácido nucleico diana y así sucesivamente. Por ejemplo, el procedimiento puede usar una o más NPPF diferentes diseñadas para ser específicas para cada molécula de ácido nucleico diana única, como las NPPF específicas para cada uno de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. En algunos ejemplos, la pluralidad de NPPF puede incluir más de una (como 2, 3, 4, 5, 10, 20, 50 o más) NPPF específicas para una sola molécula de ácido nucleico diana (que se conoce como un conjunto en mosaico de NPPF). La reacción también incluye moléculas de ácido nucleico que son complementarias a las secuencias flanqueantes (CFS) 208, 210. Por lo tanto, si la NPPF tiene una secuencia flanqueante 5' 204, la reacción incluirá una secuencia complementaria a la secuencia flanqueante 5' (5CFS) 208 y si la NPPF tiene una secuencia flanqueante 3' 206, la reacción incluirá una secuencia complementaria a la secuencia flanqueante 3' (3CFS) 210. En algunos ejemplos, la 5CFS 208 tiene un fosfato en el extremo 5', que permite la unión a la diana 200 en una etapa posterior. En algunos ejemplos, al menos una de las CFS incluye un residuo de captura 212, que permite la recuperación de la diana en el momento deseado. Aunque el residuo de captura 212 se muestra en la 5CFS 208, se apreciará que en las realizaciones donde está presente, puede estar alternativamente en la 3CFS 210, la NPPF 202 o la diana 200. Un experto en la materia apreciará que la secuencia de las CFS variará dependiendo de la secuencia flanqueante presente. Además, se puede usar más de una CFS para garantizar que una región de flanqueo esté protegida (*por ejemplo*, se pueden usar al menos dos CFS que se unan a diferentes regiones de una sola secuencia flanqueante). La CFS puede incluir bases naturales o no naturales.

El residuo de captura 212 en la FIG. 2A es opcional, o puede estar en la 3CFS 210 o la diana 200. La muestra, las NPPF y las CFS se incuban en condiciones suficientes para que las NPPF se unan específicamente a (*por ejemplo*, hibridarse con) su molécula de ácido nucleico diana respectiva, y para que las CFS se unan a (*por ejemplo*, se hibriden con) su secuencia complementaria en la secuencia flanqueante de NPPF. En algunos ejemplos, las CFS 208, 210 se adicionan en exceso de las NPPF 202, por ejemplo, al menos 5 veces más CFS que NPPF (exceso molar), como al menos 6, 7, 8, 9, 10, 20, 40, 50 o 100 veces más CFS que NPPF. En algunos ejemplos, las NPPF 202 se adicionan en exceso de las moléculas de ácido nucleico totales en la muestra, por ejemplo, al menos 50 veces más NPPF que las moléculas de ácido nucleico totales en la muestra (exceso molar), como al menos 75, 100, 200, 500 o 1000 veces más NPPF que las moléculas de ácido nucleico totales en la muestra. Por conveniencia experimental, se puede incluir una concentración similar de cada NPPF para hacer un cóctel, de modo que para la diana de ácido nucleico más abundante medida habrá al menos 50 veces más NPPF para esa diana de ácido nucleico, por ejemplo, al menos 100 veces más. El exceso real y la cantidad total de todas las NPPF utilizadas solo se limitan según la capacidad de la nucleasa (*por ejemplo*, nucleasa S1) para destruir todas las NPPF que no se hibridan con dianas de ácido nucleico diana. En algunos ejemplos, la reacción se calienta, por ejemplo, se incuba durante la noche (por ejemplo, durante 16 horas) a 50°C.

Como se muestra en la etapa 2 en la FIG. 9A, después de permitir que ocurran las reacciones de unión/hibridación, la muestra se pone en contacto con un reactivo (como una nucleasa) específico para moléculas de ácido nucleico de una sola cadena (ss) en condiciones suficientes para eliminar (o digerir) las moléculas de ácido nucleico ss, como las moléculas de ácido nucleico no unidas (como las NPPF no unidas, las CFS no unidas y moléculas de ácido nucleico diana no unidas, o porciones de tales moléculas que permanecen de una sola cadena, por ejemplo, las porciones de una molécula de ácido nucleico diana no unida a la NPPF). Esto da como resultado la generación de un complejo ds NPPF/diana hibridado (o dúplex) 214. Como se muestra en la FIG. 9A, la incubación de la muestra con una nucleasa específica para las moléculas de ácido nucleico ss da como resultado la degradación de cualquier molécula de ácido nucleico ss presente, dejando moléculas de ácido nucleico bicatenario intactas, incluyendo las NPPF que se han unido a las mismas, las CFS y la molécula de ácido nucleico diana. Por ejemplo, la reacción puede incubarse a 50°C durante 1,5 horas con nucleasa S1 (aunque la hidrólisis puede ocurrir a otras temperaturas y llevarse a cabo durante otros períodos de tiempo, y, en parte, el tiempo y la temperatura requeridos serán una función de la cantidad de nucleasa y la cantidad de ácido nucleico que se requiere hidrolizar, así como la Tm de la región bicatenaria que se protege).

Como se muestra en la etapa 3 de la FIG. 9A, el complejo NPPF/diana 214 (que se captura o recupera en algún momento antes de la ligadura, véanse las FIG. 9B-9D, y el material no hibridado eliminado), se expone a condiciones de ligadura, que permiten que la secuencia diana 200 ( que puede ser ADN o ARN) se ligue a las CFS (*por ejemplo*, a la 5CFS 208, la 3CFS 210 o ambas), generando así una diana ligada 216. Puede usarse cualquier ligasa capaz de

- unirse covalentemente a la secuencia diana y la o las CFS, como una ligasa de ADN o ARN, como una ligasa de ADN T4, una ligasa de ARN T4 o una ligasa Taq (por lo tanto, tales ligasas pueden estar presentes en un kit proporcionado en esta invención, así como también estar individualmente o en un tampón de ligadura). La ligasa utilizada puede depender del nucleótido (ribonucleótido o desoxirribonucleótido) presente en la CFS más cercana a la secuencia diana.
- 5 En algunos ejemplos, una ligasa se diluye en un tampón (por ejemplo, uno que contiene cofactores esenciales como NAD<sup>+</sup> o ATP, un agente tamponador y cationes se adiciona a la mezcla que contiene los dúplex NPPF/diana y se incuba a una temperatura de aproximadamente 4 a 60°C (dependiendo de la ligasa) (por ejemplo, al menos 4, 20, 25, 30, 37, 50 o 60°C, como de 4 a 37°C, de 4 a 25°C o de 20 a 25°C) durante al menos 15 o 30 minutos o al menos 1, 2, 4, 6, 12 o 16 horas, para ligar las CFS a la molécula de ácido nucleico diana (donde las CFS y la diana ya están
- 10 hibridadas con una NPPF). En un ejemplo, un tampón para la ADN ligasa T4 incluye Tris-HCl 50 mM, pH 7,5, MgCl<sub>2</sub> 10 mM, Ditiotreitól 10 mM y ATP 1 mM. La 5CFS 208 puede incluir un fosfato en el extremo 5', que puede ligarse al extremo 3' de la diana 200. La 3CFS 210 puede incluir un extremo 3' (como un nucleótido terminal 3'), que puede ligarse al fosfato del extremo 5' de la diana. Si solo una secuencia flanqueante está presente en la NPPF, solo una CFS se habrá unido en el complejo NPPF/diana, y la diana se ligará a una sola CFS. Las CFS pueden ser ADN o ARN
- 15 (o una mezcla de ambos tipos de nucleótidos). Si la CFS o la diana es el ADN, el donante de fosfato del extremo 5' para la ligadura al ADN no puede ser un ribonucleótido. En un ejemplo, si la diana es ARN, la 5CFS 208 es ADN o ARN y la 3CFS 210 es ARN. En un ejemplo, si la diana es ADN, la 5CFS 208 es ADN (o al menos el nucleótido del extremo 5' de la 5CFS 208 es un dNTP que se unirá al extremo 3' de la diana) o ARN y la 3CFS 210 es ARN
- 20 Como se muestra en la etapa 4 de la FIG. 9A, después de ligar las CFS a la diana, la NPPF 202 se separa de la diana ligada 216. Es decir, el complejo NPPF/diana 214 (*por ejemplo*, la molécula de ácido nucleico bicatenario) se separa en dos moléculas de ácido nucleico de una sola cadena, la NPPF 202 y la diana ligada 216. Esta etapa también puede eliminar las moléculas de ácido nucleico no ligadas (*por ejemplo*, la CFS no ligada). Por ejemplo, se puede calentar la reacción o alterar el pH (*por ejemplo*, para que la reacción tenga un pH básico) en condiciones que permitan que la
- 25 NPPF 202 se disocie de la diana ligada 216, lo que da como resultado una población mixta de NPPF 202 de una sola cadena y dianas ligadas de una sola cadena 216. En algunos ejemplos, la diana ligada 216 se separa de esta mezcla usando el residuo de captura 212 en la CFS. Los procedimientos de captura descritos en esta invención pueden usarse para capturar la diana ligada 216. Por ejemplo, si el residuo de captura 212 es una microesfera, la diana ligada 216 puede recuperarse mediante el uso de una centrifugación. Si el residuo de captura 212 es una microesfera de metal,
- 30 la diana ligada 216 puede recuperarse usando un imán. Si el residuo de captura 212 incluye un grupo carboxilo, la diana ligada 216 puede capturarse usando un sustrato sólido marcado adecuadamente con amina. Después de la captura de la diana ligada 216, el residuo de la reacción (*por ejemplo*, NPPF 202) puede eliminarse (*por ejemplo*, mediante un lavado). En algunos ejemplos, la diana ligada 216 se captura mediante filtración.
- 35 Como se muestra en la etapa 5 de la FIG. 9A, la diana ligada aislada 216 (o sus amplicones) se secuencia a continuación. En algunos ejemplos, una pluralidad de dianas ligadas se secuencia en paralelo, por ejemplo, de manera simultánea o contemporánea. Por lo tanto, este procedimiento puede usarse para secuenciar una pluralidad de secuencias diana ligadas.
- 40 Opcionalmente, las dianas ligadas 216 se amplifican (*por ejemplo*, usando una PCR), se lavan, o ambos, antes de la secuenciación. En algunos ejemplos, las NPPF 202 se degradan con uracilo ADN glicosilasa (UDG), por ejemplo, si al menos una secuencia flanqueante, o la secuencia de captura, incluye dUTP. Tal degradación puede ocurrir antes o durante la amplificación. En algunos ejemplos, como si la diana es ARN, se puede incluir una etapa de transcripción inversa como parte de la amplificación. Por lo tanto, los amplicones diana ligados resultantes 226 entonces pueden
- 45 secuenciarse. Los cebadores o sondas de PCR pueden incluir una o más etiquetas de experimento 222, 224 y/o adaptadores de secuenciación 218, 220 (*por ejemplo*, que permiten que las dianas ligadas sean secuenciadas por una plataforma de secuenciación particular y, por lo tanto, tales adaptadores son complementarios para capturar secuencias en un chip de secuenciación). Al menos una porción de los cebadores/sondas de PCR es específica para la 5CFS 208 y/o la 3CFS 210. En algunos ejemplos, la concentración de los cebadores supera la diana ligada 216,
- 50 por ejemplo, en exceso, al menos 10.000, 50.000, 100.000, 150.000, 200.000, 400.000, 500.000, 600.000, 800.000 o 1.000.000 veces. En algunos ejemplos, la concentración de cebadores 208 en la reacción es de al menos 200 nM (tal como al menos 400, 500, 600, 750 o 1000 nM).
- Como se muestra en las FIG. 9B a 9D, en uno o varios puntos antes de la ligadura (*por ejemplo*, en la etapa 3 de la
- 55 FIG. 9A), la diana (200 de la FIG. 9A) y otras moléculas unidas/hibridadas a esta, se recuperan o capturan (etapa 320), permitiendo así la eliminación de los agentes restantes en la muestra (por ejemplo, el material no hibridado) y la adición de nuevos reactivos. Además, la captura de la diana permite el lavado posterior de las moléculas de ácido nucleico hibridadas o unidas de otro modo a la diana (*por ejemplo*, para inactivar o eliminar enzimas residuales).
- 60 En un ejemplo, la diana se recupera o captura en la hibridación (FIG. 9B). Como se muestra en la FIG. 9B, después de la desnaturalización opcional 300, la muestra que contiene la diana se incuba con NPPF y CFS en condiciones que permiten la hibridación 310, produciendo así un complejo hibridado. En tal ejemplo, la diana, junto con la NPPF con la

que se hibridará, y las CFS que pueden hibridarse con la NPPF, se incuban en presencia de un soporte sólido que contiene un material que permite la captura del complejo hibridado resultante 320. Por ejemplo, el residuo de la captura (*por ejemplo*, 212 en la FIG. 9A) presente en la 3CSF, la 5CSF y/o la NPPF puede utilizarse como soporte sólido. Por ejemplo, si el residuo de la captura es una microesfera o placa (*por ejemplo*, unida a una CFS usando un enlace de carboxi-amina, como un grupo amino en la CFS y un carboxi en la microesfera), el soporte sólido es la microesfera o placa. En este caso, el residuo de captura/soporte sólido permite que el complejo hibridado se una al soporte sólido (*por ejemplo*, porque la NPPF se hibridará con la CFS unida al soporte sólido, y la diana y otras CFS se hibridarán con la NPPF). El procedimiento puede incluir etapas de lavado después de la hibridación. Una o más de las etapas posteriores, como la digestión de nucleasa 330, los lavados 340, la ligadura 350, la digestión UDG 370 (opcional, solo si la NPPF incluye dUTP) y/o la amplificación mediante PCR 380 pueden tener lugar en presencia de soporte sólido, lo que permite eliminar los reactivos y adicionarlos a la diana capturada.

En un ejemplo, la diana se recupera o captura después del tratamiento con nucleasa (*por ejemplo*, después de la etapa 2 en la FIG. 9A), pero antes de la ligadura (*por ejemplo*, antes de la etapa 3 en la FIG. 9A) (FIG. 9C). Como se muestra en la FIG. 9C, después de la desnaturalización opcional 400, la hibridación 410 y la digestión de nucleasas 430, que produce un complejo NPPF/diana (214 en la figura 9A), el complejo NPPF/diana se recupera de la mezcla de muestra en la etapa 420. En algunos ejemplos, el complejo NPPF/diana se recupera utilizando un residuo de captura en la CFS, sin separar el dúplex NPPF/diana en moléculas de ácido nucleico ss. Por ejemplo, el residuo de la captura (*por ejemplo*, 212 en la FIG. 9A) presente en la 3CSF, la 5CSF y/o la NPPF puede utilizarse como soporte sólido. Por ejemplo, si el residuo de la captura es una microesfera o placa (*por ejemplo*, unida a una CFS usando un enlace de carboxi-amina, como un grupo amino en la CFS y un carboxi en la microesfera), el soporte sólido es la microesfera o placa. En este caso, el residuo de captura/soporte sólido es parte del complejo NPPF/diana, lo que permite su captura o recuperación. El procedimiento puede incluir etapas 440 de lavado después de la recuperación. Una o más de las etapas posteriores, como la ligadura 450, el lavado 460, la UDG 470 (opcional, solo si la NPPF incluye dUTP) y/o la amplificación mediante PCR 480 pueden tener lugar en presencia de soporte sólido, lo que permite eliminar los reactivos y adicionarlos a la diana capturada.

En un ejemplo, la diana se recupera o captura después de la hibridación (*por ejemplo*, después de la etapa 1 en la FIG. 9A), pero antes de la digestión de la nucleasa (*por ejemplo*, antes de la etapa 2 en la FIG. 9A) (FIG. 9D). Como se muestra en la FIG. 9D, después de la desnaturalización opcional 500 y la hibridación 510, que produce un complejo hibridado, el complejo hibridado resultante se recupera de la mezcla de muestra en la etapa 520. En algunos ejemplos, el complejo hibridado se recupera utilizando un residuo de captura en la CFS. Por ejemplo, el residuo de la captura (*por ejemplo*, 212 en la FIG. 9A) presente en la 3CSF, la 5CSF y/o la NPPF puede utilizarse como soporte sólido. Por ejemplo, si el residuo de la captura es una microesfera o placa (*por ejemplo*, unida a una CFS usando un enlace de carboxi-amina, como un grupo amino en la CFS y un carboxi en la microesfera), el soporte sólido es la microesfera o placa. En este caso, el residuo de captura/soporte sólido es parte del complejo hibridado, lo que permite su captura o recuperación. El procedimiento puede incluir etapas de lavado después de la recuperación. Una o más de las etapas posteriores, como la digestión de nucleasa 530, la ligadura 550, los lavados 540, 560, la digestión UDG 570 (opcional, solo si la NPPF incluye dUTP) y/o la amplificación mediante PCR 580 pueden tener lugar en presencia de soporte sólido, lo que permite eliminar los reactivos y adicionarlos a la diana capturada.

### **B. Proteínas para detectar la expresión génica**

En algunas realizaciones de los procedimientos descritos, determinar el nivel de expresión génica en una muestra incluye detectar una o más proteínas (*por ejemplo*, determinando las cantidades relativas o reales de tales proteínas) en la muestra. Los procedimientos rutinarios de detección de proteínas son conocidos en la técnica, y la descripción no se limita a procedimientos particulares de detección de proteínas.

Los productos genéticos proteicos (*por ejemplo*, los de la Tabla 1) se pueden detectar y el nivel de expresión de las proteínas en la muestra se puede determinar a través de nuevos epítomos reconocidos por agentes de unión específicos de proteínas (como anticuerpos o aptámeros) específicos para la proteína diana (como las de la Tabla 1) utilizados en inmunoensayos, tales como los ensayos ELISA, los ensayos de inmunotransferencia, los ensayos de citometría de flujo, los ensayos inmunohistoquímicos, el inmunoensayo enzimático, los ensayos inmunoradiométricos, los ensayos de transferencia Western, los ensayos inmunofluorescentes, los ensayos quimioluminiscentes y otras estrategias de detección de péptidos. En general, estos procedimientos utilizan anticuerpos monoclonales o policlonales.

Por lo tanto, el nivel de expresión de la proteína diana (como las de la Tabla 1) presente en la muestra biológica y, por lo tanto, la cantidad de proteína expresada se detecta usando un agente de unión específico de la proteína diana, como un anticuerpo de un fragmento de la misma, que puede marcarse de manera detectable. En algunas realizaciones, el agente de unión específico es un anticuerpo, tal como un anticuerpo policlonal o monoclonal, que se une específicamente a la proteína diana (como las de la Tabla 1). Por lo tanto, determinar el nivel o la cantidad de



proteína en una muestra biológica incluye poner en contacto una muestra del sujeto con un agente de unión específico de proteína (como un anticuerpo que se une específicamente a una proteína que se muestra en la Tabla 1), detectando si el agente de unión está unido por la muestra, y, de ese modo, medir la cantidad de proteína presente en la muestra.

Por ejemplo, el agente de unión específico es un anticuerpo policlonal o monoclonal, que se une específicamente a la proteína diana (como las de la Tabla 1). Un experto en la materia apreciará que existen fuentes comerciales de anticuerpos para proteínas diana, como las de la Tabla 1.

La presencia de una proteína diana (como las de la Tabla 1) se puede detectar con múltiples agentes de unión específicos, como uno, dos, tres o más agentes de unión específicos. Por lo tanto, los procedimientos pueden utilizar más de un anticuerpo. En algunas realizaciones, uno de los anticuerpos está unido a un soporte sólido, tal como una placa multipocillo (por ejemplo, una placa de microtitulación), una microesfera, una membrana o similares. En la práctica, las placas de microtitulación pueden utilizarse convenientemente como la fase sólida. Sin embargo, las reacciones de anticuerpos también pueden realizarse en una fase líquida.

En algunos ejemplos, el procedimiento puede incluir poner en contacto la muestra con un segundo anticuerpo que se une específicamente al primer anticuerpo que se une específicamente a la proteína diana (como las de la Tabla 1). En algunos ejemplos, el segundo anticuerpo se marca de manera detectable, por ejemplo, con un fluoróforo (como FITC, PE, una proteína fluorescente y similares), una enzima (como HRP), un marcador de radio o una nanopartícula (como una partícula de oro o un nanocrystal semiconductor, como un punto cuántico (QDOT®)). En este procedimiento, una enzima que se une al anticuerpo reaccionará con un sustrato adecuado, tal como un sustrato cromogénico, de modo tal que se produzca un residuo químico que pueda detectarse, por ejemplo, mediante espectrofotometría, fluorimetría o a través de un medio visual. Las enzimas que se pueden usar para marcar de manera detectable el anticuerpo incluyen, de modo no taxativo, malato deshidrogenasa, nucleasa estafilocócica, delta-5-esteroide isomerasa, levadura alcohol deshidrogenasa, alfa-glicerofosfato, deshidrogenasa, triosa fosfato isomerasa, peroxidasa de rábano picante, fosfatasa alcalina, asparaginasa, glucosa oxidasa, beta-galactosidasa, ribonucleasa, ureasa, catalasa, glucosa-6-fosfato deshidrogenasa, glucoamilasa y acetilcolinesterasa. Adicionalmente, la detección se puede lograr mediante procedimientos colorimétricos que emplean un sustrato cromogénico para la enzima.

