

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 441 374**

51 Int. Cl.:

C12Q 1/68 (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **17.06.2004 E 10175210 (3)**

97 Fecha y número de publicación de la concesión europea: **11.12.2013 EP 2264188**

54 Título: **Métodos y composiciones para la amplificación y genotificación de genomas completos**

30 Prioridad:

20.06.2003 US 600634
08.10.2003 US 681800

45 Fecha de publicación y mención en BOPI de la traducción de la patente:
04.02.2014

73 Titular/es:

ILLUMINA, INC. (100.0%)
9885 Towne Centre Drive
San Diego, California 92121-1975, US

72 Inventor/es:

GUNDERSON, KEVIN y
STEEMERS, FRANK

74 Agente/Representante:

UNGRÍA LÓPEZ, Javier

ES 2 441 374 T3

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

DESCRIPCIÓN

Métodos y composiciones para la amplificación y genotificación de genomas completos

5 **Campo de la invención**

La presente invención se refiere al análisis genético en general, y más concretamente, a la amplificación de genomas completos y a la genotificación basándose en pluralidades de marcadores genéticos que abarcan genomas.

10

Antecedentes de la invención

La mayor parte del ADN de una persona cualquiera, un 99,9 por ciento, es exactamente igual al ADN del resto de las personas. La diferencia del aproximadamente 0,1 % en la secuencia del genoma representa una amplia variedad de diferencias entre las personas, tales como el color de los ojos y el grupo sanguíneo. La variación genética también desempeña un papel en si una persona está en riesgo de contraer determinadas enfermedades o si una persona puede tener una respuesta favorable o adversa a un fármaco en particular. Las diferencias genéticas individuales se han asociado con un alto riesgo de adquirir una variedad de enfermedades tales como fibrosis quística y enfermedad de células falciformes. Las interrelaciones más complejas entre múltiples genes y el medio ambiente son responsables de muchos rasgos como el riesgo de padecer algunas enfermedades comunes tales como diabetes, cáncer, apoplejía, enfermedad de Alzheimer, enfermedad de Parkinson, depresión, alcoholismo, enfermedad cardíaca, artritis y asma.

15

20

25

30

Hay disponibles ensayos de diagnóstico basados en la genética para varias enfermedades muy penetrantes causadas por un solo gen, como la fibrosis quística. Tales ensayos se pueden realizar mediante el sondeo de mutaciones o polimorfismos particulares en los respectivos genes. Por consiguiente, es posible determinar el riesgo de contraer una enfermedad en particular mucho antes de que aparezcan los síntomas y, si se desea, se pueden tomar medidas preventivas. Sin embargo, se cree que la mayoría de las enfermedades, incluyendo muchas enfermedades comunes tales como la diabetes, enfermedades cardíacas, cáncer y trastornos psiquiátricos, se ven afectadas por múltiples genes, así como por condiciones ambientales. Por lo tanto, el diagnóstico de tales enfermedades basándose en la genética es considerablemente más complejo a medida que aumenta el número de genes por estudiar.

35

40

Recientemente, a través de una variedad de esfuerzos de genotipificación, se ha identificado un gran número de marcadores polimórficos de ADN, muchos de los cuales se cree que están asociados con la probabilidad de desarrollar determinados rasgos tales como el riesgo de adquirir enfermedades conocidas. Los marcadores de ADN polimórficos a modo de ejemplo que están disponibles incluyen polimorfismos de un solo nucleótido (SNP) que se producen con una frecuencia media de más de un 1 por kilobase en el ADN genómico humano. Es probable que muchos de estos SNP sean variantes genéticas terapéuticamente relevantes y/o que participen en la predisposición genética a padecer una enfermedad. Sin embargo, los métodos actuales para la consulta en todo el genoma de los SNP y otros marcadores no son eficaces, convirtiendo así la identificación de conjuntos de marcadores de diagnóstico útiles en poco práctica.

45

El documento WO02101022 divulga un método para la genotificación de polimorfismos de un solo nucleótido usando la amplificación de genomas completos y la extensión de una sola base o la extensión de cebadores específicos del alelo en una micromatriz.

50

El documento WO9923256 divulga un análisis en micromatriz de ADN genómico obtenido de la digestión de una endonucleasa de restricción, seguido de la unión de adaptadores y la amplificación usando cebadores complementarios a las secuencias de los adaptadores.

Syvanen *et al.*, (*Human Mutation*, 1999, vol 13, N° 1, páginas 1-10) y Zhao *et al.* (*Nucleic Acid Res.* 2001, vol 29, N° 4, páginas 955-959) divulgan la extensión de una sola base en una micromatriz.

55

El documento US2002039728 divulga sustratos y formatos de matrices a base de perlas.

Dean *et al.* (*Proc Nat Academy Sci*, 2002, vol 99, N° 8, páginas 5261-5266) y Hosono S. *et al.* (*Genome Research*, 2003, vol 13, N° 5, páginas 954-964) divulgan la amplificación de genomas completos usando la amplificación de múltiples desplazamientos seguida del análisis de los polimorfismos de ácidos nucleicos.

60

65

La capacidad de determinar simultáneamente el genotipo de grandes números de marcadores de SNP a través de una muestra de ADN cada vez está cobrando mayor importancia para los estudios de vinculación y asociación genética. Una limitación importante para los estudios de asociación de genomas completos es la falta de una tecnología para llevar a cabo la genotificación de SNP altamente multiplexados. La generación del mapa de haplotipos completo del genoma humano a través de los principales grupos étnicos proporcionará el contenido de SNP para los estudios de asociación de todo el genoma (estimado en aproximadamente 200.000-300.000 SNP). Sin

embargo, los métodos de genotipificación actualmente disponibles son tediosos y poco eficaces en la clasificación del gran número de SNP necesarios para generar un mapa de haplotipos.

5 Por lo tanto, existe la necesidad en la técnica de métodos para la consulta simultánea de grandes números de locus de genes a nivel del genoma completo. Tales beneficios afectarán al proceso de descubrimiento genómico y al análisis genético de enfermedades, así como al análisis genético a nivel individual. La presente invención satisface esta necesidad y proporciona otras ventajas adicionales. La presente invención describe y muestra un método para llevar a cabo reacciones de multiplexación a gran escala que darán paso a una nueva era en la genómica.

10 **Resumen de la invención**

En un aspecto, la presente invención presenta un método de acuerdo con la reivindicación 1. Las sondas de ácido nucleico son, como máximo, de 125 nucleótidos de longitud. Sin embargo, se pueden usar sondas que tengan cualquiera de una variedad de longitudes o secuencias como se expone en más detalle a continuación.

15 En otro aspecto, la presente divulgación presenta un método para la detección de locus tipificables de un genoma que incluye las etapas de proporcionar una población representativa amplificada de fragmentos de genoma que tengan tales locus tipificables; poner en contacto los fragmentos de genoma con una pluralidad de sondas de ácido nucleico que tengan secuencias correspondientes a los locus tipificables en condiciones donde se formen híbridos de sonda-fragmento; y detectar directamente locus tipificables de los híbridos de sonda-fragmento.

20 En un aspecto adicional, la presente divulgación presenta un método para la detección de locus tipificables de un genoma que incluye las etapas de proporcionar una población representativa amplificada de fragmentos del genoma que tienen los locus tipificables; poner en contacto los fragmentos de genoma de una pluralidad de sondas de ácido nucleico inmovilizadas que tienen secuencias correspondientes a los locus tipificables en condiciones donde se formen híbridos de sonda-fragmento inmovilizados; modificar los híbridos de sonda-fragmento inmovilizados; y detectar una sonda o un fragmento que se haya modificado, detectando de ese modo los locus tipificables del genoma.

30 La divulgación también proporciona un método que incluye las etapas de (a) proporcionar una pluralidad de fragmentos de genoma, donde la pluralidad de fragmentos de genoma tiene al menos 100 ug de ADN que tiene una complejidad de al menos 1 gigabase; (b) poner en contacto la pluralidad de fragmentos de genoma con una pluralidad de diferentes sondas de ácido nucleico inmovilizadas, donde al menos 500 de las diferentes sondas de ácido nucleico se hibridan con fragmentos de genoma para formar híbridos de sonda-fragmento; y (c) detectar locus tipificables de los híbridos de sonda-fragmento.