La detección también puede lograrse mediante la comparación visual de la medida de la reacción enzimática de un sustrato contra estándares preparados de manera similar. También es posible marcar el anticuerpo con un compuesto fluorescente. Los ejemplos de compuestos de marcaje fluorescente incluyen isotiocianato de fluoresceína, rodamina, ficoeritrina, ficocianina, alofocianina, o-ftaldehído, Cy3, Cy5, Cy7, isotiocianato de tetrametilrodamina, ficoeritrina, alofocianinas, Texas Red y fluorescamina. El anticuerpo también puede marcarse de forma detectable usando metales emisores de fluorescencia, tales como  $^{152}\text{Eu}$ , u otros de la serie de los lantánidos. Otros compuestos metálicos que pueden conjugarse con los anticuerpos incluyen, entre otros, ferritina y oro coloidal, como en el caso de las microesferas superparamagnéticas coloidales. Estos metales se pueden unir al anticuerpo usando tales grupos quelantes de metales como el ácido dietilentriaminopentaacético (DTPA) o el ácido etilendiaminotetraacético (EDTA). El anticuerpo también se puede marcar de manera detectable al acoplarlo a un compuesto quimioluminiscente. Los ejemplos de compuestos de marcaje quimioluminiscente particularmente útiles son: luminol, isoluminol, éster de acridinio teromático, imidazol, sal de acridinio y éster de oxalato. Del mismo modo, se puede usar un compuesto bioluminiscente para marcar el anticuerpo. En un ejemplo, el anticuerpo está marcado con un compuesto de bioluminiscencia, como luciferina, luciferasa o aequorina. Los haptenos que pueden conjugarse con los anticuerpos incluyen, entre otros, biotina, digoxigenina, oxazalona y nitrofenol. Los compuestos radiactivos que se pueden conjugar o incorporar a los anticuerpos incluyen, pero no se limitan a, tecnecio 99m ( $^{99}\text{Tc}$ ),  $^{125}\text{I}$  y aminoácidos, incluidos los radionucleótidos, incluidos, entre otros,  $^{14}\text{C}$ ,  $^3\text{H}$  y  $^{35}\text{S}$ .

En general, los inmunoensayos para proteínas (por ejemplo, según se muestran en la Tabla 1) generalmente incluyen incubar una muestra biológica en presencia de anticuerpos y detectar el anticuerpo unido mediante cualquiera de una serie de técnicas bien conocidas en la técnica. En un ejemplo, la muestra biológica (como una que contiene melanocitos) puede ponerse en contacto e inmovilizarse en un soporte o vehículo de fase sólida como nitrocelulosa o una placa de pocillos múltiples u otro soporte sólido que sea capaz de inmovilizar células, partículas celulares o proteínas solubles. El soporte se puede lavar con tampones adecuados, lo cual es seguido por un tratamiento con el anticuerpo que se une específicamente a la proteína diana (como las de la Tabla 1). El soporte de fase sólida se puede lavar con el tampón por segunda vez para eliminar el anticuerpo no unido. Si el anticuerpo se marca directamente, la cantidad de marcador unido en el soporte sólido se puede detectar por medios convencionales. Si el anticuerpo no está marcado, se puede usar un segundo anticuerpo marcado, que detecta ese anticuerpo que se une específicamente a la proteína diana (como las de la Tabla 1).

Alternativamente, los anticuerpos se inmovilizan en un soporte sólido y después se ponen en contacto con proteínas aisladas de una muestra biológica, como una biopsia de tejido, en condiciones que permiten que el anticuerpo y la proteína se unan específicamente entre sí. El complejo anticuerpo:proteína resultante puede detectarse, por ejemplo, adicionando otro anticuerpo específico para la proteína (formando así un sándwich anticuerpo:proteína:anticuerpo). Si

el segundo anticuerpo adicionado está marcado, se puede detectar el complejo o, alternativamente, se puede usar un antigay secundario marcado que sea específico para el segundo anticuerpo adicionado.

Un soporte o vehículo en fase sólida incluye materiales capaces de unir una muestra, un antígeno o un anticuerpo.

- 5 Los soportes ejemplares incluyen vidrio, poliestireno, polipropileno, polietileno, dextrano, nailon, amilasas, celulosas naturales o modificadas, poliácridamidas, gabros y magnetita. La naturaleza del vehículo puede ser soluble hasta cierto punto o insoluble. El material de soporte puede tener prácticamente cualquier configuración estructural posible siempre que la molécula acoplada sea capaz de unirse a su diana (como un anticuerpo o proteína). Por consiguiente, la configuración del soporte puede ser esférica, como en una microesfera, o cilíndricos, como el interior de la superficie de un tubo de ensayo o la superficie externa de una varilla. Alternativamente, la superficie puede ser plana, como una lámina o una tira de prueba.

- En algunos ejemplos, se utiliza un ensayo inmunoabsorbente ligado a enzimas (ELISA) para detectar las proteínas diana. El ensayo ELISA puede usarse para detectar la presencia de una proteína en una muestra, por ejemplo, mediante el uso de un anticuerpo que se une específicamente a una proteína diana (como las de la Tabla 1). En algunos ejemplos, el anticuerpo se puede unir a una enzima, por ejemplo, directamente conjugado o a través de un anticuerpo secundario, y se adiciona una sustancia que la enzima puede convertir en una señal detectable. Por lo tanto, en el caso del ELISA de fluorescencia, cuando la luz de la longitud de onda adecuada se ilumina sobre la muestra, cualquier complejo antígeno:anticuerpo fluorescerá para que la cantidad de antígeno en la muestra pueda inferirse a través de la magnitud de la fluorescencia. La proteína (como las proteínas extraídas o aisladas de una muestra que contiene melanocitos) generalmente se inmoviliza en un soporte sólido (por ejemplo, una placa de microtitulación de poliestireno) de manera no específica (por ejemplo, mediante la adsorción a la superficie) o específicamente (por ejemplo, mediante la captura por otro anticuerpo específico para el mismo antígeno, en un ELISA "sándwich"). Entre cada etapa, la placa se lava generalmente con una solución de detergente suave, como solución salina tamponada con fosfato con o sin NP40 o TWEEN para eliminar cualquier proteína o anticuerpo que no esté específicamente unido. Después de la etapa final de lavado, la placa se desarrolla adicionando un sustrato enzimático para producir una señal visible, que indica la cantidad de proteína en la muestra.

- La detección también se puede lograr mediante el uso de cualquier otro inmunoensayo de entre una variedad de los mismos. Por ejemplo, al marcar radiactivamente los anticuerpos o fragmentos de anticuerpos, es posible detectar el gen de huellas dactilares de péptidos mutantes o de tipo salvaje mediante el uso de un radioinmunoensayo (RIA). En otro ejemplo, se puede usar un ensayo inmunoradiométrico tándem sensible y específico (véase Shen y Tai, J. Biol. Chem., 261:25, 11585-11591, 1986). El isótopo radiactivo puede detectarse por medios tales como el uso de un contador gamma o un contador de centelleo o mediante autorradiografía.

- En un ejemplo, se utiliza un procedimiento espectrométrico para detectar o cuantificar un nivel de expresión de una proteína diana (como las de la Tabla 1). Los procedimientos espectrométricos ejemplares incluyen la espectrometría de masas, la espectrometría de resonancia magnética nuclear y combinaciones de las mismas. En un ejemplo, la espectrometría de masas se usa para detectar la presencia de una proteína diana (como las de la Tabla 1) en una muestra biológica.

- Una proteína diana (como las de la Tabla 1) también puede detectarse mediante ensayos de espectrometría de masas junto con ensayos de inmunoafinidad, el uso de mapeo de masas por desorción/ionización mediante láser asistida por matriz/acoplada a un tiempo de vuelo (MALDI-TOF) y la cromatografía líquida/tiempo de vuelo de cuadrupolo/ionización por electropulverización/espectrometría de masa tándem (LC/Q-TOF-ESI-MS/MS) para etiquetar secuencias de proteínas separadas mediante electroforesis en gel de poliácridamida bidimensional (2D-PAGE).

- Los procedimientos espectroscópicos de masa cuantitativa, como SELDI, se pueden utilizar para analizar la expresión de proteínas en una muestra. En un ejemplo, la espectrometría de masas de tiempo de vuelo de desorción-ionización láser de superficie mejorada (SELDI-TOF) se utiliza para detectar la expresión de proteínas, por ejemplo, utilizando ProteinChip (CIPHERGEN Biosystems, Palo Alto, CA). Tales procedimientos son bien conocidos en la técnica. SELDI es un procedimiento de fase sólida para la desorción en el que el analito se presenta a la corriente de energía en una superficie que mejora la captura o la desorción del analito.

- Brevemente, una versión de SELDI utiliza una superficie cromatográfica con una química que captura selectivamente analitos de interés, como los de cualquiera de las Tablas 2-8. Las superficies cromatográficas pueden estar compuestas de químicas hidrofóbicas, hidrofílicas, de intercambio iónico, de metal inmovilizado u otras. Por ejemplo, la química de la superficie puede incluir funcionalidades de unión basadas en medios de inmovilización de analitos covalentes o no covalentes dependientes de oxígeno, carbono, azufre y/o nitrógeno. Las superficies activadas se usan para inmovilizar de forma covalente moléculas específicas de "cebo", como anticuerpos, receptores u oligonucleótidos que se usan a menudo para estudios de interacción biomolecular, como las interacciones proteína-proteína y proteína-

ADN.

La química de la superficie permite que los analitos unidos se retengan y los materiales no unidos se eliminen mediante lavado. Posteriormente, los analitos unidos a la superficie (como los de la Tabla 1) pueden ser desorbidos y analizados por cualquiera de varios medios, por ejemplo, utilizando espectrometría de masas. Cuando el analito se ioniza en el proceso de desorción, como en la espectrometría de masas por desorción/ionización por láser, el detector puede ser un detector de iones. Los espectrómetros de masas generalmente incluyen medios para determinar el tiempo de vuelo de los iones desorbidos. Esta información se convierte en masa. Sin embargo, no es necesario determinar la masa de iones desorbidos para resolverlos y detectarlos: el hecho de que los analitos ionizados golpeen el detector en diferentes momentos proporciona su detección y resolución. Alternativamente, el analito puede marcarse de forma detectable (por ejemplo, con un fluoróforo o un isótopo radiactivo). En estos casos, el detector puede ser un detector de fluorescencia o radiactividad. Se puede implementar una serie de medios de detección en serie para interrogar completamente los componentes del analito y la función asociada con las moléculas retenidas en cada ubicación de la matriz.

Por lo tanto, en un ejemplo particular, la superficie cromatográfica incluye anticuerpos que se unen específicamente a una proteína diana (como las de la Tabla 1). En otros ejemplos, la superficie cromatográfica consiste esencialmente en, o consiste en, anticuerpos que se unen específicamente a una proteína diana (como las de la Tabla 1). En algunos ejemplos, la superficie cromatográfica incluye anticuerpos que se unen a otras moléculas, como las proteínas de normalización (*por ejemplo*, las de la Tabla 1).

En otro ejemplo, los anticuerpos se inmovilizan en la superficie usando un soporte de unión a Fc bacteriano. La superficie cromatográfica se incuba con una muestra, como una muestra de un nevo. Los antígenos presentes en la muestra pueden reconocer los anticuerpos en la superficie cromatográfica. Las proteínas no unidas y los compuestos de interferencia espectrométrica de masas se lavan y las proteínas que se retienen en la superficie cromatográfica son analizadas y detectadas mediante SELDI-TOF. El perfil de MS de la muestra se puede comparar utilizando un mapeo diferencial de expresión de proteínas, por lo que los niveles de expresión relativos de proteínas a pesos moleculares específicos se comparan mediante una variedad de técnicas estadísticas y sistemas de software bioinformático.

Alternativamente, la cantidad de proteína diana se puede determinar usando procedimientos fluorescentes. Por ejemplo, puntos cuánticos (*por ejemplo*, los marcadores Qdots®) son útiles en una lista creciente de aplicaciones que incluyen la inmunohistoquímica, la citometría de flujo y los ensayos basados en placas y, por lo tanto, se pueden usar junto con esta descripción. Los nanocristales de puntos cuánticos tienen propiedades ópticas únicas que incluyen una señal extremadamente brillante para la sensibilidad y la cuantificación; y una alta fotoestabilidad para imágenes y análisis. Se necesita una sola fuente de excitación y un intervalo creciente de conjugados (*por ejemplo*, conjugados de anticuerpos) los hace útiles en una amplia gama de aplicaciones. La emisión de puntos cuánticos es estrecha y simétrica, lo que significa que se minimiza la superposición con otros colores, lo que resulta en un sangrado mínimo en los canales de detección adyacentes y la diafonía atenuada, a pesar de que se pueden usar muchos más colores simultáneamente. Por ejemplo, la IHC se puede realizar con anticuerpos secundarios conjugados con puntos cuánticos o puntos cuánticos conjugados con estreptavidina en combinación con anticuerpos primarios o secundarios marcados con biotina.

### **C. Medidas de control de ensayo opcionales**

Opcionalmente, los ensayos utilizados para detectar los productos de expresión génica (por ejemplo, los ácidos nucleicos (como ARNm, ARNnc) o las proteínas) incluyen elementos de control de proceso positivos y negativos utilizados para evaluar el rendimiento del ensayo.

Un control positivo puede ser cualquier elemento conocido, preferiblemente de naturaleza similar a la diana (*por ejemplo*, una diana ARN, después un control positivo de ARN (o ADNc)), que puede incluirse en un ensayo (o muestra) y detectarse en paralelo con la o las dianas y que no interfiere (*por ejemplo*, en reacción cruzada) con tales dianas de detección. En un ejemplo, el control positivo es una transcripción *in vitro* (IVT) que se ejecuta en paralelo como una muestra separada o se "adiciona" a cada muestra en una cantidad conocida. Los agentes de unión específicos de IVT (*por ejemplo*, las sondas de oligonucleótidos, tales como las NPP o NPPF) y, si corresponde, también se incluyen agentes de detección específicos de IVT en cada ensayo para asegurar un resultado positivo para tales transcripciones *in vitro*.

En algunas situaciones, las señales anómalas pueden resultar de problemas inesperados relacionados con el proceso que de otro modo no se controlan, *por ejemplo*, mediante el análisis de normalizadores; por lo tanto, en algunas realizaciones, es útil incluir uno o más elementos de control de proceso independientes de la muestra para indicar un ensayo exitoso o fallido en cualquier muestra, independientemente de la estabilidad, integridad o nivel de entrada de

la muestra. Las realizaciones de procedimiento en las que se detectan productos de expresión génica de ácido nucleico pueden incluir una concentración conocida de una muestra de ARN (por ejemplo, la transcripción *in vitro* de ARN o IVT) en cada ensayo. Tal elemento de control (*por ejemplo*, la IVT) se puede medir en cada ensayo y actuar como un control de calidad del proceso de ensayo.

5 Algunas realizaciones de procedimientos descritas que involucran productos de expresión génica de ARN pueden, aunque no necesariamente, incluir una muestra procesada en paralelo que contenga ARN de referencia humana universal. Si dicha muestra de ARN universal incluye todos o algunos de los ARN destinados a la detección mediante el ensayo aplicable, se puede esperar una señal positiva para dichos ARN incluidos, que pueden servir como un (u  
10 otro) control de calidad del proceso de ensayo.

Los elementos de control de proceso negativos pueden incluir agentes de unión específicos de analito (por ejemplo, oligonucleótidos o anticuerpos) diseñados o seleccionados para detectar un producto génico que no se espera expresar en la muestra de prueba. Por ejemplo, un agente de unión específico de analito que reconoce cualquier  
15 producto de expresión génica en el transcriptoma o proteoma humano puede incluirse en un ensayo multiplexado (como una sonda de oligonucleótidos o un anticuerpo específico para un ARN o proteína de planta o insecto o nematodo, respectivamente), donde los productos de expresión génica humana son las dianas deseados). Este elemento de control negativo no debe generar una señal en el ensayo aplicable. Cualquier señal por encima del fondo para dicho elemento de control de proceso negativo es un indicador de falla del ensayo. En un ejemplo, el control  
20 negativo es ANT.

#### IV. Datos de expresión génica

Los datos de expresión génica "contienen las claves para abordar problemas fundamentales relacionados con la  
25 prevención y cura de enfermedades, mecanismos de evolución biológica y descubrimiento de fármacos" (Lu y Han, Information Systems, 28:243-268 (2003)). En algunos ejemplos, destilar la información de dichos datos es tan simple como hacer una determinación cualitativa a partir de la presencia, ausencia o cantidad cualitativa (*por ejemplo*, alta, media o baja) de uno o más productos genéticos detectados. En otros ejemplos, los datos de expresión génica sin procesar pueden procesarse previamente (*por ejemplo*, ser sustraídos del fondo, transformados en su logaritmo y/o  
30 corregidos), normalizarse y/o aplicarse en algoritmos de clasificación. Sin embargo, dicho procesamiento previo de datos es opcional.

Muchas variables biológicas (*por ejemplo*, los datos de expresión génica) no cumplen con los supuestos de las pruebas estadísticas paramétricas, *por ejemplo*, tales variables no se distribuyen normalmente, las variaciones no son  
35 homogéneas, o ambos (Durbin y col., Bioinformática, 18:S105 (2002)). En algunos casos, la transformación de los datos hará que se ajusten mejor a los supuestos estadísticos. En algunas realizaciones de procedimientos, la transformación de datos útil puede incluir (i) la transformación de logaritmos, que consiste en tomar el logaritmo de cada observación, *por ejemplo*, los logaritmos de base 10, los logaritmos de base 2, los logaritmos de base e (también conocidos como logaritmos naturales); la selección de logaritmos no hace ninguna diferencia porque dichos logaritmos  
40 difieren en un factor constante; o la transformación estabilizadora de varianza, *por ejemplo*, según lo descrito por Durbin (véase arriba). En ejemplos específicos, los valores de expresión sin procesar para cada biomarcador detectado en dicho procedimiento (*por ejemplo*, al menos dos biomarcadores en la Tabla 1) son logaritmos (*por ejemplo*, log 2 o log 10) transformados. Cuando se calcula la expresión de recuento/lectura, la transformación puede incluir lecturas totales para cada muestra, este valor se escala por millón de lecturas: las lecturas a nivel de gen se  
45 evalúan proporcionalmente en comparación con las lecturas totales para una muestra y el log<sub>2</sub> transformado resultando

$$\log_2\left(\frac{r_{gi}+0.5}{R_i+1} \times 1,000,000\right)$$

en recuentos log-base-2 por millón (CPM), donde  $r_{gi}$  es el número de lecturas de secuencia para cada sonda (g) y muestra (i), (escalado para evitar recuentos cero), ajustado por el número de lecturas mapeadas (recuento de biblioteca) para cada muestra ( $R_i$ , escalado mediante una constante 1 para garantizar que la relación proporcional de lectura a tamaño de biblioteca sea mayor que cero).  
50

#### V. Implementación del procedimiento

Los procedimientos, como los que involucran clasificadores, descritos en esta invención, pueden implementarse de numerosas maneras. A continuación, se describen varias realizaciones no limitativas representativas. En algunas  
55 realizaciones de procedimientos, los datos de expresión génica se ingresan (por ejemplo, manual o automáticamente) en un ordenador u otro dispositivo, máquina o aparato para la aplicación de los diversos algoritmos descritos en esta invención, lo cual es particularmente ventajoso cuando hay un gran número de puntos de datos de expresión génica recolectado y procesado. Por lo tanto, en algunos ejemplos, los datos de expresión génica o el valor de probabilidad obtenidos usando los procedimientos descritos se almacenan en la memoria de del ordenador. Otras realizaciones

involucran el uso de una infraestructura de comunicaciones, por ejemplo, Internet. Diversas formas de hardware, software, firmware, procesadores o una combinación de los mismos son útiles para implementar clasificaciones específicas y realizaciones de procedimientos. El software se puede implementar como un programa de aplicación incorporado de manera tangible en un dispositivo de almacenamiento de programas, o diferentes partes del software implementado en el entorno informático del usuario (*por ejemplo*, como un applet) y en el entorno informático del revisor, donde el revisor puede estar ubicado en un sitio remoto asociado (*por ejemplo*, en las instalaciones de un proveedor de servicios).