35 Un método de la divulgación también puede incluir las etapas de (a) proporcionar una pluralidad de fragmentos de genoma, donde la pluralidad de fragmentos de genoma tiene una concentración de al menos 1 ug/ul de ADN que tiene una complejidad de al menos 1 gigabase; (b) poner en contacto la pluralidad de fragmentos de genoma con una pluralidad de diferentes sondas de ácido nucleico inmovilizadas, donde al menos 500 de las diferentes sondas de ácido nucleico se hibridan con fragmentos de genoma para formar híbridos de sonda-fragmento; y (c) detectar locus tipificables de la los híbridos de sonda-fragmento.

45 En un aspecto adicional, la presente divulgación describe un método de amplificación de ADN genómico que incluye las etapas de proporcionar ADN genómico de doble cadena aislado; producir ADN mellado poniendo en contacto el ADN genómico de doble cadena con un agente de formación de muescas; poner en contacto este ADN mellado con una polimerasa de desplazamiento de cadena y una pluralidad de cebadores, con el fin de amplificar el ADN genómico.

50 La divulgación proporciona además un método para la detección de locus tipificables de un genoma. El método incluye las etapas de (a) transcribir *in vitro* una pluralidad de fragmentos de ADN_g amplificados, obteniéndose de este modo fragmentos de ARN genómico (ARN_g); (b) hibridar los fragmentos de ARN_g con una pluralidad de sondas de ácido nucleico que tengan secuencias correspondientes a los locus tipificables; y (c) detectar locus tipificables de los fragmentos de ARN_g que se hibriden con las sondas.

55 La divulgación proporciona además un método de producción de una población representativa amplificada, específica de los locus, de complejidad reducida de fragmentos de genoma. El método incluye las etapas de (a) replicar un genoma nativo con una pluralidad de cebadores aleatorios, produciendo de este modo una población representativa amplificada de fragmentos de genoma; (b) replicar una subpoblación de la población representativa amplificada de fragmentos de genoma con una pluralidad de diferentes cebadores específicos de los locus, produciendo de este modo una población representativa amplificada, específica de los locus, de fragmentos de genoma; y (c) aislar la subpoblación, produciendo así una población representativa amplificada, específica de los locus, de complejidad reducida de fragmentos de genoma.

65 La divulgación también proporciona un método para inhibir la extensión ectópica de sondas en un ensayo de extensión de cebadores. El método incluye las etapas de (a) poner en contacto una pluralidad de sondas de ácido

nucleico con una pluralidad de ácidos nucleicos diana en condiciones donde se formen híbridos de sonda-diana; (b) poner en contacto la pluralidad de sondas de ácido nucleico con un inhibidor de la extensión ectópica en condiciones donde se formen híbridos de inhibidor de la extensión ectópica-sonda; y (c) modificar selectivamente las sondas de los híbridos de sonda-diana en comparación con las sondas de los híbridos de inhibidor de la extensión ectópica-sonda.

También se proporciona un método que incluye las etapas de (a) poner en contacto una pluralidad de fragmentos de genoma con una pluralidad de diferentes sondas de ácido nucleico inmovilizadas en condiciones donde se formen híbridos de sonda inmovilizada-fragmento; (b) modificar las sondas inmovilizadas mientras están hibridadas con los fragmentos de genoma, formando de este modo sondas inmovilizadas modificadas; (c) retirar dichos fragmentos de genoma de dichos híbridos de sonda-fragmento; y (d) detectar las sondas inmovilizadas modificadas tras la eliminación de los fragmentos de genoma, detectando de este modo locus tipificables de los fragmentos de genoma.

La divulgación también proporciona un método que incluye las etapas de (a) amplificación de manera representativa de un genoma nativo, donde se produce una población representativa amplificada de fragmentos de genoma que tienen los locus tipificables en condiciones isotérmicas; (b) poner en contacto los fragmentos de genoma con una pluralidad de sondas de ácido nucleico que tienen secuencias correspondientes a los locus tipificables en condiciones donde se formen híbridos de sonda-fragmento; y (c) detectar locus tipificables de los híbridos de sonda-fragmento.

Breve descripción de las figuras

La Figura 1 muestra un diagrama de un método de genotipificación de genomas completos (GGC) de la invención.

La Figura 2 muestra ejemplos de sondas útiles para la detección de locus tipificables usando la extensión de cebadores específicos del alelo (ASPE) o la extensión de una sola base (SBE).

La Figura 3 muestra, en el Panel A, geles de agarosa cargados con productos de amplificación de reacciones de amplificación del genoma completo llevadas a cabo en diversas condiciones, y en el Panel B, una tabla de rendimientos calculados para las reacciones.

La Figura 4 muestra una imagen de la señal de una matriz de ADN genómico de levadura analizada en una BeadArray® (Panel A) y un subconjunto de intensidades de coincidencia perfecta (CP) y de falta de coincidencia (FC) para 18 locus de los 192 analizados de cuatro matrices diferentes por cuadruplicado (RSC1, RSC2, R6C1, R6C2) (Panel B). Las sondas de CP son el primer conjunto de cuatro valores de intensidad y las sondas de FC son el segundo conjunto de cuatro valores de intensidad indicados por cada marcador de tipo sonda del eje inferior.

La Figura 5 muestra la genotipificación mediante SBE basada en matriz realizada en ADN_g humano hibridado directamente a BeadArray®.

La Figura 6 muestra la genotipificación mediante ASPE basada en matriz realizada en ADN_g humano hibridado directamente a una BeadArray®. El Panel A muestra los valores de intensidad bruta a través de los 77 pares de sondas, y el Panel B muestra los coeficientes de discriminación (CP/CP + FC) representados para los 77 locus.

La Figura 7 muestra las puntuaciones de la genotipificación de ADN genómico no amplificado en comparación con ADN genómico amplificado por cebadores aleatorios (RPA) usando el ensayo GoldenGate® (bajo cada barra, se muestra la cantidad de entrada de ADN en la reacción de RPA; las reacciones de RPA emplearon oligonucleótidos 9-méricos aleatorios, excepto donde se especifica el uso de hexanucleótidos (6-méricos) o dodecanucleótidos (12-méricos)).

La Figura 8 muestra un diagrama de un método ilustrativo para la generación de ARN genómico como un ácido nucleico diana para la amplificación o detección.

La Figura 9 muestra un diagrama de un método ilustrativo para la generación de una población representativa específica del locus, de complejidad reducida, de fragmentos de genoma.

La Figura 10 muestra un esquema ilustrativo de amplificación de señales.

La Figura 11 muestra, en el Panel A, una imagen de una BeadArray® hibridada con fragmentos de ADN genómico y detectada con ASPE, y en el Panel B, una gráfica de GenTrain donde se diferencian dos agrupaciones homocigóticas (B/B y A/D) y una agrupación heterocigótica (A/B) en un locus.

La Figura 12 muestra, en el Panel A, una tabla de estadísticas de precisión de la genotipificación; en los Paneles B y C, representaciones de GenCall para dos muestras (la línea de 0,45 indica un umbral inferior usado para filtrar los datos a los que se vaya a acceder), y en los Paneles D y E, representaciones de GenTrain para dos locus (las flechas indican los puntos de datos cuestionables que no se necesitaron al estar bajo un umbral de 0,45 en las representaciones de GenCall).

La Figura 13 muestra diagramas que ilustran la extensión ectópica (Panel A) y métodos para inhibir la extensión ectópica incluyendo la inhibición mediante la unión de sondas de una sola cadena a SSB (Panel B); el bloqueo del extremo 3' de las sondas con ácidos nucleicos que tienen secuencias complementarias (Panel C); y la formación de horquillas no extensibles (Panel D).