Por ejemplo, durante o después de la entrada de datos por parte del usuario, se pueden realizar partes del procesamiento de datos en el entorno informático del lado del usuario. Por ejemplo, el entorno informático del lado del usuario puede programarse para proporcionar códigos de prueba definidos para denotar una "puntuación" de probabilidad, donde la puntuación se transmite como respuestas procesadas o parcialmente procesadas al entorno informático del revisor en forma de código de prueba para su posterior ejecución de uno o más algoritmos para proporcionar resultados y/o generar un informe en el entorno informático del revisor. La puntuación puede ser una puntuación numérica (representativa de un valor numérico) o no numérica, representativa de un valor numérico o un intervalo de valores numéricos (*por ejemplo*, "A", que representa de un 90 a un 95% de probabilidad de un resultado).

El programa de aplicación para ejecutar los algoritmos descritos en esta invención puede cargarse y ejecutarse en una máquina que comprenda cualquier arquitectura adecuada. Preferentemente, la máquina se implementa en una plataforma informática que tiene hardware como una o más unidades centrales de procesamiento (CPU), una memoria de acceso aleatorio (RAM) y una o más interfaces de entrada/salida. La plataforma informática también incluye un sistema operativo y un código de microinstrucción. Los diversos procesos y funciones descritos en esta invención pueden formar parte del código de microinstrucción o parte del programa de aplicación (o una combinación de los mismos) que se ejecuta mediante el sistema operativo. Además, se pueden conectar otros dispositivos periféricos a la plataforma informática, como un dispositivo adicional de almacenamiento de datos y un dispositivo de impresión.

Como sistema informático, el sistema generalmente incluye una unidad de procesador. La unidad de procesador funciona para recibir información, que puede incluir datos de prueba (*por ejemplo*, el nivel de un gen de respuesta, el nivel de un producto o genes de referencia; el nivel normalizado de un gen de respuesta; y también puede incluir otros datos, como los correspondientes al paciente. Esta información recibida puede almacenarse al menos temporalmente en una base de datos, y los datos pueden analizarse para generar un informe como se describió anteriormente.

Parte o la totalidad de los datos de entrada y salida también se pueden enviar electrónicamente; ciertos datos de salida (*por ejemplo*, los informes) se pueden enviar por vía electrónica o telefónica (*por ejemplo*, por facsímil, utilizando dispositivos como la devolución al remitente por fax). Los dispositivos receptores de salida ejemplares pueden incluir un elemento de visualización, una impresora, un dispositivo facsímil y similares. Las formas electrónicas de transmisión y/o visualización pueden incluir el correo electrónico, la televisión interactiva y similares. En una realización, todos o una parte de los datos de entrada y/o todos o una parte de los datos de salida (*por ejemplo*, generalmente al menos el informe final) se mantienen en un servidor web para su acceso, preferiblemente un acceso confidencial, con navegadores típicos. Se puede acceder a los datos o enviarlos a profesionales de la salud como se desee. Los datos de entrada y salida, incluyendo todo o parte del informe final, se pueden utilizar para completar el registro médico de un paciente que puede existir en una base de datos confidencial en el centro de salud. En algunos ejemplos, el procedimiento incluye generar un informe. En algunos ejemplos, el informe incluye un icono que indica la clasificación de una muestra, como "ABC" para el ABC de DLBCL o "GCB" para el GCB de DLBCL.

Un sistema para usar en los procedimientos descritos aquí generalmente incluye al menos un procesador de ordenador (*por ejemplo*, donde el procedimiento se lleva a cabo en su totalidad en un solo sitio) o al menos dos procesadores informáticos en red (*por ejemplo*, donde los datos deben ser ingresados por un usuario (también denominado "cliente" en esta invención) y transmitidos a un sitio remoto a un segundo procesador de ordenador para su análisis, donde el primer y el segundo procesador de ordenador están conectados por una red, por ejemplo, a través de una intranet o Internet). El sistema también puede incluir uno o más componentes de usuario para la entrada; y uno o varios componentes de revisión para la revisión de datos, de los informes generados y la intervención manual. Los componentes adicionales del sistema pueden incluir uno o varios componentes del servidor; y una o más bases de datos para almacenar datos (*por ejemplo*, como en una base de datos de elementos de informe, *por ejemplo*, elementos de informe interpretativos, o una base de datos relacional (RDB) que puede incluir la entrada de datos por parte del usuario y la salida de datos). Los procesadores del ordenador pueden ser procesadores que normalmente se encuentran en los ordenadores personales de escritorio (*por ejemplo*, de IBM, Dell, Macintosh), en los ordenadores portátiles, ordenadores principales, mini ordenadores, tabletas u otros dispositivos informáticos.

La arquitectura de cliente/servidor en red se puede seleccionar como se desee, y puede ser, por ejemplo, un modelo clásico de servidor de cliente de dos o tres niveles. Un sistema de gestión de bases de datos relacionales (RDMS), como parte de un componente del servidor de aplicaciones o como un componente separado (máquina RDB)

proporciona la interfaz a la base de datos.

En un ejemplo, la arquitectura se proporciona como una arquitectura cliente/servidor centrada en la base de datos, en la cual la aplicación cliente generalmente solicita servicios del servidor de aplicaciones que realiza solicitudes a la base de datos (o al servidor de la base de datos) para llenar el informe con los diversos elementos de informes según sea necesario, en particular los elementos del informe interpretativo, especialmente el texto de interpretación y las alertas. El o los servidores (*por ejemplo*, ya sea como parte de la máquina del servidor de aplicaciones o como una máquina de base de datos relacional/RDB separada), responden a las solicitudes del cliente.

10 Los componentes del cliente de entrada pueden ser ordenadores personales completos e independientes que ofrecen una gama completa de potencia y características para ejecutar aplicaciones. El componente del cliente generalmente opera bajo cualquier sistema operativo deseado e incluye un elemento de comunicación (*por ejemplo*, un módem u otro hardware para conectarse a una red), uno o más dispositivos de entrada (*por ejemplo*, un teclado, mouse, teclado u otro dispositivo utilizado para transferir información o comandos), un elemento de almacenamiento (*por ejemplo*, un disco duro u otro medio de almacenamiento legible por ordenador, grabable por ordenador) y un elemento de visualización (*por ejemplo*, un monitor, televisor, LCD, LED u otro dispositivo de visualización que transmite información al usuario). El usuario ingresa comandos de entrada en el procesador del ordenador a través de un dispositivo de entrada. En general, la interfaz de usuario es una interfaz gráfica de usuario (GUI) escrita para aplicaciones de navegador web.

20 Los componentes del servidor pueden ser un ordenador personal, una mini ordenador o un ordenador principal y ofrecen administración de datos, intercambio de información entre clientes, administración de red y seguridad. La aplicación y las bases de datos utilizadas pueden estar en el mismo servidor o en servidores diferentes.

25 Se contemplan otras disposiciones informáticas para el cliente y los servidores, incluyendo el procesamiento en una sola máquina, como un ordenador principal, una colección de máquinas u otra configuración adecuada. En general, las máquinas cliente y servidor trabajan juntas para lograr el procesamiento de la presente descripción.

30 Cuando se usan, las bases de datos generalmente están conectadas al componente del servidor de bases de datos y pueden ser cualquier dispositivo que contenga datos. Por ejemplo, la base de datos puede ser cualquier dispositivo de almacenamiento magnético u óptico para un ordenador (*por ejemplo*, un CDROM, disco duro interno o unidad de cinta). La base de datos se puede ubicar remotamente respecto del componente del servidor (con acceso a través de una red, módem, etc.) o localmente respecto del componente del servidor.

35 Cuando se utiliza en el sistema y los procedimientos, la base de datos puede ser una base de datos relacional que está organizada y a la que se accede según las relaciones entre los elementos de datos. La base de datos relacional generalmente se compone de una pluralidad de tablas (entidades). Las filas de una tabla representan registros (colecciones de información sobre elementos separados) y las columnas representan campos (atributos particulares de un registro). En su concepción más simple, la base de datos relacional es una colección de entradas de datos que 40 "se relacionan" entre sí a través de al menos un campo común.

Se pueden usar estaciones de trabajo adicionales equipadas con ordenadores e impresoras en el punto de servicio para ingresar datos y, en algunas realizaciones, generar informes adecuados, si así se lo desea. Los ordenadores pueden tener un acceso directo (*por ejemplo*, en el escritorio) para iniciar la aplicación a fin de facilitar el inicio de la 45 entrada de datos, la transmisión, el análisis, la recepción de informes, etc., como se desee.

### **1. Medios de almacenamiento legibles por ordenador**

La presente descripción también contempla un medio de almacenamiento legible por ordenador (*por ejemplo*, un 50 CDROM, un dispositivo de memoria, una tarjeta de memoria flash, un disquete, etc.) habiendo almacenado en él un programa que, cuando se ejecuta en un entorno informático, proporciona la implementación de algoritmos para llevar a cabo todos o una parte de los resultados de una evaluación de probabilidad de respuesta como se describe en esta invención. Cuando el medio legible por ordenador contiene un programa completo para llevar a cabo los procedimientos descritos en esta invención, el programa incluye instrucciones del programa para recopilar, analizar y 55 generar resultados, y generalmente incluye dispositivos de código legibles por ordenador para interactuar con un usuario como se describe en esta invención, procesando esos datos en conjunto con la información analítica y generando medios impresos o electrónicos únicos para ese usuario.

Donde el medio de almacenamiento proporciona un programa que proporciona la implementación de una parte de los 60 procedimientos descritos en esta invención (*por ejemplo*, el aspecto del lado del usuario de los procedimientos (*por ejemplo*, la entrada de datos, las capacidades de recepción de informes, etc.), el programa proporciona la transmisión de los datos ingresados por el usuario (*por ejemplo*, a través de Internet, a través de una intranet, etc.) a un entorno

informático en un sitio remoto. El procesamiento o finalización del procesamiento de los datos se puede llevar a cabo en el sitio remoto para generar un informe. Después de revisar el informe y completar cualquier intervención manual necesaria, para proporcionar un informe completo, el informe completo puede transmitirse al usuario como documento electrónico o documento impreso (*por ejemplo*, por fax o como informe en papel enviado por correo). El medio de almacenamiento que contiene un programa como se describe en esta invención se puede empaquetar con instrucciones (*por ejemplo*, para la instalación del programa, uso, *etc.*) registradas en un sustrato adecuado o una dirección web donde se pueden obtener dichas instrucciones. El medio de almacenamiento legible por ordenador también se puede proporcionar en combinación con uno o más reactivos para llevar a cabo la evaluación de probabilidad de respuesta (*por ejemplo*, cebadores, sondas, matrices u otros componentes del kit).

10 **2. Resultado:**

En algunos casos, una vez que se determina una puntuación para una muestra particular (paciente), se puede mostrar y/o transmitir una indicación de esa puntuación a un médico u otro cuidador. Por ejemplo, los resultados de la prueba se proporcionan a un usuario (como un médico u otro trabajador de atención médica, personal de laboratorio o paciente) en un resultado perceptible que proporciona información sobre los resultados de la prueba. En algunos ejemplos, el resultado es un resultado en papel (por ejemplo, un resultado escrito o impreso), una visualización en una pantalla, un resultado gráfico (por ejemplo, un gráfico, un cuadro u otro diagrama) o un resultado audible. Por lo tanto, el resultado puede incluir un informe generado.

Por ejemplo, el resultado puede ser textual (opcionalmente, con una puntuación correspondiente). Por ejemplo, los resultados textuales pueden ser "ABC" o similar, o "GCB" o similar, o "indeterminante" o similar. Dicho resultado textual puede usarse, por ejemplo, para proporcionar un diagnóstico del subtipo GCB o ABC de DLBCL, o simplemente puede usarse para ayudar a un médico a distinguir los subtipos ABC y GCB.

En otros ejemplos, el resultado es un valor numérico (*por ejemplo*, un resultado cuantitativo), como un valor de probabilidad pronosticado. En ejemplos adicionales, el resultado es una representación gráfica, por ejemplo, un gráfico que indica la probabilidad predicha. En un ejemplo particular, el resultado (como un resultado gráfico) muestra o proporciona un valor de corte o nivel que caracteriza la muestra analizada como ABC o GCB. En otros ejemplos, el resultado es un icono, como un "ABC", si la muestra se clasifica como ABC, "GCB", si la muestra se clasifica como GCB, "U" o "?", si la muestra se clasifica como sin clasificar (*por ejemplo*, cuando no es consistente con ABC o GCB). En algunos ejemplos, el resultado se comunica al usuario, por ejemplo, proporcionando un resultado a través de medios físicos, audibles o electrónicos (por ejemplo, por correo, teléfono, transmisión por fax, correo electrónico o comunicación a un registro médico electrónico).

En algunos ejemplos, el resultado se acompaña de pautas para interpretar los datos, por ejemplo, límites numéricos u otros límites que indican que el DLBCL es ABC o GCB. Las pautas no necesitan especificar si un DLBCL es ABC o GCB, aunque puede incluir dicho diagnóstico. Los indicios en el resultado pueden, por ejemplo, incluir intervalos normales o anormales o un punto límite, que el receptor del resultado puede usar para interpretar los resultados, por ejemplo, para llegar a un diagnóstico o plan de tratamiento. En otros ejemplos, el resultado puede proporcionar un régimen terapéutico recomendado. En algunos ejemplos, la prueba puede incluir la determinación de otra información clínica (como determinar la cantidad de uno o más biomarcadores de DLBCL adicionales en la muestra).

45 **VI. Uso clínico de conjuntos de genes y resultados del clasificador**

Los conjuntos o clasificadores de genes descritos pueden resultar en la caracterización de la muestra de DLBCL (*por ejemplo*, diagnosticada) como un subtipo GCB o un subtipo no GCB (como ABC) o indeterminado. Cada uno de estos (y otros posibles) resultados es útil para el profesional clínico capacitado. Algunos usos clínicos representativos se describen con más detalle a continuación. Se puede proporcionar un diagnóstico o pronóstico a un sujeto (*por ejemplo*, paciente o médico) por cualquier medio adecuado, incluidos, entre otros, medios físicos, audibles o electrónicos (por ejemplo, por correo, teléfono, transmisión por fax, correo electrónico o comunicación a un registro médico electrónico).

**A. Indicaciones de diagnóstico**

55 Un diagnóstico informa a un sujeto (*por ejemplo*, paciente) qué enfermedad o afección tiene o puede tener. Como se describe más particularmente a lo largo de esta descripción, se puede proporcionar cualquier resultado de cualquier procedimiento descrito que clasifique un subtipo de DLBCL, *por ejemplo*, para un sujeto o profesional de la salud, como diagnóstico. Por lo tanto, un diagnóstico incluye determinar si un DLBCL es un subtipo GCB o un subtipo no GCB (como ABC) o indeterminado.

60 **B. Indicaciones de pronóstico**

El pronóstico es el resultado de salud probable para un sujeto cuya muestra recibió un resultado de prueba particular (*por ejemplo*, subtipo GCB o subtipo no GCB (como ABC) de DLBCL). Un mal pronóstico significa que la perspectiva a largo plazo para el sujeto no es buena, *por ejemplo*, la supervivencia a 1, 2, 3 o 5 años es del 50% o menos (por ejemplo, 40, 30, 25, 20, 15, 10, 5, 2, 1% o menos). Por otro lado, un buen pronóstico significa que las perspectivas a largo plazo para el sujeto son de regulares a buenas, *por ejemplo*, la supervivencia a 1, 2, 3 o 5 años es superior al 30, 40, 50, 60, 70, 75, 80 o 90%.

Se ha demostrado que el subtipo no GCB de DLBCL, como el subtipo ABC de DLBCL, tiene un pronóstico que es peor que el subtipo GCB de DLBCL. En consecuencia, un hallazgo del subtipo no GCB de DLBCL, como el subtipo ABC de DLBCL por cualquiera de los procedimientos descritos, puede usarse para predecir un pronóstico comparativamente malo para un sujeto del que se toma la muestra de prueba. Por el contrario, un hallazgo del subtipo GCB de DLBCL por cualquiera de los procedimientos descritos puede usarse para predecir un pronóstico comparativamente bueno para el sujeto correspondiente.

### 15 C. Indicaciones terapéuticas (predictivas)

Los procedimientos descritos pueden incluir además seleccionar (o no) sujetos para el tratamiento, para el subtipo GCB o no GCB de DLBCL, como el subtipo ABC de DLBCL, si su muestra correspondiente está subtipada de ese modo. En la sección a continuación, se proporcionan procedimientos de tratamiento ejemplares. Por lo tanto, en un ejemplo, si se determina que la muestra es un subtipo GCB de DLBCL, el sujeto de quien se obtuvo la muestra se trata con (1) una quimioterapia con ciclofosfamida, doxorubicina, vincristina y prednisona o prednisolona (CHOP), (2) quimioterapia con rituximab más CHOP (R-CHOP) o (3) etopósido más R-CHOP (R-EPOCH). En otro ejemplo, si se determina que la muestra es del subtipo ABC de DLBCL, es menos probable que el sujeto del que se obtuvo la muestra responda a la terapia CHOP o R-CHOP, y dicho sujeto puede seleccionarse para el tratamiento con una o más terapias alternativas (como las descritas en esta invención).

En algunos ejemplos, los procedimientos descritos también incluyen uno o más de los siguientes, según el diagnóstico del paciente: a) prescribir un régimen de tratamiento para el sujeto, si el diagnóstico determinado del sujeto es el subtipo GCB de DLBCL (como el tratamiento con uno o más agentes de quimioterapia o terapia sistémica, como CHOP, R-CHOP o R-EPOCH); b) prescribir un régimen de tratamiento para el sujeto, si el diagnóstico determinado del sujeto es el subtipo ABC de DLBCL; o c) no prescribir un régimen de tratamiento (como no prescribir CHOP, R-CHOP o R-EPOCH) para el sujeto, si el diagnóstico determinado del sujeto es el subtipo ABC de DLBCL.

### VII. Procedimientos de tratamiento de un sujeto con DLBCL

En algunos ejemplos, siguiendo la clasificación de un subtipo de DLBCL, se puede seleccionar un sujeto para el tratamiento con una o más terapias, por ejemplo, si la muestra del sujeto se califica como GCB o no GCB (*por ejemplo*, ABC) de DLBCL. Además, dicha una o más terapias pueden administrarse posteriormente al sujeto seleccionado.

En algunos ejemplos, al sujeto se le administra una quimioterapia con ciclofosfamida, doxorubicina, vincristina y prednisona o prednisolona (CHOP) o una quimioterapia con rituximab más CHOP (R-CHOP). En un ejemplo, al sujeto también se le administra etopósido para R-CHOP, lo que da como resultado una combinación de fármacos llamada R-EPOCH.

En otros ejemplos, al sujeto se le administra uno o más anticuerpos monoclonales, tales como los siguientes: anti-CD20 (por ejemplo, rituximab, ofatumumab, ocrelizumab, GA101, tositumomab o ibritumomab tiuxetán), anti-CD22 (por ejemplo, epratuzumab), anti-CD 19 (por ejemplo, SAR3419 o blinatumumab), anti-CD30 (por ejemplo, brentuximab vedotina), anti-CD40 (por ejemplo, dacetuzumab o lucatumumab) o anti-CD70 (por ejemplo, SGN-75). En algunos ejemplos, un anticuerpo monoclonal se conjuga con una toxina, radiomarcador o fármaco. En ejemplos adicionales, al sujeto se le administran uno o más inhibidores de Bcl-2 (por ejemplo, ABT-737, obatoclaix u oblimersen sódico), inhibidores de mTOR (por ejemplo, rapamicina, everolimus, temsirolimus o deforolimus (ridaforolimus)), Syk inhibidores (por ejemplo, fostamatinib (tamatatinib)), inhibidores del proteasoma (por ejemplo, bortezomib), inhibidores de la proteína quinasa C (por ejemplo, enzastaurina o sotrastaurina), agentes inmunomoduladores (por ejemplo, lenalidomida y pomalidomida), inhibidores de la aurora A quinasa (por ejemplo, alisertib) o inhibidores de histona desacetilasa (por ejemplo, vorinostat, romidepsin, panobinostat, belinostat, mocetinostat, abexinostat o entinostat).

En algunos ejemplos, a un sujeto se le administra CHOP, R-CHOP o R-EPOCH como terapia de primera línea, y se le administran terapias adicionales, como las descritas anteriormente (y, en algunos ejemplos, un trasplante de células madre) tras una recaída o cuando el DLBCL es refractario al tratamiento CHOP, R-CHOP o R-EPOCH. En otros ejemplos, R-CHOP o R-EPOCH y una o más de las terapias adicionales enumeradas anteriormente se pueden administrar a un sujeto con DLBCL. El rituximab y una o más de las terapias adicionales mencionadas anteriormente también se pueden administrar al sujeto.



- Las terapias adicionales que pueden usarse para tratar a un sujeto con DLBCL GCB o no GCB incluyen la radioterapia y/o uno o más tratamientos de quimioterapia. En algunos ejemplos, los agentes de quimioterapia incluyen, entre otros, agentes alquilantes, como mostazas nitrogenadas (por ejemplo, clorambucilo, clormetina, ciclofosfamida, ifosfamida y melfalan), nitrosoureas (por ejemplo, carmustina, fotemustina, lomustina y estreptozocina), compuestos de platino (por ejemplo, carboplatino, cisplatino, oxaliplatino y BBR3464), busulfano, dacarbazina, mecloretamina, procarbazona, temozolomida, tiotepa y uramustina; antimetabolitos, como el ácido fólico (por ejemplo, metotrexato, pemetrexed y raltitrexed), purina (por ejemplo, cladribina, clofarabina, fludarabina, mercaptopurina y tioguanina), pirimidina (por ejemplo, capecitabina), citarabina, fluoritaracilo, fluoritaracilo, fluoritaracilo, fluoritaracilo, gemouracilo y fluorouracilo;
- 5  
10 alcaloides vegetales, tales como podófilo (por ejemplo, etopósido y tenipósido), taxano (por ejemplo, docetaxel y paclitaxel), vinca (por ejemplo, vinblastina, vincristina, vindesina y vinorelbina); antibióticos citotóxicos/antitumorales, como los miembros de la familia de la antraciclina (por ejemplo, daunorrubicina, doxorubicina, epirubicina, idarrubicina, mitoxantrona y valrubicina), bleomicina, hidroxurea y mitomicina; inhibidores de topoisomerasa, tales como topotecán e irinotecán; anticuerpos monoclonales, como alemtuzumab, bevacizumab, cetuximab, gemtuzumab, rituximab, panitumumab y trastuzumab; fotosensibilizadores, tales como el ácido aminolevulínico, metil aminolevulinato, porfímero de sodio y verteporfina; y otros agentes, como alitretinoína, altretamina, amsacrina, anagrelida, trióxido de arsénico, asparaginasa, bexaroteno, bortezomib, celecoxib, denileukin diftitox, erlotinib, estramustina, gefitinib, hidroxycarbamida, imatinib, petatata, petatata, petatata, petaco, petaco y pentosina.
- 15  
20 Un sujeto puede ser tratado con una combinación de terapias, incluyendo una o más de las descritas anteriormente. En un ejemplo específico, a un sujeto se le puede administrar un inhibidor de la proteína quinasa C en combinación con rituximab, gemcitabina y oxaliplatino (R-GEMOX). En otro ejemplo específico, a un sujeto se le puede administrar un inhibidor de proteasoma (como bortezomib) en combinación con R-CHOP o en combinación con etopósido, vincristina y doxorubicina más ciclofosfamida y prednisona (EPOCH). En un ejemplo específico, a un sujeto se le
- 25 administra R-ACVBP (rituximab más doxorubicina, ciclofosfamida, vindesina, bleomicina y prednisona).

Una persona con conocimiento ordinario en la materia puede seleccionar las terapias, las dosis y los horarios de dosificación adecuados para un sujeto. La dosis requerida variará de un sujeto a otro dependiendo de la especie, la edad, el peso y el estado general del sujeto, los agentes terapéuticos particulares que se utilizan y los modos de

30 administración.

### Matrices para detectar expresiones

- Aquí se describen conjuntos que se pueden usar para detectar o medir la expresión génica (como la expresión de dos o más de los biomarcadores en la Tabla 1), por ejemplo, para determinar el subtipo de DLBCL en un sujeto con DLBCL. En algunas realizaciones, las matrices descritas también se pueden usar para detectar la expresión de uno o más genes de control (*por ejemplo*, los biomarcadores de normalización). En ejemplos particulares, la superficie de la matriz incluye una placa, una microesfera o celda de flujo.
- 40 En algunos ejemplos, una matriz puede incluir una superficie sólida que incluye regiones específicamente discretas o ubicaciones direccionables, cada región tiene al menos un oligonucleótido inmovilizado capaz de hibridarse directa o indirectamente a un biomarcador en la Tabla 1 (por lo tanto, la matriz puede tener una pluralidad de regiones, cada una específica para un biomarcador en la Tabla 1, por ejemplo). Por ejemplo, la matriz puede incluir regiones específicamente discretas, cada región tiene al menos una o al menos dos sondas de captura inmovilizadas capaces
- 45 de hibridarse directa o indirectamente específicamente con al menos 2, 3, 5, 10 o todos los 16 biomarcadores de la Tabla 1 (como sondas de captura capaces de hibridarse directa o indirectamente específicamente con cada uno de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8, y TYMS). En algunos ejemplos, los oligonucleótidos pueden identificarse por su posición en la matriz. En otros ejemplos, los oligonucleótidos pueden identificarse por un marcador detectable (ya sea directa o indirectamente
- 50 asociado con el oligonucleótido). En otro ejemplo, una matriz puede incluir regiones específicamente discretas, donde cada región tiene al menos uno o dos oligonucleótidos inmovilizados.
- En algunos ejemplos, la matriz incluye dos o más sondas de captura capaces de hibridarse directa o indirectamente específicamente con una NPP de un gen de firma de DLBCL o un gen de control. Por ejemplo, la matriz puede incluir
- 55 oligonucleótidos que incluyen o consisten en oligonucleótidos que son complementarios de al menos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o los 16 genes de firma de DLBCL en la Tabla 1 (como los oligonucleótidos que incluyen o consisten en oligonucleótidos que son complementarios a cada uno de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, por ejemplo, los que se muestran en las SEQ ID NO: 1-16). En algunos ejemplos, la matriz (como una que incluye sondas de captura
- 60 capaces de hibridarse directa o indirectamente específicamente con cada uno de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS) incluye además oligonucleótidos que incluyen o consisten en oligonucleótidos que son complementarios a al menos 1, 2, 3,

4, 5, 6, 7, 8 o 9 genes de control.

- En ejemplos particulares, la matriz incluye oligonucleótidos (por ejemplo, enlazadores de programación) que incluyen secuencias de ácido nucleico diana, tales como ácidos nucleicos que incluyen todas o una parte de las regiones de ácidos nucleicos diana descritas en la Tabla 1 (o un oligonucleótido complementario a tales secuencias). En algunos ejemplos, los enlazadores de programación son oligonucleótidos bifuncionales que incluyen una porción que es complementaria a un oligonucleótido (como un anclaje) asociado con un sustrato y una porción que es complementaria a al menos una porción de una NPP.
- 10 En algunos ejemplos, la matriz puede incluir una superficie que tiene regiones espacialmente discretas (como pocillos en una superficie de pocillos múltiples, microesferas o canales en una celda de flujo), cada región incluye un anclaje estable (*por ejemplo*, covalentemente) unido a la superficie y un oligonucleótido (*por ejemplo*, un enlazador de programación), donde el enlazador de programación es un enlazador heterobifuncional que tiene una primera porción complementaria al anclaje y una segunda porción complementaria a una NPP, donde la NPP es complementaria a un ácido nucleico diana (como dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; como todos los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS). En algunos ejemplos, la matriz incluye o consiste esencialmente en enlazadores bifuncionales, donde la primera porción es complementaria a un anclaje y la segunda porción es complementaria a una NPP. En algunos ejemplos, la matriz incluye además conectores bifuncionales, en los que la primera porción es complementaria a un anclaje y la segunda porción es complementaria a una NPP que es complementaria a un marcador de control. En un ejemplo adicional, la matriz también incluye además conectores bifuncionales, donde la primera porción es complementaria a un anclaje y la segunda porción es complementaria a una NPP que es complementaria a un control negativo (tal como ANT).
- 25 Dichas matrices se han unido al anclaje híbrido al menos a un segmento del conector bifuncional que no es complementario a la NPP. Dichas matrices pueden incluir además: (1) la sonda de anclaje híbrida a la primera porción del enlazador de programación; (2) NPP híbridadas a la segunda porción del enlazador de programación; (3) enlazadores de detección bifuncionales que tienen una primera porción híbrida a las NPP y una segunda porción híbrida con una sonda de detección; (4) una sonda de detección; (5) un marcador (tal como avidina HRP); o combinaciones de los mismos.

### Kits

- En esta invención, también se describen kits que pueden usarse para detectar o medir los niveles de genes de firma de DLBCL (como la expresión de dos o más de los biomarcadores en la Tabla 1), por ejemplo, para caracterizar una muestra como ABC, GCB o sin clasificar DLBCL como se analiza en esta invención. En algunas realizaciones, los kits descritos también se pueden usar para detectar la expresión de uno o más marcadores de control (*por ejemplo*, ANT). En ejemplos particulares, el kit incluye una o más de las matrices proporcionadas en esta invención, como una matriz que incluye sondas de captura capaces de hibridarse directa o indirectamente específicamente con cada uno de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS.

- En algunos ejemplos, los kits incluyen sondas (*por ejemplo*, NPP o NPPF) y/o cebadores para la detección de la expresión de ácido nucleico, como la detección de dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; o, por ejemplo, todos los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. Por ejemplo, los kits pueden incluir al menos dos NPPF diferentes y las CFS correspondientes, donde las al menos dos NPPF diferentes son específicas para dos o más (como al menos 3, 4, 5, 6, 7, 8, 9 o los 10) de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3 o TNFRSF8; o específicas para dos o más (como al menos 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 o los 16) de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS. Por ejemplo, los kits pueden incluir NPP (o NPPF) específicas para cada uno de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; como las NPP que comprenden o consisten en las secuencias mostradas en las SEQ ID NO; 1-16.

En algunos ejemplos, los kits incluyen además uno o más biomarcadores de control. En algunos ejemplos, los kits incluyen además sondas y/o cebadores para detectar uno o más biomarcadores de control (*por ejemplo*, ANT).

- 60 Las sondas incluidas en el kit pueden incluir una o más NPP o NPPF que son complementarias o se hibridan específicamente con las secuencias diana descritas en esta invención, como sondas que incluyen una secuencia que se muestra en las SEC ID NO: 1-16. En algunos ejemplos, los kits pueden incluir una o más sondas de ácido nucleico

necesarias para construir una matriz para detectar los biomarcadores descritos aquí.

En algunos ejemplos, los kits incluyen anticuerpos que se unen específicamente a uno o más biomarcadores enumerados en la Tabla 1, y opcionalmente anticuerpos que se unen específicamente a uno o más biomarcadores de control. Por ejemplo, los kits pueden incluir anticuerpos (*por ejemplo*, anticuerpos monoclonales, anticuerpos policlonales y/o fragmentos de anticuerpos) para la detección de la expresión de proteínas, como la detección de dos o más de los siguientes: CD47, CD86, IL16, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS; como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS (como todos los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS) y, en algunos ejemplos, uno o más biomarcadores de control.

En algunos ejemplos, el kit incluye adicionalmente uno o más de: un recipiente que contiene un tampón (tal como un tampón de lisis, un tampón de hibridación, un tampón de lavado y/o un tampón de secuenciación); un recipiente que contiene una nucleasa específica para ácidos nucleicos de una sola cadena (como una nucleasa S1); un recipiente que contiene enlazadores de programación de ácido nucleico; un recipiente que contiene etanol; un recipiente que contiene aceite de desnaturalización; un recipiente que contiene proteinasa K; un contenedor que contiene microesferas que pueden unirse específicamente a amplicones de al menos dos NPPF diferentes; y uno o más recipientes que contienen reactivos para PCR (como un tampón de PCR, polimerasa, dNTP). En algunos ejemplos, los kits incluyen además muestras de control, tales como cantidades particulares de proteínas o ácidos nucleicos CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS

En un ejemplo, un kit incluye NPP o NPPF que son complementarias o se hibridan específicamente con CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, como las NPP que comprenden o consisten en las secuencias mostradas en las SEQ ID NO: 1-16 Si el kit incluye NPPF, el kit puede incluir además CFS correspondientes. En algunos ejemplos, dicho kit incluye además un tampón de lisis, una solución de nucleasa (como una que incluye S1), un tampón de terminación (*por ejemplo*, uno que pueda neutralizar o desactivar la nucleasa, tal como uno que pueda aumentar el pH por encima de 6, por ejemplo, a alrededor de 9 a 12). En algunos ejemplos, dicho kit incluye además proteinasa K; aceite de desnaturalización; y opcionalmente placas de micromatrices y puntas de pipeta. En algunos ejemplos, dicho kit incluye además un tampón de ligadura (*por ejemplo*, uno que incluye ligasa). En algunos ejemplos, el kit también incluye cebadores específicos para una porción de la NPP o NPPF, o CFS, como un cebador que puede adicionar una etiqueta de experimento y/o un adaptador de secuenciación en el extremo 3' o 5' o en ambos extremos de la NPP o NPPF.

En un ejemplo, el kit incluye un sistema informático que implementa un clasificador para subtipos de DLBCL, como el software o el medio legible por ordenador descrito en esta invención que recibe múltiples valores que miden o detectan la expresión de ARNm de los genes de la firma de DLBCL (como dos o más de los siguientes: CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS), procesa los valores múltiples usando la Ecuación 1, calcula una probabilidad pronosticada y clasifica una muestra en un marco que indica el subtipo de DLBCL de la muestra (*por ejemplo*, ABC, GCB, o sin clasificar).

En un ejemplo, el kit incluye un gráfico o una tabla que muestra los valores esperados o los intervalos de valores (como los valores de probabilidad predictiva) esperados en los subtipos ABC y/o GCB de DLBCL y, opcionalmente, las muestras sin clasificar.

Los kits pueden incluir además componentes adicionales tales como materiales de instrucción y reactivos adicionales, por ejemplo, reactivos de detección, tales como un sistema de detección basado en enzimas (por ejemplo, reactivos de detección que incluyen peroxidasa de rábano picante o fosfatasa alcalina y un sustrato adecuado). Los kits también pueden incluir componentes adicionales para facilitar la aplicación particular para la cual está diseñado el kit (por ejemplo, placas de microtitulación, puntas de pipeta y similares). Los kits también pueden incluir anticuerpos secundarios (por ejemplo, anticuerpos que se unen específicamente a los anticuerpos primarios que se unen específicamente a las proteínas en la Tabla 1, o anticuerpos que se unen específicamente a los anticuerpos primarios que se unen específicamente a una proteína de control), o un medio para marcar anticuerpos. En un ejemplo, el kit incluye además ácidos nucleicos de control y/o proteínas de control. Dichos kits y contenidos adecuados son bien conocidos por los expertos en la materia. Los materiales de instrucción se pueden encontrar en forma escrita, en forma electrónica (como un disquete de ordenador o disco compacto) o pueden ser visuales (como archivos de vídeo).

La descripción se ilustra adicionalmente mediante los siguientes ejemplos no limitantes.

## 60 Ejemplo 1

### Identificación de la lista de genes de firma de DLBCL

Este ejemplo describe procedimientos estadísticos utilizados para construir un clasificador para el subtipo de DLBCL en dos clases, ABC y GCB. Además, se agregó a la clasificación un tercer subtipo que incluye una región de incertidumbre, denotada como sin clasificar. Todos los análisis de datos se realizaron utilizando el lenguaje de programación R versión 3.2.2 y el paquete Bioconductor versión 3.1 (versiones de 64 bits).

Los procedimientos de diseño estadístico general se desarrollaron utilizando una estrategia similar, como se describe en Hastie, Tibshirani, y Freidman (The Elements of Statistical Learning: Data Mining, Inference and Prediction Springer Science and Business Media, Inc., 2009) cuando se utiliza un único conjunto de datos para el desarrollo y la prueba del rendimiento estimado. Para desarrollar el clasificador de subtipo, el modelo y evaluar este último, se utilizaron dos etapas generales. El desarrollo del modelo incluye seleccionar de modelo/variable, ajustar los parámetros y evaluar de manera interna las pruebas del modelo final que define los subtipos. La evaluación incluye la estimación del error de predicción. Ambas etapas se realizaron utilizando una muestra de conveniencia de un solo sitio obtenida prospectivamente de espirales FFPE de DLBCL con el estado de DLBCL asociado siendo evaluado por un médico experto.

### **Control de calidad**

Se aplicaron tres métricas de control de calidad (QC) a cada muestra después de completar la secuenciación. Se identificaron modos de falla para cada una de las métricas de QC mutuamente excluyentes evaluadas en orden secuencial. La primera de las métricas de QC es la tasa de alineación (basada en Bowtie) de las lecturas mapeadas para las secuencias de la sonda. La expectativa es que la alineación general sea  $\geq 85\%$ . Si esto no se cumple, una muestra tiene un modo de falla de QC1. La segunda métrica de QC aplicada a todos los casos que cumplen con la métrica de calidad QC1 es que la muestra tiene una expresión diferencial. La expresión se establece evaluando las diferencias de expresión entre las sondas en el primer cuartil de expresión (Q1) y el tercer cuartil de expresión (Q3); si la relación entre la expresión promedio es menor a 2,0 ( $Q1/Q3 < 2,0$ ), entonces la muestra no aprobará la segunda métrica de QC (QC2) (a lo que se hace referencia como falla S1 o una pared de números). La tercera métrica de QC (QC3) aplicada a las muestras/casos que pasaban QC1 y QC2 involucraba los controles negativos: sondas ANT (ANT1, ANT2, ANT3). Las métricas de control del proceso estadístico (SPC) para controles negativos se establecieron en un estudio de línea de base que secuenciaba 90 pocillos de una única réplica técnica del lisado de la línea celular SUDHL6. El estudio de caracterización SPC estableció una tolerancia superior para los controles negativos, esta tolerancia se aplicó a todas las muestras nuevas para demostrar el rendimiento dentro de un rendimiento esperado. El rendimiento esperado se estableció evaluando la variabilidad promedio de la señal ANT. La expectativa es que la ANT promedio en comparación con los niveles medios no debería fluctuar más que un límite predeterminado. El límite era 7,75 en el  $\log_2(\text{CPM})$ ; cualquier muestra con un ANT de promedio de muestra mayor que este valor no aprobaría la tercera métrica de QC (QC3). Todas las muestras con cualquiera de estos modos de falla podrían haber comprometido los perfiles del clasificador.

### **Muestras**

Un total de 257 casos tenían suficiente tejido representativo para la secuenciación, pasaron los procedimientos de control de calidad de secuenciación descritos anteriormente y se determinó el estado de DLBCL mediante el perfil de expresión génica (GEP). Las muestras utilizadas en el análisis pasaron las métricas de control de calidad para los valores promedio de ANT (control negativo) y tuvieron alineamientos de sonda a secuencia  $\geq 85\%$ . El ARNseq es una metodología que cuenta el número de lecturas que se asignan a una secuencia específica. La tasa de alineación es una forma de cuantificar cuántos de los pares de bases se asignan correctamente a las copias de ARN generadas a partir de la PCR. Se espera que al menos el 85% de las lecturas disponibles se asignen dentro de una muestra. Menos de esta cantidad involucra que hay un producto sin mapear en la muestra. La FIG. 2 muestra el diagrama de flujo de datos para la selección de casos del total de casos disponibles.

Los lisados de tejido DLBCL se prepararon a partir de secciones de tejido FFPE. La sección de tejido se midió y después se raspó en un tubo eppendorf marcado usando una cuchilla de afeitar y evitando cualquier exceso de parafina en el portaobjetos. El área de la muestra se usó para calcular la cantidad de tampón de lisis adicionado:  $\text{área total (cm}^2\text{)} \times /28 \mu\text{l}/0,25 \text{ cm}^2$  de tejido o  $\text{área total (cm}^2\text{)} \times /25 \mu\text{l}/0,25 \text{ cm}^2$  de tejido. La muestra se suspendió en el tampón de lisis (tampón de SSC precalentado (50°C) que incluía formamida y SDS). Quinientos (500)  $\mu\text{l}$  de aceite mineral que contenía un tensioactivo (*por ejemplo*, Brij-97) ("Capa no acuosa") después se superpusieron sobre la suspensión de tejido, y esta reacción de lisis se incubó a 95°C durante 10 a 20 minutos. Después de enfriar brevemente la mezcla de reacción, se adicionó proteinasa K en una proporción de 1/20<sup>th</sup>. El volumen del tampón de lisis y la incubación continuaron a 50°C durante 60 a 120 minutos. La reacción de lisis puede usarse inmediatamente o congelarse y almacenarse a una temperatura de -70 a -80°C. Las reacciones de lisis congeladas se descongelaron a 50C durante 10 a 15 minutos antes de un uso posterior. La entrada de muestra fue de 0,02  $\text{cm}^2/28 \mu\text{l}$  para las muestras con entrada de muestra conocida.