La Figura 14 muestra diagramas de dispersión para las reacciones de ASPE cebadas por Klenow en BeadArray® que comparan la señal del ensayo en presencia y en ausencia de la proteína de unión de cadena simple (SSB). El diagrama de dispersión del Panel A muestra el efecto de la SSB en la intensidad de la señal

ectópica en ausencia de ADN genómico amplificado, mientras que el diagrama de dispersión del Panel B muestra el efecto de SSB sobre la intensidad de la señal en presencia de ADN genómico amplificado. Los Paneles C y D muestran representaciones de la intensidad de los locus (ordenados en orden creciente de intensidad), para reacciones de ASPE bien de Klenow (Panel C) o KlenTaq (Panel D) realizadas en BeadArray® en ausencia de una población amplificada de fragmentos de genoma (el ntc, control sin diana, proporciona una medida de la extensión "ectópica").

Figura 15 muestra diagramas de dispersión que comparan los valores de intensidad para las sondas después de la detección mediante ASPE de las poblaciones de fragmentos de genoma producidas mediante amplificación por cebadores aleatorios (amplificadas) y/o ADN genómico no amplificado (no amplificadas).

La Figura 16 muestra una distribución del número de sondas (recuentos) que tienen determinadas proporciones de intensidades de señal para las entradas de ADN no amplificado con respecto al amplificado (proporción de amplificado:no amplificado).

La Figura 17 muestra Genoplots ilustrativas para cuatro locus (1824, 2706, 3633 y 6126) detectados a partir de poblaciones amplificadas de manera representativa de fragmentos de genoma usando el ensayo GoldenGate®.

Las poblaciones amplificadas de manera representativa de fragmentos de genoma se produjeron por separado de muestras de ADN genómico en las tres cantidades diferentes indicadas en la leyenda. Los puntos de datos de control se obtuvieron de ADN genómico no amplificado detectado en las mismas condiciones usando el ensayo GoldenGate®. Las agrupaciones de puntos de datos de control identificadas por el algoritmo GenTrain están rodeadas y el número de puntos de datos de cada agrupación se indica bajo el eje x. Para el locus 2706, la agrupación vacía indica una ubicación de la agrupación predicha para el genotipo AA en base a las ubicaciones de las agrupaciones AB y BB.

La Figura 18 muestra (A) un gráfico de barras que representa la intensidad media detectada para todas las sondas en cada matriz (LOD) después de la hibridación y la detección por ASPE de mezclas de reacción RPA generadas a partir de diferentes cantidades de ADN genómico de entrada (entrada); y (B) un gráfico de barras que representa la proporción (intensidad de la señal de CP/(intensidad de la señal de CP + intensidad de la señal de FC) para todas las sondas de una matriz (proporción) usadas para sondear mezclas RPA producidas a partir de cantidades variables de ADN genómico de entrada (entrada).

La Figura 19 muestra Genoplots representativas del locus 860 (Panel A) y locus 954 (Panel B) para fragmentos de genoma humano amplificados por cebadores aleatorios producidos a partir de 95 muestras humanas de CEPH y detectados mediante extensión de cebadores específicos del alelo de sondas en una matriz que tiene sondas específicas del conjunto de locus 1500 HapMap QC. El panel C muestra la distribución de los locus de acuerdo con la puntuación de separación de las agrupaciones de genotipos.

La Figura 20 muestra la intensidad de señal para las sondas de coincidencia perfecta (CP) y de falta de coincidencia (FC) tras la detección por extensión de cebadores específicos del alelo y el tratamiento con o sin NaOH 0,1 N.

La Figura 21 muestra (A) el tratamiento con fosfatasa alcalina y ADN quinasa de T4 de fragmentos de ADN generados con bisulfito para generar productos bien completamente desfosforilados o 3'-desfosforilados, respectivamente; (B) el tratamiento de ADN 3'-desfosforilado con ARN ligasa de T4 para producir ADN concatenado seguido de la amplificación en una reacción de amplificación por cebadores aleatorios del genoma completo de desplazamiento de cadena; (C) el tratamiento con desoxinucleótido transferasa terminal (TdT) y ARN ligasa de T4 de fragmentos de ADN generados con bisulfito para añadir secuencias de cola universales a los fragmentos, seguido de amplificación por PCR; (D) el tratamiento con ARN ligasa de T4 de fragmentos de ADN generados con bisulfito para añadir colas de secuencia de cola universal 5' y 3' al producto de bisulfito seguido de amplificación por