**Ensayo qNPS**

Se colocaron veintiocho (28) µl de cada mezcla de reacción lisada en un pocillo de una placa de 96 pocillos y se recubrieron con 70 µl de capa no acuosa. A cada pocillo se añadieron 5 µl de una mezcla que contenía una sonda de protección de nucleasa con una secuencia flanqueante 5' y 3' (NPPF) y secuencias flanqueantes complementarias (CFS) adecuadas para cada secuencia flanqueante. Las secuencias de las NPPF incluían una secuencia de captura que se unía específicamente a la diana, así como secuencias flanqueantes 3' y 5'. La Tabla 2 muestra la porción de secuencia de captura de la NPPF para los 16 genes del clasificador de DLBCL. Por lo tanto, la secuencia diana es el complemento de la secuencia mostrada. Un (1) nM (un exceso) de NPPF complementaria para cada uno de la pluralidad de dianas de ARNm a detectar estaba presente en la mezcla. Cada placa de 96 pocillos incluía 30 (cada uno por triplicado) diferentes lisados de DLBCL y seis muestras de control (dos sin control de muestra y 4 lisados de línea celular). Las muestras de DLBCL se aleatorizaron en las placas de manera ciega ( *es decir*, mediante la anotación de muestra cegada).

15

Tabla 2. Porción de captura de la NPPF utilizada para detectar la secuencia diana

Objetivo	Secuencia de captura de la NPPF	SEQ NO:	ID
CD47	AAACTGGTTACCTAGAGGCCAGAGGCCCTAGGACCTGA AAGGCATCATTCC	1	
CD86	GTCTCCTCTTGGCATAACGGAGCAGAGCTGGAGTTACAG GGAGGCTATTCC	2	
ENTPD1	CCATGAGGAAGACATAGGTGGAGTGGGAGAGAGGTGTG GACAATGGTTGC	3	
FOXP1	CTCTTCCCGTATTGCGCTGGCTAAGTTGCCAGAGTGGG ATTTCCCATGG	4	
FUT8	CAGTTTGGCTGAAATGGTTTGAAGTGGTTCGAAAC TGAGTTTGGTC	5	
IL16	GTCCATCAGGCAGTGCCTTGATGATGTTCCAGGCTTCAA ACCGTGTGAGG	6	
ITPKB	GCGCCTCAAACATGCCACTTTCTGGTTCACCTGCACGT TCTGCAACTCG	7	
LRMP	GCTATTTCTACAGCGGTTAAAGTCCTGATATGTCAAGTT GCCTCTTTCTG	8	
MME	GGCCTTGCGGAAAGCATTTCTGGACTCCTTGTAGGTTCG GCTGAGGCTGC	9	
NF2	GAGAGAAATCTCTACAGGGTCGTAGTTCAAGGCAATTG CACATAAGAGGG	10	
PIM1	GCTCACCTTCTTCAGCAGGACCACTTCCATGGGCACTCG AGTGCCATTAG	11	
PTPRC	GGAACAATTTCTCCTCTGTTACCCTAAGAACAACCAC TTGCTAGCTGC	12	
REL	CAGAGGTGCCATTGAGGCATGATGTGACAATCCACTTG AGATGGGCCCAG	13	
STAT3	CAGCTTCAGGATGCTCCTGGCTCTCTGGCCGACAATACT	14	

Objetivo	Secuencia de captura de la NPPF	SEQ NO:	ID
	TTCCGAATGCC		
TNFRSF8	CTGGGACCAATGCTGTTCTCGGCAGTGCCCATCCTGCCA TCTGTTTGCTG	15	
TYMS	CTTCAGGCCCGTGATGTGCGCAATCATGTACGTGAGCAG GGCGTAGCTGG	16	

La placa de prueba de 96 pocillos se calentó a 85°C durante 10 minutos para desnaturalizar los ácidos nucleicos y, después, se dejó incubar a 52°C durante 16 horas para permitir la hibridación de la NPPF con sus respectivas dianas de ARNm y permitir la hibridación de las CFS a su respectiva diana de secuencia flanqueante.

Después de la etapa de hibridación, se añadieron 20 µl de exceso de nucleasa S1 (2,5 U/µl) en un tampón de acetato de sodio a la fase acuosa de cada pocillo. La reacción S1 continuó a 52°C durante 90 minutos para digerir el ARNm no unido y tanto las NPPF como las CFS no unidas.

10

Durante la etapa de digestión S1, se preparó una placa de parada de 96 pocillos adicionando 5 µl de solución que contenía Tris-HCL 1 M (pH 9,0) en cada pocillo correspondiente a las reacciones en la placa de prueba de 96 pocillos. Se transfirieron 40 µl de cada reacción en la placa de prueba de 96 pocillos a un pocillo correspondiente en la segunda placa de parada de 96 pocillos y se procedió a la superposición con 70 µl de la capa no acuosa. La placa de parada se incubó a 100°C durante 20 minutos y, después, se enfrió durante 30 minutos a temperatura ambiente.

15

Una vez completada la reacción de captura, la placa puede almacenarse a -20°C o usarse directamente para el etiquetado por PCR. Si las placas se almacenaron congeladas, se pueden descongelar a temperatura ambiente durante 30 a 40 minutos con agitación suave antes de usarlas en la PCR.

20

El etiquetado por PCR se realizó para la preparación de la biblioteca de secuenciación como se indica a continuación. Los reactivos de PCR, que incluyen 6 µl de agua, 15 µl de tampón GC NEB OneTaq® HotStart X2 Master Mix, 3 µl de cebador directo, 3 µl de cebador inverso y 3 µl de muestra se adicionaron a un tubo limpio. Como control negativo, algunas muestras simplemente recibieron agua. La PCR se realizó durante 20 ciclos

25

Para eliminar el exceso de cebadores después de la PCR, se agruparon 15 µl de cada muestra de PCR (excluyendo los controles negativos de PCR) en un tubo de microcentrifugadora. Se mezclaron 350 µl de las reacciones de PCR agrupadas con 135 µl de NaCl 5 M, 137 µl de solución de PEG al 40% y 35 µl de microesferas AMPure. El tubo se agitó vorticialmente y se incubó a los 5 minutos a temperatura ambiente. El tubo se colocó en un soporte magnético durante 5 minutos para permitir que se recojan las microesferas. El sobrenadante se descartó. A las microesferas, se adicionaron 30 µl de Tris-HCl 10 mM de pH 8,0 y 75 µl de AMPure XP. El tubo se agitó vorticialmente y se incubó durante 5 minutos a temperatura ambiente. El tubo se colocó en un soporte magnético durante 5 minutos para permitir que se recojan las microesferas. El sobrenadante se descartó. Las microesferas se lavaron dos veces con etanol al 80% y las microesferas se dejaron secar a 37°C durante 5 minutos. Las microesferas se suspendieron en 40 10 mM Tris-HCl de pH 8.0. Se adicionaron 30 µl de sobrenadante a un nuevo tubo que contenía el material limpio. La eliminación de los cebadores se confirmó ejecutando 10 µl de la biblioteca en gel de agarosa al 2%.

30

35

Los amplicones limpios se diluyeron en una relación de 1:10,000 usando el kit KapaQuant (Kapa Biosystems, Catálogo No. 10176100), siguiendo las instrucciones del fabricante.

40

Los amplicones limpios se secuenciaron usando un secuenciador MiSeq (Illumina). Se cargaron 30 pM de la biblioteca en un kit MiSeq 150bp V3, siguiendo las instrucciones del fabricante para la desnaturalización de la biblioteca y la configuración y carga del secuenciador.

#### 45 **Desarrollo del modelo**

Como primer etapa de la construcción del algoritmo clasificador, 257 muestras de DLBCL con control de calidad para las que estaban disponibles los perfiles de expresión génica (GEP) se dividieron aleatoriamente en dos conjuntos de datos no superpuestos que comprendían 172 muestras aleatorias para el desarrollo del clasificador; siendo que los 86 restantes se usaron como un conjunto de validación. Se evaluó la similitud de estas dos muestras a fin de garantizar la ausencia de un sesgo a raíz de la selección aleatoria. La Tabla 3 describe la distribución de estos dos subconjuntos

50

según los perfiles GEP.

Inicialmente, se determinó la expresión de 96 genes en todas las muestras de DLBCL usando el ensayo qNPS. La FIG. 3 muestra la naturaleza aleatoria de los posibles genes clasificadores, después de la eliminación del gen negativo, el positivo y el de mantenimiento en los dos conjuntos; el sesgo aparecería en las diferencias entre los recuentos de  $\log_2(\text{CPM})$  en lugar de la superposición que se muestra en la FIG. 3. Antes de realizar los análisis estadísticos, se eliminaron las sondas de limpieza y control (Tabla 4). La eliminación de estas sondas dio como resultado 83 genes considerados para el clasificador.

10 Tabla 3. Distribución de los perfiles GEP entre los conjuntos de desarrollo y validación.

	ABC	GCB
Desarrollo	86	86
Validación	36	49

Tabla 4. Sondas de limpieza y control por nombres de analizador

Sondas de limpieza	Sondas de control
ACTB	POS_CTRL_POS1
EEF1G	NFG_CTRL_ANT1
EIF4A1	NEG_CTRL_ANT2
GAPDH	NEG_CTRL_ANT3
RPL19	
RPL4	
RPL6	
RPS29	
TBP	

15 El desarrollo del clasificador se realizó utilizando un modelo lineal generalizado con una función de enlace logit, también conocida como regresión logística. La regresión logística es un modelo que se utiliza cuando los datos de respuesta (resultado) tienen dos categorías, también llamadas binarias; en este caso, las dos categorías son la clasificación/subtipado ABC y GCB. El modelo estadístico que se utiliza para estimar la probabilidad de una clasificación dada se escribe como se indica a continuación:

$$Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}} \quad \text{(Ecuación 1)}$$

20 donde  $\beta_0$  denota una intercepción y  $\beta$  son las estimaciones de parámetros asociadas con el vector de sondas,  $x^T$ . La muestra de desarrollo se utilizó para ajustar un modelo logístico descrito en (1) para obtener estimaciones de parámetros,  $\hat{\beta}_0$  y  $\hat{\beta}$ , siendo que a éstos se los denomina "ponderaciones". Estas ponderaciones se utilizan, junto con los valores de recuento de los niveles de expresión para cada gen, para determinar una probabilidad predicha,  $Pr(Y = y|x)$ , de ser clasificado como ABC o GCB en cualquier conjunto de datos nuevo (independiente del conjunto de datos de desarrollo). El modelo final puede tener tantos valores  $\hat{\beta}$  como sondas utilizadas para calcular las probabilidades. Las probabilidades estimadas se limitan a un valor entre 0 y 1 y, para clasificar a los pacientes futuros, se utiliza un punto límite óptimo para las probabilidades.

30 La porción de desarrollo del modelo incluyó dos etapas, la capacitación y las pruebas. La fase de prueba se refiere a las métricas internas que se utilizaron para identificar las sondas/dianas óptimas y obtener parámetros (ponderaciones) después de la selección de la diana/sonda y la reducción/escala del coeficiente. Esto es ligeramente diferente de alguna literatura estadística (Tibshirani, J. Royal Statistical Society (Series B) 58:267-288, 2006) que sugiere que se deben utilizar tres conjuntos de datos aleatorios: capacitación, pruebas y validación. Para los propósitos de este clasificador, la capacitación y las pruebas se realizaron en el mismo conjunto de desarrollo. La determinación de los puntos límites óptimos en las probabilidades de clasificación de pacientes como ABC, GCB y sin clasificar se realizó

en la fase de prueba. Todas estas etapas se realizaron utilizando el conjunto de casos de desarrollo.

La selección de las 83 dianas utilizadas para el cálculo probabilístico, junto con la optimización de las ponderaciones, se realizó utilizando un algoritmo llamado "el lazo". Estadísticamente, estos procedimientos se agrupan por el tipo de "compensación" que se aplica como parte de la optimización. Esta compensación reduce las ponderaciones para ajustar el sesgo potencial en función de tamaños de muestra pequeños. Se han publicado procedimientos que describen este enfoque (*por ejemplo*, véase Tibshirani, J. Royal Statistical Society (Serie B) 58:267-288, 2006). El algoritmo de lazo empleado para el desarrollo de este clasificador se implementa en el paquete "glmnet" en R y se describe en Friedman y col. (J. Statistical Software, 33(1):1-22, 2010). Las variables óptimas se determinaron utilizando diez veces la validación cruzada como un nuevo muestreo. Este es un procedimiento donde el conjunto de datos de desarrollo inicial se divide en submuestras de igual tamaño, correspondiente a 10. Se crearon 100 submuestras de este tipo. Estas submuestras se usaron para encontrar el modelo (*por ejemplo*, los parámetros de regresión ajustados) que mejor se clasifican en los dos subtipos (*es decir*, ABC y GCB), así como también para determinar los valores que se utilizarán como ponderaciones (*por ejemplo*, los coeficientes después de cualquier escalado y contracción que pueda ocurrir).

Usando estos procedimientos, el campo de 83 dianas se redujo a las 16 dianas de la Tabla 5. La Tabla 5 también enumera las "ponderaciones" para el conjunto óptimo de coeficientes ( $\beta$ ) (*es decir*, para cada diana). Aquellas dianas que tienen ponderaciones inferiores a 1 (*es decir*, negativa) están asociados con la clasificación ABC y aquellos con ponderaciones superiores a 1 están asociados con la clasificación GCB.

**Tabla 5. Coeficientes para el clasificador de DLBCL**

Gen	Coeficientes (ponderaciones)
CD47	-0,16
CD86	0,83
ENTPD1	-0,71
FOXP1	-0,56
FUT8	-0,30
IL16	-0,51
ITPKB	1,16
LRMP	0,16
MME	0,59
NF2	-0,70
PIM1	-0,13
PTPRC	0,07
REL	0,20
STAT3	-0,17
TNFRSF8	-0,01
TYMS	-0,25

## 25 Puntos límites de probabilidad pronosticados

Una vez obtenidas las dianas y ponderaciones óptimas, se obtuvo el punto límite para la probabilidad pronosticada. Cuando hay exactamente dos clasificaciones, la probabilidad que define las clases se obtiene maximizando la sensibilidad (probabilidad de especificar correctamente una muestra como ABC) y la especificidad (probabilidad de especificar correctamente una muestra como GCB) en todo el intervalo de las probabilidades predichas. El clasificador de DLBCL también incluyó un subconjunto de muestras "sin clasificar". Estas muestras se identifican como el conjunto de muestras que posiblemente no sean ni ABC ni GCB y, por lo tanto, pueden clasificarse en un subtipo separado (*por ejemplo*, para el cual no se puede realizar el perfil de ABC o GCB). El punto en el que la sensibilidad + especificidad (cuya combinación se denomina precisión) comienza a caer desde el punto máximo en el mismo punto en el que se determinaron los aumentos de la tasa de clasificación errónea. A esto se lo denominó límite inferior; el límite superior



es el punto alto simétrico asociado con este valor. La FIG. 4 representa este punto de decisión gráficamente; numéricamente, los puntos límites son 0,43 para el punto inferior y 0,57 para el punto superior. Los perfiles ABC/GCB resultantes son los siguientes: aquellos casos con una probabilidad estimada < 0,43 son ABC y aquellos con una probabilidad estimada > 0,57 son GCB, mientras que todo lo que se encuentre dentro de 0,43 y 0,57 inclusive, se cataloga como SIN CLASIFICAR.

**Ejemplo 2**

**Aplicación del clasificador de DLBCL a muestras independientes**

Este ejemplo demuestra que el clasificador de DLBCL desarrollado en el Ejemplo 1 permite la identificación de los subtipos GCB y ABC de DLBCL en muestras fijadas en formalina e incrustadas en parafina (FFPE).

Las pruebas finales del clasificador mostraron una baja propensión de "sobreajuste" (por ejemplo, de sobreoptimismo) y las ponderaciones resultantes se aplicaron al conjunto de validación para demostrar el rendimiento. Por consiguiente, el desarrollo del clasificador se completó en este conjunto de datos.

Las 86 muestras de DLBCL restantes que no se usaron para desarrollar el clasificador se usaron como un conjunto de validación. El análisis de qNPS de los 16 genes en la Tabla 5 se realizó como se describió anteriormente. Los datos resultantes se sometieron al clasificador (por ejemplo, el valor para cada gen obtenido usando qNPS se analizó con la Ecuación 1 y se determinó la probabilidad predicha). Como se menciona en esta invención, CD86, ITPKAB, LRMP, MME, PTPRC y REL son genes que tienden a estar asociados con el subtipo GCB de DLBCL y CD47, ENTPD1, FOXP1, FUT8, IL16, NF2, PIM1, STAT3, TNFRSF8 y TYMS son genes que tienden a estar asociados con el subtipo ABC de DLBCL. El valor de resultado para cada uno de los 16 genes obtenidos de qNPS se multiplicó por el valor de ponderación que se muestra en la Tabla 5, se determinó la suma adicionada de estos 16 genes y se determinó la probabilidad predicha, donde el valor de probabilidad pronosticado resultante determina si la muestra es ABC, GCB o sin clasificar.

Las métricas para el rendimiento del clasificador incluyen la tasa de clasificación errónea, definida como la proporción de casos para los cuales el clasificador pudo obtener un perfil de ABC/GCB que fue el mismo que los perfiles GEP. Las métricas adicionales incluyeron el valor predictivo positivo y negativo del valor predictivo (PPV y NPV, respectivamente; es decir, los casos que se clasifican correctamente como ABC o GCB). Cabe señalar que las muestras sin clasificar no se incluyen en la evaluación de estas métricas. La Tabla 6 muestra un resumen de la validación del clasificador aplicado al conjunto de validación, y la FIG. 5 representa la clasificación para cada muestra.

Tabla 6. Métricas de rendimiento para el clasificador en el conjunto de validación

	Proporción de casos
Clasificado correctamente como GCB (PPV)	0,95 (42/44)
Clasificado correctamente como ABC (NPV)	0,84 (32/38)
Clasificado erróneamente	0,10 (8/82)
Perfiles correctos en general	0,96 (74/82)

**Ejemplo 3**

**Ensayo de diagnóstico in vitro de DLBCL**

Este ejemplo describe procedimientos que se pueden usar para determinar el subtipo de células de origen (COO) de los tumores difusos de linfoma de células B grandes (DLBCL) del tejido FFPE utilizando el sistema HTG EdgeSeq con detección realizada en un secuenciador de próxima generación Illumina MiSeq. Los datos perfilados se evalúan mediante un algoritmo de clasificación y se determina que el tumor es de tipo células B activadas (ABC), tipo células B germinales (GCB) o subtipo sin clasificar.

El sistema HTG EdgeSeq combina una química del ensayo cuantitativo de protección de nucleasa (qNPA) con una plataforma de secuenciación de próxima generación (NGS) para permitir el análisis semicuantitativo de 96 genes dirigidos en un solo ensayo (Figura 1). La química qNPA no requiere la extracción de ácido nucleico de la mayoría de las muestras, lo que reduce significativamente los requisitos de entrada de muestras. También elimina los sesgos asociados con la extracción de ácido nucleico, la selección del tamaño, la síntesis de ADNc y la ligadura del adaptador. El procedimiento automatizado reduce los errores de operación y, por lo tanto, ayuda a garantizar que las bibliotecas

de secuenciación preparadas a partir de muestras preciosas, como el tejido FFPE, se puedan generar de forma reproducible. El uso del sistema HTG EdgeSeq permite a los laboratoristas pasar de muestras sin procesar a bibliotecas listas para secuenciar en aproximadamente 36 horas con menos de tres horas de tiempo práctico.

- 5 La tecnología qNPA (véase, por ejemplo, la Patente de los EE.UU. No. 8.741.564) permite la cuantificación de ARNm en un formato de amplificación de ARN sin extracción y sin formato de amplificación de ARN. Dos elementos primarios del proceso químico HTG EdgeSeq son la hibridación de ADN a ARN y la digestión de nucleasa S1. Las sondas funcionales de protección de nucleasa de ADN flanqueadas por secuencias de ala universales se hibridan con ARN diana, que pueden ser tanto solubles como reticulados en la matriz biológica. Las sondas universales de DNA
- 10 Wingmen se hibridan con las alas para evitar la digestión de la nucleasa S1. La nucleasa S1 se adiciona para digerir el exceso de sondas de ADN no hibridadas y ARN no hibridadas; las únicas sondas de protección de ADN funcionales restantes que permanecen completamente intactas son las hibridadas con ARN diana. Esto produce esencialmente una relación de 1:1 de sondas de detección de ADN con respecto al ARN inicialmente dirigido en la muestra. La desnaturalización por calor libera las sondas de protección del ADN: los dúplex de ARN. Las sondas de protección de
- 15 ADN liberadas están listas para su enumeración.

Las sondas de protección de ADN están marcadas con adaptadores de secuenciación y etiquetas en un termociclador. Las sondas de protección de ADN marcadas se concentran, se agrupan y están listas para secuenciar utilizando protocolos NGS estándar en la plataforma Illumina MiSeq. Los datos de expresión génica del instrumento NGS se

20 procesan e informan, por ejemplo, a través del software del sistema host HTG EdgeSeq.

Los oligonucleótidos (sondas) especialmente diseñados que se hibridan con secuencias complementarias en el ARNm diana se usan como un medio para medir los niveles de ARNm. Estos oligonucleótidos están diseñados para secuencias que se verifican a fin de garantizar la ausencia de una estructura secundaria potencial y se someten a

25 búsquedas BLAST contra otras secuencias humanas informadas con la intención de garantizar una falta de homología significativa o complementariedad con otros genes en el ensayo. Cada oligonucleótido es un 100-mer que comprende un "ala" de 25 mer en el extremo 5', un "ala" de 25 mer en el extremo 3' y una secuencia de 50 mer que es complementaria al ARNm diana.

- 30 El ensayo de células de origen HTG EdgeSeq de DLBCL incluye cuatro etapas principales. (1) Preparación de la muestra (etapa manual): Se adiciona un tampón de lisis de HTG para lisar y/o permeabilizar la muestra de tejido FFPE haciendo que el ARN esté disponible para unirse posteriormente a las Sondas de protección de nucleasa (NPP) correspondientes a una diana. La muestra lisada se transfiere a una placa estándar de microvaloración de 96 pocillos, denominada Placa de muestra. (2) Captura de dianas a través de la química de HTG EdgeSeq (etapa automatizada):
- 35 La captura de dianas se realiza mediante la química de HTG EdgeSeq. Brevemente, las sondas de protección de nucleasa (NPP) se adicionan a las muestras lisadas en la placa de muestra, en una cantidad en exceso, y se hibridan con el ARNm diana. Después, se adiciona la nucleasa S1 para digerir el ARNm no hibridado y las NPP en exceso, produciendo así una cantidad estequiométrica de dúplex de ARNm/NPP diana. Una vez que se completa la digestión de S1, la muestra procesada se transfiere a una nueva placa de microvaloración de 96 pocillos con fondo en V,
- 40 denominada Placa de detención, y la digestión de S1 se termina mediante la adición de una solución de terminación seguida de la desnaturalización por calor de la enzima S1. (3) Amplificación por PCR para la adición de adaptadores y códigos de barras (etapa manual): Cada muestra procesada de la placa de parada se usa como plantilla para configurar reacciones de PCR con cebadores especialmente diseñados, a los que se hace referencia como etiquetas. Estas etiquetas comparten secuencias comunes que son complementarias a las secuencias "ala" de extremo 5' y 3'
- 45 de las sondas y los adaptadores comunes necesarios para la generación de un clúster en una plataforma de secuenciación Illumina MiSeq. Además, cada etiqueta contiene un código de barras único que se utiliza para la identificación y multiplexación de muestras. Se pueden usar 12 etiquetas de cebador directo y 8 inversas, y por lo tanto se pueden analizar hasta 96 muestras simultáneamente en una plataforma de secuenciación Illumina. El ensayo de células de origen HTG EdgeSeq de DLBCL está configurado para analizar 8, 24 o 96 muestras simultáneamente en
- 50 el instrumento Illumina MiSeq. Una vez finalizada la amplificación mediante PCR, se realiza un procedimiento de limpieza para eliminar las etiquetas no incorporadas de los productos de PCR (a lo que se hace referencia como biblioteca). Después, se realiza la limpieza mediante PCR con reacciones de PCR individuales. (4) Cuantificación de bibliotecas y secuenciación (etapa semiautomática): La secuencia se realiza en la plataforma Illumina MiSeq utilizando el kit MiSeq v3. Las concentraciones de las bibliotecas de muestras se equilibrarán y ajustarán para garantizar la
- 55 generación adecuada de grupos. Los datos de secuenciación sobre la expresión de ARNm de genes diana se importan al software de análisis HTG EdgeSeq, donde el subtipo COO de DLBCL se puede informar en formato tabular al laboratorio.

Procesamiento de muestras: Para las muestras FFPE, las secciones probadas deben tener un grosor de 4 a 6 micras

60 y deben montarse en un portaobjetos de microscopio de vidrio. El o las áreas en la sección a ser probada deben determinarse como DLBCL. Para cada muestra, se recomienda firmemente un portaobjetos teñido con H&E o una imagen escaneada con el área de interés claramente marcada para guiar la macrodissección del área de interés del

portaobjetos. Por lo general, se requiere una sección sin manchas. Si la porción tumoral contenida en la sección es extremadamente pequeña, se pueden raspar múltiples secciones y combinarlas en un solo tubo para la prueba.

Al medir el tamaño del tumor, se debe excluir cualquier tejido necrótico o no tumoral normal de la sección de tejido.

- 5 En circunstancias donde la exclusión no es práctica, no se debe incluir tejido no tumoral a la hora de calcular el tamaño. Cuando la sección de tejido proporciona más área de superficie que la entrada de muestra máxima recomendada, se debe raspar la porción tumoral relevante completa y diluir la concentración de trabajo adecuada usando un tampón de lisis. Si se analizan tejidos extremadamente pequeños, se pueden combinar múltiples secciones en un solo tubo para lograr la cantidad deseada de entrada de tejido y procesarla como una sola muestra. El aceite de desnaturalización
- 10 puede formar un precipitado blanco cuando se almacena a temperatura ambiente. Caliente el aceite de desnaturalización a 50°C durante 10 minutos y mezcle brevemente la botella para disolver el precipitado. Las muestras lisadas que no se utilizarán de inmediato pueden congelarse a -80°C durante un máximo de una (1) semana en tubos de microcentrifugadora. Las muestras no deben congelarse en la placa de muestra de 96 pocillos.
- 15 El procesamiento de tejido FFPE se puede realizar de la siguiente manera. El área o áreas adecuadas para la prueba se identifican con tinción H&E, IHC u otros medios adecuados; marque esta área en la parte posterior del portaobjetos con un marcador indeleble. Mida y calcule el área de superficie total (en mm<sup>2</sup>) abarcada por el área marcada. Excluya las áreas de tejido que son de baja celularidad, tejido no diana, necrótico o que podría comprometer los resultados de la prueba. Las cantidades de entrada aceptables para el ensayo de célula de origen HTG EdgeSeq de DLBCL son
- 20 1,56 - 12,5 mm<sup>2</sup> de tejido por prueba. En el caso de muestras extremadamente pequeñas (por ejemplo, biopsias con núcleo de aguja), se pueden combinar múltiples secciones de tejido en el mismo tubo para lograr la cantidad de muestra requerida. Descongele el tampón de lisis y el aceite de desnaturalización a 50°C durante 30 minutos. La Proteinasa K debe mantenerse a -20°C hasta que esté lista para su uso. Comenzando con una sección sin manchar, use el portaobjetos marcado como guía y raspe el área adecuada de tejido del portaobjetos sin manchar con una
- 25 cuchilla de afeitar. Coloque el tejido en un tubo de microcentrifugadora libre de ARNse debidamente marcado (se puede usar una pipeta para ayudar a extraerlo de la cuchilla de afeitar y colocarlo en el tubo de microfuga). Centrifugue brevemente (durante ~20 segundos) para tirar del tejido FFPE al fondo del tubo. Agregue 35 µL de tampón de lisis HTG al tubo por área diana de tejido. Por ejemplo:

30 Tabla 7. Ejemplos de cantidades de tejido y tampón de lisis

Cantidad de tejido	Volumen de tampón de lisis (µl)
1,56 - 12,5 mm <sup>2</sup>	35
12,6 - 25 mm <sup>2</sup>	70
25 - 50 mm <sup>2</sup>	140
50 - 100 mm <sup>2</sup>	280

- Se pueden procesar secciones extremadamente grandes utilizando los múltiples volúmenes de tampón de lisis enumerados anteriormente. Agregue 500 µL de aceite de desnaturalización a cada tubo y tape este último. No agite
- 35 vorticialmente. Caliente las muestras a 95°C durante 15 minutos y deje que las muestras se enfríen durante 10 minutos a temperatura ambiente. Agregue Proteinasa K a la capa inferior (color rojo) de cada tubo en una proporción de 1/20 del volumen del tampón de lisis. Por ejemplo:

Tabla 8. Ejemplos de cantidades de tampón de lisis y proteinasa K

Volumen de tampón de lisis (µl)	Proteinasa K (µl)
35	1,75
70	3,5
140	7
280	14

- 40 Mezcle la Proteinasa K pipeteando la solución seis veces. Incube las muestras lisadas a 50°C durante 3 horas mezclando en un agitador orbital o cada 30 minutos con una pipeta. Las muestras pueden procesarse o almacenarse congeladas a una temperatura de -70 a -80°C por hasta 4 semanas.

- Inicio de una ejecución en el procesador HTG EdgeSeq: Elimine el paquete de reactivos del ensayo HTG EdgeSeq
- 45 del paquete y descongélelo a temperatura ambiente durante 80 minutos. No agite ni invierta la bandeja de reactivos, los reactivos se mezclarán en el instrumento antes de su uso. Limpie el paquete de reactivos del ensayo HTG EdgeSeq

con lejía al 10% y después con alcohol isopropílico al 70%. Si las muestras lisadas se congelan, descongélelas a 50°C durante 45 minutos y mezcle con la pipeta la fase acuosa de cada muestra antes de usar. Centrifugue brevemente las muestras para devolver todo el líquido al fondo del tubo. Pipetee 35 µL de cada muestra en el pocillo adecuado de la placa de muestra. Con el escáner de código de barras del sistema HTG EdgeSeq, escanee la placa de muestra del paquete de placas HTG EdgeSeq. Después se inicia el procesador HTG EdgeSeq, el cual se ejecutará durante aproximadamente 20 horas. Después del procesamiento, elimine la placa de parada y analice.

Procesamiento de la biblioteca: Las mezclas maestras de PCR se pueden preparar, por ejemplo, usando el tampón GC NEB OneTaq®HotStart 2X Master Mix, agua de grado molecular y etiquetas de secuenciación. Esta mezcla, junto con un cebador inverso, se adiciona a una porción de las muestras adecuadas en la placa de parada, y la PCR se realiza de la siguiente manera: 95°C durante 4 minutos, después 20 ciclos de: 95°C durante 15 segundos, 56°C durante 45 segundos y 68°C durante 45 segundos, después 68°C durante 10 minutos y mantener a 4°C. La reacción resultante puede procesarse inmediatamente o almacenarse a -20°C. La biblioteca se puede agrupar y limpiar como se indica a continuación. Transfiera 15 µL de cada producto de PCR a una nueva placa de PCR de 96 pocillos previamente marcada. Agregue 37.5 µL de microesferas AMPure XP por pocillo (precalentado a temperatura ambiente y mezcle bien mediante agitación vorticial durante 1 minuto). Mezcle bien e incube a temperatura ambiente (RT) durante 5 minutos. Coloque la muestra en el soporte magnético y espere 3 minutos para que se acumulen las microesferas. Se debe formar una bolita visible de microesferas. Elimine el sobrenadante y no perturbe la bolita de microesferas. Para evitar la contaminación con microesferas, se pueden dejar de 5 a 10 µL de líquido residual si el gránulo de microesferas está suelto. Agregue 200 µL de etanol al 80% recién preparado y deje reposar durante 30 segundos a 1 minuto. Elimine el etanol de cada pocillo. Repita el lavado con etanol para un total de dos lavados. Elimine la placa del soporte del imán y deje secar a temperatura ambiente durante 5 minutos. Agregue 40 µL de Tris-HCl 10 mM, pH 8,0 a cada pocillo y mezcle la pipeta para volver a suspender las microesferas. Deje reposar 5 minutos a temperatura ambiente. Coloque la placa en el soporte del imán durante 2 minutos para recoger las microesferas. Transfiera 30 µL de cada sobrenadante de la biblioteca a un pocillo nuevo.

Las bibliotecas de muestras individuales creadas pueden cuantificarse y la concentración de la biblioteca individual puede equilibrarse antes de la secuenciación, por ejemplo, utilizando el sistema de cuantificación KAPA de la siguiente manera. Realice una dilución en serie con una relación de 1:100 (por ejemplo, por triplicado). Agregue 297 µL de Tris 10 mM/Tween al 0,05% a cada pocillo de la dilución en una relación de 1:100. Agregue 3 µL de biblioteca limpia a los pocillos correspondientes en la placa de pocillos profundos. Mezcle las muestras. Realice una segunda dilución en una relación de 1:100 para lograr un factor de dilución final de 1:10.000. Agregue 297 µL de Tris 10 mM/Tween al 0,05% a cada pocillo de la dilución en una relación de 1:10.000. Agregue 3 µL de la muestra de la placa "Dilución de 1:100" a los pocillos correspondientes en la placa "Dilución 1:10.000". Mezcle bien. Si se requiere una dilución de 1:100.000, realice una dilución de 1:10 adicional. Descongele el kit de cuantificación de la biblioteca KAPA a temperatura ambiente durante 30 minutos. Para el primer uso, adicione 1 ml de 10X de premezcla de cebador a la botella de 2X KAPA SYBR FAST qPCR Master Mix (5 ml) y mezcle. Calcule los volúmenes de 2X Mastermix, 50X ROX High Dye (si es necesario) y el grado de biología molecular H<sub>2</sub>O para las muestras y estándares HTG EdgeSeq según las instrucciones del fabricante. Cree la entrada de orden y el protocolo adecuados para el sistema qPCR que se está utilizando. Cargue el instrumento qPCR y realice un ciclo de las muestras según las instrucciones del fabricante. Después de completar el ciclo, exporte los resultados como un archivo de Microsoft Excel®. Asegúrese de que las métricas de QC para la curva estándar sean compatibles con las especificaciones del fabricante.

La biblioteca agrupada se puede crear, por ejemplo, utilizando la calculadora de la biblioteca de HTG EdgeSeq IVD.

Secuenciación en el sistema Illumina MiSeq: La secuenciación de la biblioteca agrupada se puede realizar de la siguiente manera: por ejemplo, utilizando el instrumento Illumina MiSeq según las instrucciones del fabricante. La concentración óptima final de la biblioteca agrupada para cada instrumento Illumina MiSeq puede variar. Se pueden analizar los datos de secuenciación resultantes.

El sistema de clasificación del ensayo de células de origen HTG EdgeSeq de DLBCL realiza una evaluación de la calidad de los datos antes de producir la clasificación de la muestra. Las muestras que no superan estos requisitos se informan como un "error de QC" con un número adjunto. Si se obtuvieron lecturas de secuenciación insuficientes para la clasificación, asegúrese de que la muestra se cuantificó correctamente y se adicionó suficiente biblioteca durante la etapa de agrupación y cuantificación de la biblioteca; vuelva a ejecutar la muestra de cualquier lisado de muestra restante, vuelva a lisar la muestra de una sección nueva de tejido FFPE. Vuelva a evaluar el tejido para asegurarse de que el área utilizada sea un tumor, o combinaciones del mismo. Si el perfil de los datos no refleja los patrones típicos de expresión obtenidos de DLBCL, esto indica un posible problema con la preparación de la muestra o el procesador. En esta situación, vuelva a analizar la muestra de cualquier lisado de muestra restante, y/o vuelva a analizar la muestra de una sección nueva de tejido FFPE y vuelva a evaluar el tejido para asegurarse de que el área utilizada sea un tumor.

Realización del ensayo de células de origen HTG EdgeSeq de DLBCL: Permita que los componentes congelados provistos con el ensayo de células de origen HTG EdgeSeq de DLBCL se descongelen a temperatura ambiente durante 80 minutos sin mezclar o agitar los reactivos. Descongele cualquier lisado de muestra congelado calentándolo a 50°C durante 30 minutos antes de pipetear las muestras en la placa de muestra. La ubicación de los consumibles dentro del procesador HTG EdgeSeq puede variar con el número de muestras a procesar.

Resultados del ensayo: El informe del ensayo produce tres resultados potenciales consistentes con los casos de DLBCL de: (1) "ABC" o subtipo de células B activadas; (2) "GCB" o subtipo de células B germinales; (3) "sin clasificar", donde el patrón de expresión génica no era indicativo de un ABC o GCB; o (4) "falla de QC", lo que indica que la muestra no produjo resultados que pasen las métricas de control de calidad a nivel de muestra antes de la clasificación, y las muestras se deben analizar nuevamente.

Los tumores que muestran un subtipo ABC generalmente tienen un peor pronóstico y respuesta a las terapias de DLBCL de estándar de cuidado que los del subtipo GCB. La subtipificación de los tumores de DLBCL puede dirigir al médico hacia formas alternativas de terapia en pacientes con recaídas. (Vose, J Clin Oncol 2011 29:4065-4066; Thieblemont y col., J Clin Oncol 2011 29:4079-87). Se han realizado cambios recientes en las clasificaciones de DLBCL de la Organización Mundial de la Salud para reconocer estas diferencias y facilitar las investigaciones sobre tratamientos más efectivos para el subtipo ABC (Swerdlow y col., Blood 2016; 127:2375-90). En estas opciones se incluyen ensayos de investigación para numerosas terapias nuevas.

**Características de rendimiento**

Entrada de muestras: Se analizaron lisados de muestra replicados de muestras FFPE (un ABC y un GCB) en diluciones dobles entre 12,5 y 0,39 mm<sup>2</sup> por pocillo. La clasificación se mantuvo constante en el 100% de los 6 puntos de dilución probados, incluso por debajo de los 1,56 mm<sup>2</sup> de tejido recomendados. Esta consistencia también se evaluó en dos sondas de baja expresión, CD274 (PD-L1) y PDCD1 (PD-1). Las FIG. 6 y 7 representan la expresión de muestra en muestras ABC y GCB que van desde cantidades de entrada de muestra bajas a altas.

Repetibilidad de ensayo: La repetibilidad diaria se evaluó utilizando 80 muestras (40 ABC, 40 GCB) caracterizadas previamente por el perfil de expresión génica (GEP) de Affymetrix™. Las muestras caracterizadas previamente como "sin clasificar" se excluyeron a propósito. Un equivalente de aproximadamente 5 mm<sup>2</sup> del tejido FFPE estaba presente en cada pocillo de muestra y se secuenció en tres días separados en un único procesador HTG EdgeSeq y MiSeq Sequencer. Del total de 240 conjuntos de datos producidos, 236 (98,33%) pasaron el control de calidad previo a la clasificación. Una muestra ABC y una GCB resultaron en una falla de QC en el día 1, dos muestras ABC en el día 2 y ninguna muestra en el día 3. La Tabla 9 describe los resultados finales del subtipo de muestra del Ensayo de células de origen HTG EdgeSeq de DLBCL en comparación con GEP, la tasa de acuerdo general fue del 100% en todos los días.

Tabla 9. Clasificación del Ensayo de células de origen HTG EdgeSeq de DLBCL para GEP

GEP	Día 1		Día 2		Día 3	
	ABC	GCB	ABC	GCB	ABC	GCB
ABC	39	0	38	0	40	0
GCB	0	39	0	40	0	40

40

Concordancia del ensayo con la IHC: La concordancia del ensayo de células de origen HTG EdgeSeq de DLBCL en comparación con las metodologías de IHC se realizó a través de un estudio retrospectivo utilizando dos cohortes independientes. La primera cohorte consistió en 132 muestras caracterizadas previamente como ABC o GCB a través del procedimiento Visco-Young (Lenz y col., N Engl J Med 2008; 359(22):2313-23). Una segunda cohorte de 24 muestras se caracterizó utilizando el procedimiento Choi (Rosenwald y col., N Engl J Med 2002; 346 (25):1937-47). Las tasas de acuerdo, excluyendo las muestras sin clasificar, se calcularon para cada cohorte. El Ensayo de células de origen HTG EdgeSeq de DLBCL, optimizado para GEP, funcionó de manera más consistente con el algoritmo de IHC Choi con un acuerdo general del 91% (Tabla 10). La concordancia general entre el Ensayo de células de origen HTG EdgeSeq de DLBCL y la IHC de Visco-Young fue del 83% (Tabla 11). Ninguna de las tasas de acuerdo generales incluyó las muestras sin clasificar, ya que los algoritmos de IHC no incluyen esta categoría de clasificación de subtipo.

50

Tabla 10. Clasificación del ensayo con HTG EdgeSeq de COO de DLBCL según la IHC de Choi

Choi	HTG EdgeSeq		
	ABC	GCB	UNC
	11	1	1

	GCB	1	10	0
--	-----	---	----	---

Acuerdo general con la IHC de Choi	91%
Tasa sin clasificar	4%

Tabla 11: Clasificación del ensayo con HTG EdgeSeq de células de origen de DLBCL según la IHC de Visco-Young

5

IHC de Visco-Young	Ensayo con HTGEdgeSeq de COO de DLBCL		
	ABC	GCB	UNC
ABC	60	12	6
GCB	9	40	5

Acuerdo general con la IHC de Visco-Young	83%
Tasa sin clasificar	6%

- 10 **Reproducibilidad diagnóstica:** Catorce muestras clínicas previamente caracterizadas se lisaron en dos cantidades de entrada de muestra. Se ejecutaron tres réplicas para todas las muestras en tres procesadores HTG EdgeSeq y se secuenciaron en un solo día. Se generó un total de 252 perfiles de clasificación de GCB/ABC agrupados en todas las condiciones. Un total del 97% de las muestras (244/252) pasó las métricas de QC para la clasificación posterior. Al excluir las muestras sin clasificar (1 clasificación errónea en total), se obtuvo un noventa y nueve por ciento (99%) de acuerdo; al incluir los resultados sin clasificar, se obtuvo un acuerdo del 97%. Hubo un total de 8 resultados sin clasificar, 6 de estos obtenidos de una sola muestra; estas tasas de acuerdo se resumen en la Tabla 12. La representación gráfica de la clasificación para cada réplica se muestra en la FIG. 8.
- 15

**Tabla 12:** Dentro del procesador y acuerdo de entrada de muestras

		% de aprobación del QC	% de acuerdo
<b>Procesador</b>	<b>Instrumento 1 (N=83)</b>	98,8%	94,0%
	<b>Instrumento 2 (N=81)</b>	96,4%	97,5%
	<b>Instrumento 3 (N=80)</b>	95,2%	97,5%
<b>Entrada de muestra</b>	<b>1,5 mm (N=118)</b>	93,7%	94,9%
	<b>5 mm (N=126)</b>	100,0%	97,6%

20

En vista de las muchas formas de realización posibles a las que se pueden aplicar los principios de la descripción, debe reconocerse que las formas de realización ilustradas son solo ejemplos y no deben considerarse como limitantes del alcance de la invención. Por el contrario, el alcance de la invención se define según las reivindicaciones a continuación.

25

#### LISTADO DE SECUENCIAS

<110> HTG Molecular Diagnostics, Inc.

30

<120> PROCEDIMIENTOS PARA SUBTIFICAR EL LINFOMA DE CÉLULAS B DIFUSAS (DLBCL)

<130> 8736-95651-02

35 <150> 62/234491

<151> 2015-09-29

<160> 16

<170> PatentIn versión 3.5

<210> 1

<211> 50

5 <212> ADN

<213> Secuencia artificial

<220>

<223> Porción de captura de NPPF utilizada para detectar CD47

10

<400> 1

aaactgggta cctagaggcc agaggcccta ggacctgaaa ggcattcattc 50

15 <210> 2

<211> 50

<212> ADN

<213> Secuencia artificial

20 <220>

<223> porción de captura de NPPF utilizada para detectar CD86

<400> 2

25 gtctcctctt ggcatacggg gcagagctgg agttacaggg aggctattcc 50

<210> 3

<211> 50

<212> ADN

30 <213> Secuencia artificial

<220>

<223> porción de captura de NPPF utilizada para detectar ENTPD1

35 <400> 3

ccatgaggaa gacataggtg gagtgggaga gaggtgtgga caatgggtgc 50

<210> 4

40 <211> 50

<212> ADN

<213> Secuencia artificial

<220>

45 <223> Porción de captura de NPPF utilizada para detectar FOXP1

<400> 4

ctcttcccgt attgcgctgg ctaagttgcc cagagtggga ttcccatgg 50

50

<210> 5

<211> 50

<212> ADN

<213> Secuencia artificial

55

<220>

<223> Porción de captura de NPPF utilizada para detectar FUT8

<400> 5

60

cagttggct gaaatggtt gaactgagtt tggcgaact gagtttggc 50

# ES 2 765 709 T3

<210> 6  
<211> 50  
<212> ADN  
<213> Secuencia artificial  
5  
<220>  
<223> Porción de captura de NPPF utilizada para detectar IL16  
<400> 6  
10 gtcacatcagg cagtgccctg atgatgttcc aggctcaaa ccgtgtgagg 50  
<210> 7  
<211> 50  
<212> ADN  
15 <213> Secuencia artificial  
<220>  
<223> Porción de captura de NPPF utilizada para detectar ITPKB  
20  
<400> 7  
gcgccctcaaa catgcccaact ttctggttca cctgcacggt ctgcaactcg 50  
25 <210> 8  
<211> 50  
<212> ADN  
<213> Secuencia artificial  
30 <220>  
<223> Porción de captura de NPPF utilizada para detectar LRMP  
<400> 8  
35 gctatttcta cagcgggtaa agtctcgata tgcaagttg cctctttctg 50  
<210> 9  
<211> 50  
<212> ADN  
40 <213> Secuencia artificial  
<220>  
<223> Porción de captura de NPPF utilizada para detectar MME  
45 <400> 9  
ggccttgccg aaagcatttc tggactcctt gtaggttcgg ctgaggctgc 50  
<210> 10  
50 <211> 50  
<212> ADN  
<213> Secuencia artificial  
<220>  
55 <223> Porción de captura de NPPF utilizada para detectar NF2  
<400> 10  
gagagaaatc tctacagggt cgtagttcaa ggcaattgca cataagaggg 50  
60  
<210> 11  
<211> 50



<212> ADN  
<213> Secuencia artificial

<220>

5 <223> Porción de captura de NPPF utilizada para detectar PIM1

<400> 11

gctcaccttc ttcagcagga ccacttccat gggcactcga gtgccattag 50

10

<210> 12  
<211> 50  
<212> ADN  
<213> Secuencia artificial

15

<220>  
<223> Porción de captura de NPPF utilizada para detectar PTPRC

<400> 12

20

ggaacaattt cctcctctgt taccctaaga acaaaccact tgctagctgc 50

<210> 13

<211> 50

25

<212> ADN  
<213> Secuencia artificial

<220>

<223> Porción de captura de NPPF utilizada para detectar REL

30

<400> 13

cagaggtgcc attgaggcat gatgtgacaa tccacttgag atgggccag 50

35

<210> 14  
<211> 50  
<212> ADN  
<213> Secuencia artificial

40

<220>  
<223> Porción de captura de NPPF utilizada para detectar STAT3

<400> 14

45

cagcttcagg atgctcctgg ctctctggcc gacaatactt tccgaatgcc 50

<210> 15

<211> 50

<212> ADN

50

<213> Secuencia artificial

<220>

<223> Porción de captura de NPPF utilizada para detectar TNFRSF8

55

<400> 15

ctgggaccaa tgctgttctc ggcagtgccc atcctgcat ctgtttgctg 50

60

<210> 16  
<211> 50  
<212> ADN  
<213> Secuencia artificial

<220>

<223> Porción de captura de NPPF utilizada para detectar TYMS

5 <400> 16

cttcaggccc gtgatgtgcg caatcatgta cgtgagcagg gcgtagctgg 50

**REIVINDICACIONES**

1. Un procedimiento para clasificar un linfoma difuso de células B grandes (DLBCL), que comprende:
  - 5 medir la expresión, opcionalmente la expresión de ácido nucleico, de CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, en una muestra obtenida de un sujeto, obteniendo así un valor de expresión para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;
  - ponderar cada valor de expresión, generando así valores de expresión ponderados;
  - 10 sumar los valores de expresión ponderados, generando así un valor de expresión ponderado sumado;
  - calcular una puntuación de probabilidad usando el valor de expresión ponderado sumado;
  - comparar la puntuación de probabilidad con los umbrales; y
  - clasificar el DLBCL como tipo células B activadas (ABC) cuando la puntuación de probabilidad es un valor en un intervalo para ABC;
  - 15 clasificar el DLBCL como tipo células B del centro germinal (GCB) cuando la puntuación de probabilidad es un valor en un intervalo para GCB; o
  - clasificar el DLBCL como sin clasificar cuando la puntuación de probabilidad es un valor en un intervalo para la categoría sin clasificar.
- 20 2. El procedimiento de la reivindicación 1, donde el ácido nucleico es ARNm y/o ARNm.
3. El procedimiento de la reivindicación 1 o 2, donde la muestra es una muestra fija y la expresión se analiza opcionalmente usando un ensayo basado en nucleasa.
- 25 4. El procedimiento de cualquiera de las reivindicaciones 1 a 3, donde medir de la expresión comprende secuenciación, y donde el valor de la expresión es un recuento que representa un número de moléculas presentes en la muestra y, opcionalmente, donde la expresión de medición comprende la cuantificación de la expresión.
5. El procedimiento de cualquiera de las reivindicaciones 1 a 4, donde ponderar cada valor de expresión
- 30 comprende multiplicar los valores de expresión por los coeficientes mostrados en la Tabla 5.
6. El procedimiento de cualquiera de las reivindicaciones 1 a 5, donde el cálculo de una puntuación de probabilidad usa la fórmula:

$$Pr(Y = y|x) = \frac{e^{(\beta_0 + x^T \beta)}}{1 + e^{(\beta_0 + x^T \beta)}} \quad \text{(Ecuación 1)}$$

- donde  $\beta_0$  es una constante,  $\beta$  es un vector de ponderación para los genes clasificadores, y  $x^T$  es un vector de valores medidos de los genes clasificadores para una muestra dada.
- 40 7. El procedimiento de cualquiera de las reivindicaciones 1 a 6, donde los umbrales son puntos límites t de puntuaciones de probabilidad.
  8. El procedimiento de la reivindicación 7, donde los puntos límite t de puntuaciones de probabilidad son < 0,43 para ABC, > 0,57 para GCB y 0,43 - 0,57 para sin clasificar.
  - 45 9. El procedimiento de cualquiera de las reivindicaciones 1 a 8, donde el procedimiento clasifica menos del 6% de las muestras como sin clasificar.
  10. El procedimiento de cualquiera de las reivindicaciones 1 a 9, donde los valores de expresión se obtienen
  - 50 al:
    - poner en contacto la muestra con sondas de protección de nucleasa que comprenden una secuencia flanqueante (NPPF) en condiciones suficientes para que cada NPPF se una específicamente a su molécula de ácido nucleico diana;
    - 55 poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la secuencia flanqueante (CFS) en condiciones suficientes para que la secuencia flanqueante se una específicamente a la CFS;
    - poner en contacto la muestra con una nucleasa específica para moléculas de ácido nucleico de una sola cadena en condiciones suficientes para eliminar las moléculas de ácido nucleico no unidas, generando así una muestra

digerida que comprende NPPF hibridadas para las moléculas de ácido nucleico diana y con la o las CFS; amplificar las NPPF en la muestra digerida con un cebador de amplificación, generando así amplicones de NPPF; secuenciar al menos una porción de los amplicones de NPPF; y contar el número de amplicones de NPPF; determinando así los valores de expresión para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS en la muestra.

11. El procedimiento de cualquiera de las reivindicaciones 1 a 9, donde los valores de expresión se obtienen al:

10 poner en contacto la muestra con NPPF específicas para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y moléculas de ácido nucleico diana TYMS,

donde cada NPPF comprende:

15 un extremo 5' y un extremo 3', una secuencia complementaria a una región de la molécula de ácido nucleico diana CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS, que permite la unión específica entre NPPF y la molécula de ácido nucleico diana, donde la secuencia flanqueante se ubica como 5', 3' o ambas, a la secuencia complementaria a la molécula de ácido nucleico diana, donde la secuencia flanqueante 5' se ubica como 5' en la secuencia complementaria a la molécula de ácido nucleico diana, y la secuencia flanqueante 3' se ubica como 3' en la secuencia complementaria a la molécula de ácido nucleico diana, donde la secuencia flanqueante comprende al menos 12 nucleótidos contiguos que no se encuentran en una molécula de ácido nucleico presente en la muestra,

25 si la NPPF comprende una secuencia flanqueante 5', poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la secuencia flanqueante 5' (5CFS), un fosfato del extremo 5', en condiciones suficientes para que la secuencia flanqueante 5' se hibride específicamente con la 5CFS;

30 si la NPPF comprende una secuencia flanqueante 3', poner en contacto la muestra con una molécula de ácido nucleico que comprende una secuencia complementaria a la secuencia flanqueante 3' (3CFS), en condiciones suficientes para que la secuencia flanqueante 3' hibride específicamente con la 3CFS;

donde al menos una de las 3CFS y 5CFS comprende un residuo de captura;

35 generar una NPPF hibridada con la molécula de ácido nucleico diana, hibridada con la 3CFS, hibridada con la 5CFS o hibridada con la 3CFS y la 5CFS;

poner en contacto la muestra con una nucleasa específica para moléculas de ácido nucleico de una sola cadena en condiciones suficientes para eliminar las moléculas de ácido nucleico no unidas, generando así una muestra digerida que comprende una NPPF hibridada con la molécula de ácido nucleico diana, hibridada con la 3CFS, hibridada con la

40 5CFS, o hibridada tanto con la 3CFS como con la 5CFS; capturar la NPPF hibridada con la molécula de ácido nucleico diana, hibridada con la 3CFS, hibridada con la 5CFS o hibridada con la 3CFS y la 5CFS;

ligar el fosfato 5' de la 3CFS a un extremo 3' de la molécula de ácido nucleico diana, y ligar un extremo 3' de la 5CFS a un extremo 5' de la molécula de ácido nucleico diana, generando así una molécula de ácido nucleico diana ligada;

45 separar la NPPF de la molécula de ácido nucleico diana ligada, generando así una mezcla que comprende una NPPF de una sola cadena y una molécula de ácido nucleico diana ligada de cadena simple; y

secuenciar al menos una porción de la molécula de ácido nucleico diana ligada de una sola cadena, determinando así la secuencia de moléculas de ácido nucleico diana de CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS en la muestra.

50 12. El procedimiento de la reivindicación 11, donde la NPPF comprende al menos un dUTP, y el procedimiento comprende además poner en contacto la mezcla que comprende la NPPF de una sola cadena y una molécula de ácido nucleico diana ligada de una sola cadena con uracilo ADN glicosilasa (UDG) en condiciones suficientes para degradar la NPPF de una sola cadena, después de la desnaturalización y antes de la secuenciación.

55 13. Una cantidad terapéuticamente efectiva de (1) quimioterapia con ciclofosfamida, doxorubicina, vincristina y prednisona o prednisolona (CHOP), (2) una quimioterapia con rituximab más CHOP (R-CHOP) o (3) una quimioterapia con etopósido más R-CHOP (R-EPOCH) para usar en el tratamiento de un sujeto diagnosticado por el procedimiento de cualquiera de las reivindicaciones 1 a 12 cuando el DLBCL se clasifica como GCB.

60 14. Una cantidad terapéuticamente eficaz de uno o más de bendamustina, pixantrona, gemcitabina/oxaliplatino, vincristina liposomal, mAb anti-CD20, mAb anti-CD22, mAb anti-CD74, mAb anti-CD40, una

construcción de mAb anti-CD19 y CD3 biespecífica de cadena simple, 1-131 tositumomab, Inotuzumab ozogamicina, 90Y-epratuzumab tetraxetán, talidomida, lenalidomida, Bortezomib, NPI-0052, Everolimus, Temsirolimus, Vorinostat, Oblimersen sodio, PF-3512676, 17-AAG, Bevacizumab, Aflibercept, CAL-101, ácido valproico, Dinaciclib, Fostamatinib, Dasatinib, Enzastaurin, PCI-32765, SB1518 y Sorafenib para usar en el tratamiento de un sujeto 5 diagnosticado mediante el procedimiento de cualquiera de las reivindicaciones 1 a 12 cuando el DLBCL se clasifica como ABC.

15. Uno o más medios de almacenamiento legibles por ordenador que almacenan instrucciones ejecutables por ordenador para hacer que el sistema informático, cuando así se programe, realice el procedimiento de cualquiera 10 de las reivindicaciones 1 a 12.

16. Un sistema informático adaptado para realizar el procedimiento de una cualquiera de las reivindicaciones 1 a 12.

15 17. Un kit que comprende:

un contenedor que comprende al menos 16 NPPF diferentes y CFS correspondientes, donde las al menos 16 NPPF diferentes son específicas para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;

20

uno o más de:

un recipiente que comprende microesferas que pueden unirse específicamente a amplicones de al menos 16 NPPF diferentes;

25

un recipiente que comprende una nucleasa específica para ácidos nucleicos de una sola cadena;

un recipiente que comprende un tampón de lisis;

un recipiente que comprende un tampón de lavado;

un recipiente que comprende regentes para PCR;

un recipiente que comprende etanol;

30

un recipiente que comprende una ligasa;

un recipiente que comprende un tampón de ligadura;

un recipiente que comprende aceite de desnaturalización;

un contenedor que comprende proteinasa K; y opcionalmente además comprende que el contenedor

35

comprenda al menos 16 NPPF diferentes y CFS correspondientes, donde las al menos 16 NPPF diferentes son específicas para CD47, CD86, ENTPD1, FOXP1, FUT8, IL16, ITPKB, LRMP, MME, NF2, PIM1, PTPRC, REL, STAT3, TNFRSF8 y TYMS;

una nucleasa específica para ácidos nucleicos de una sola cadena; tampón de lisis; y

un tampón de terminación de nucleasa, y el sistema informático según la reivindicación 16 que implementa un

40

clasificador para subtipos de DLBCL y opcionalmente donde el sistema informático comprende software o medio legible por computadora que recibe valores de expresión para dos o más genes de firma DLBCL, califica los valores múltiples contra umbrales correspondientes para cada gen, y clasifica una muestra en un marco que indica el subtipo de DLBCL de la muestra.

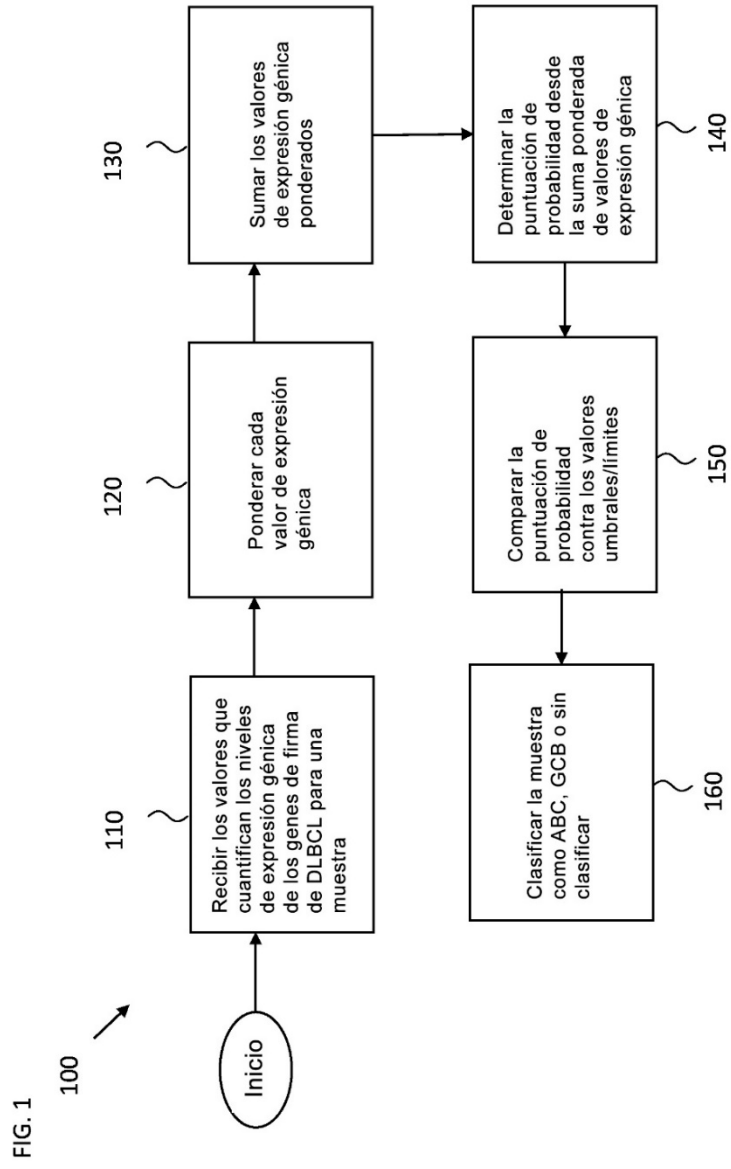
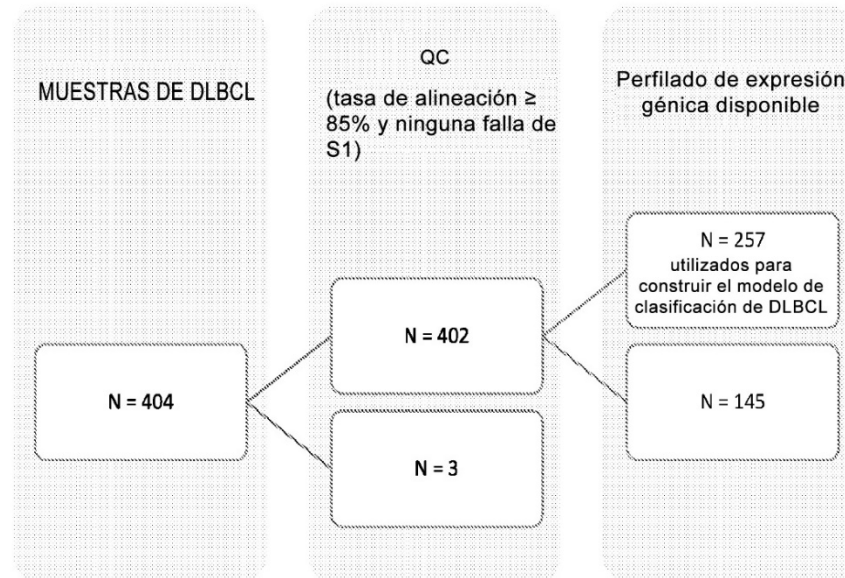
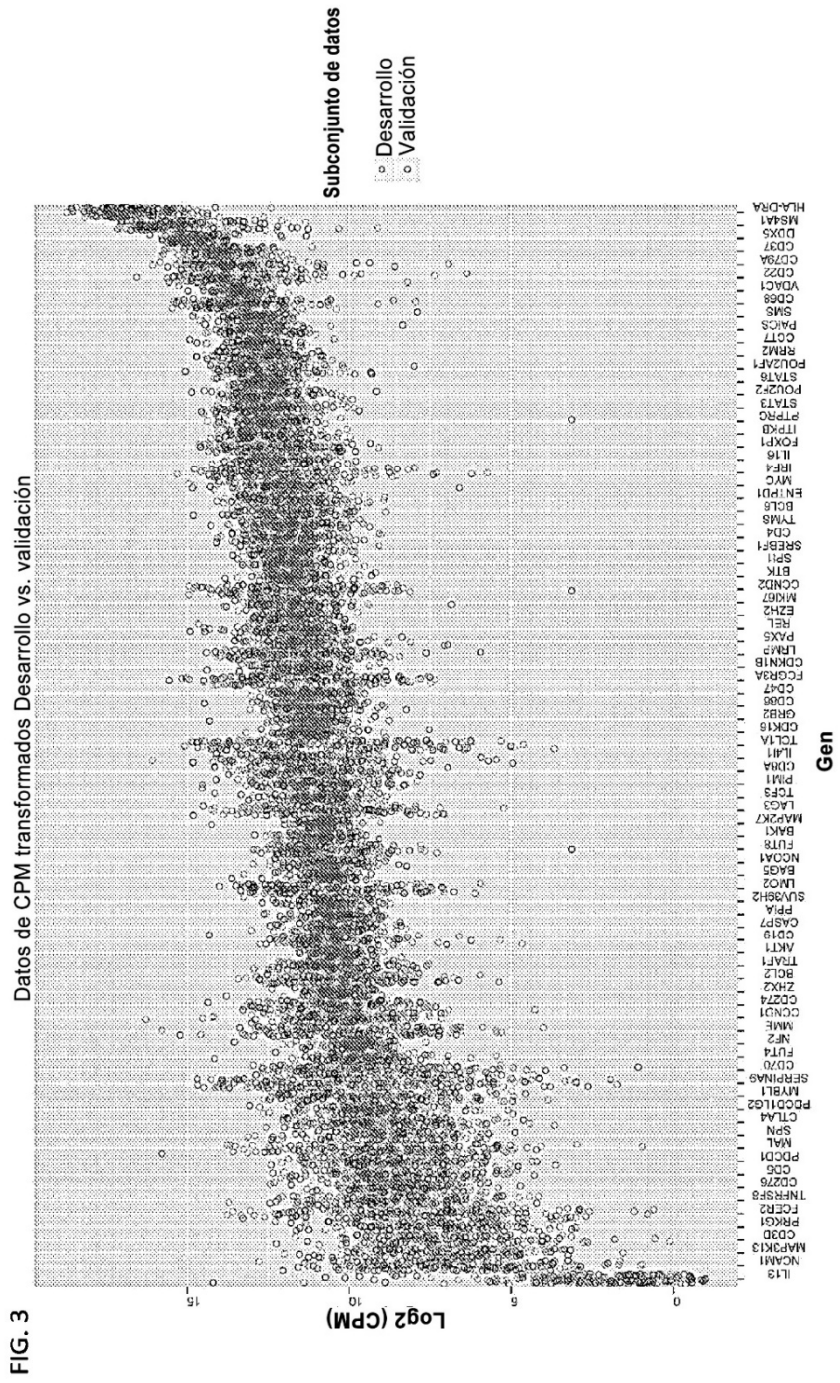
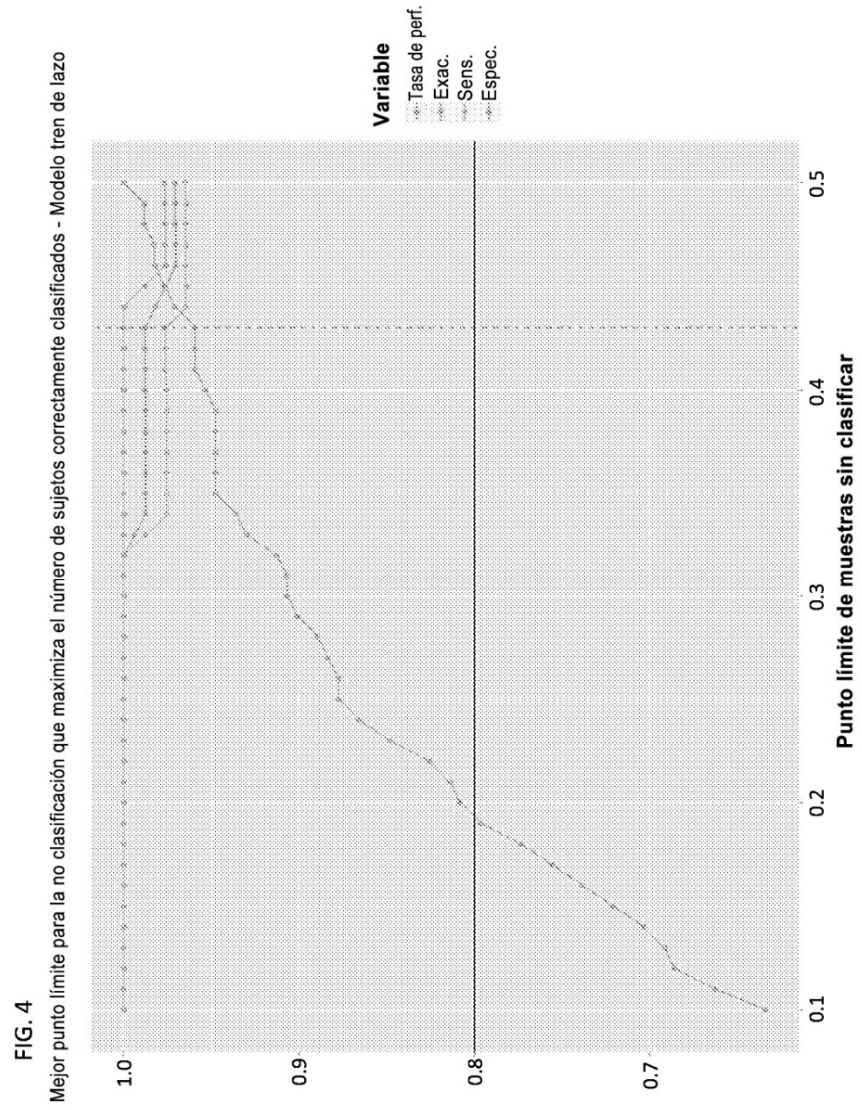


FIG. 2

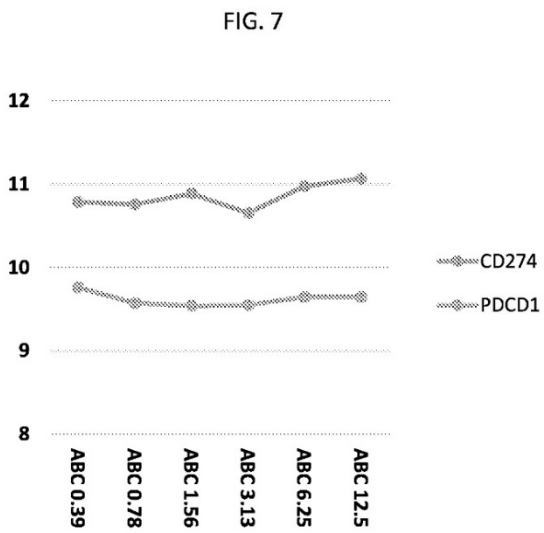
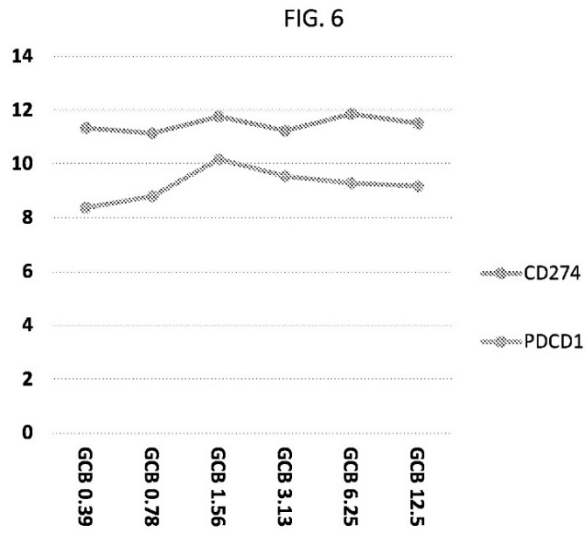


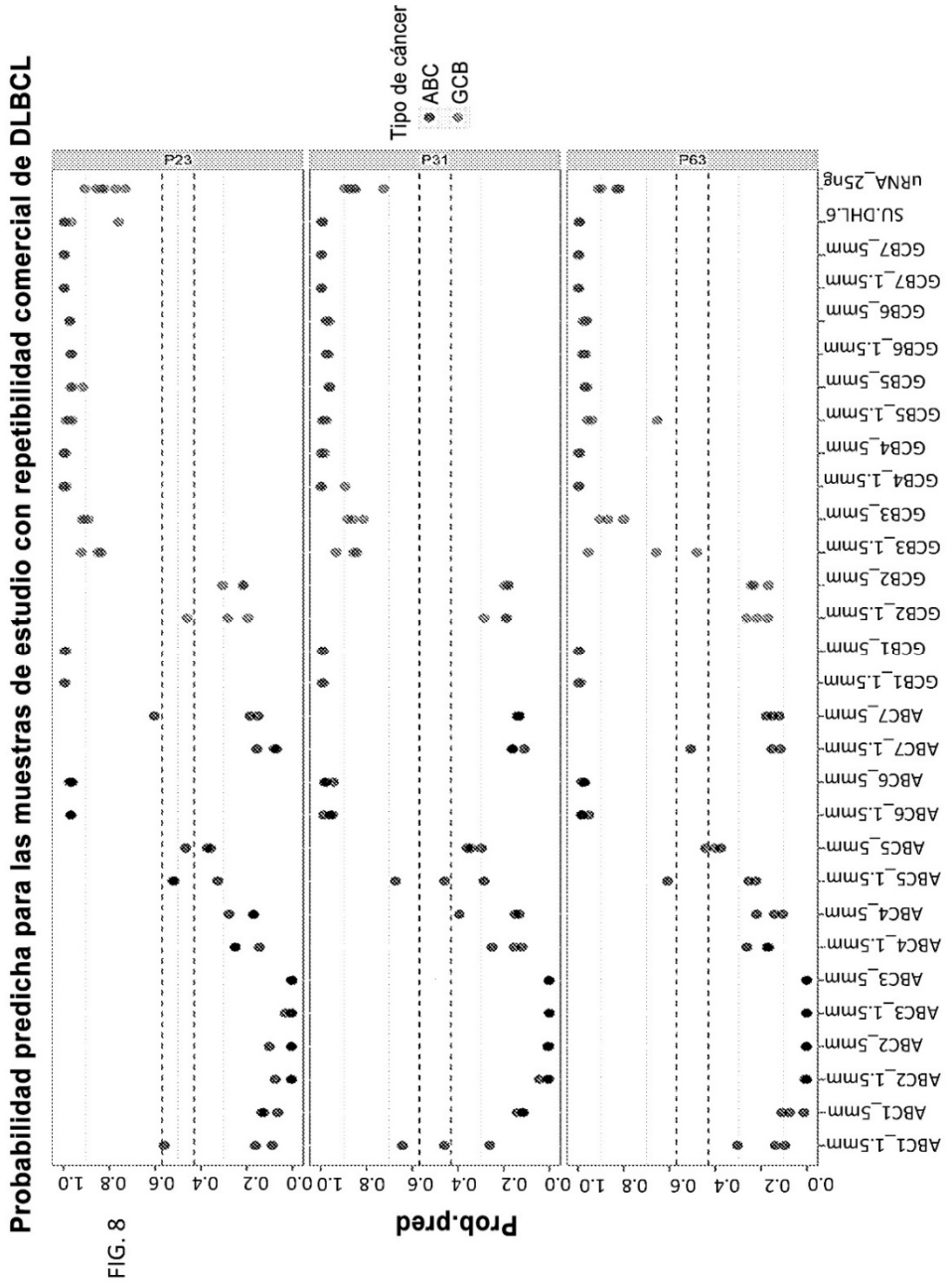












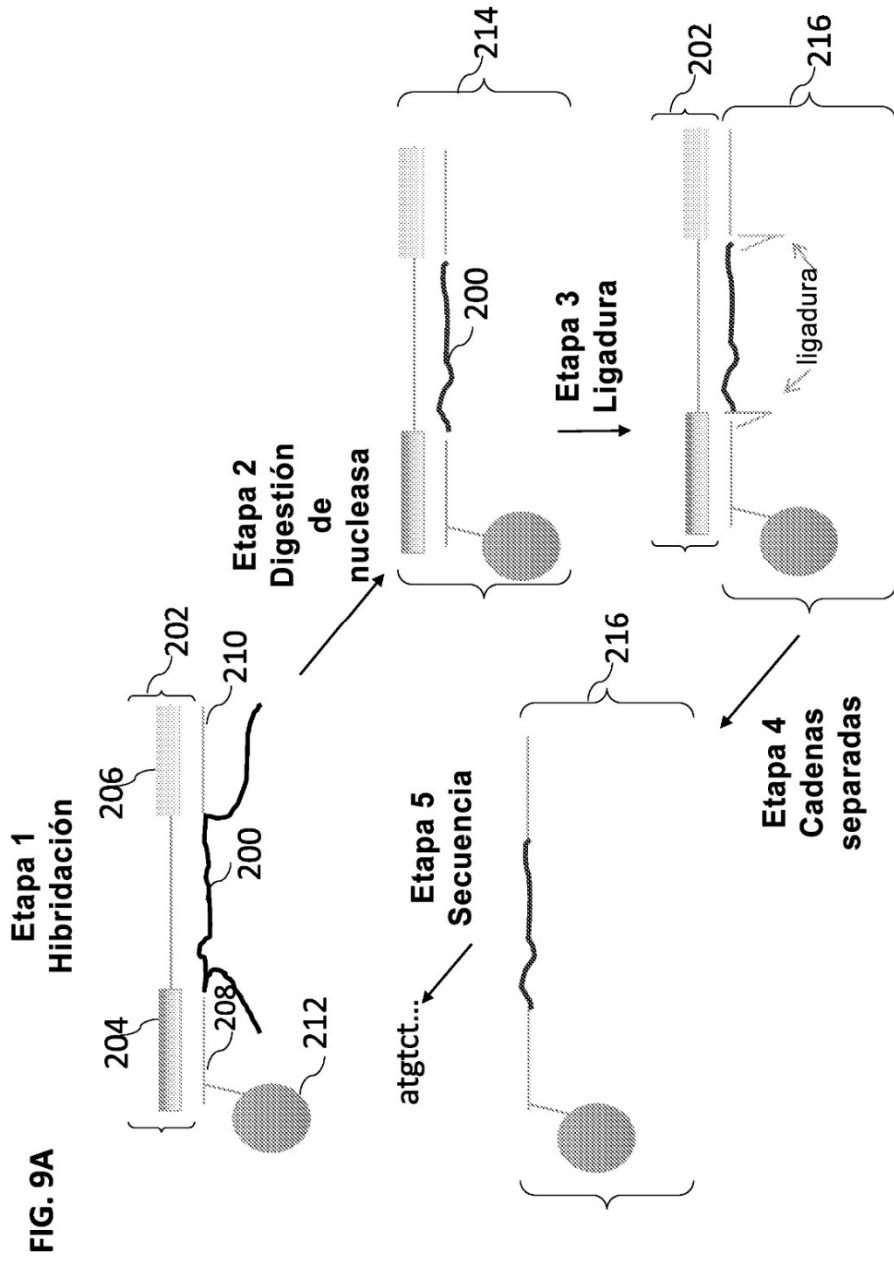


FIG. 9B Captura en la hibridación:

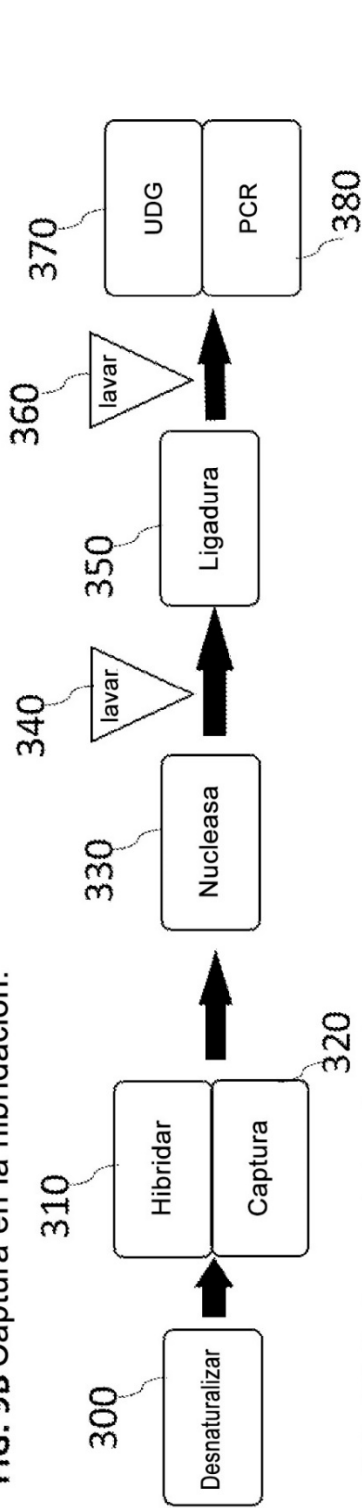


FIG. 9C Captura posterior a la nucleasa:

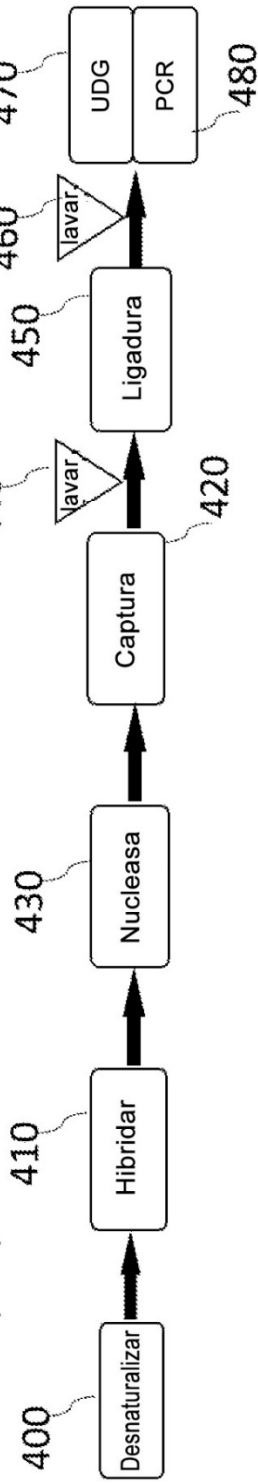


FIG. 9D Captura posterior a la hibridación:

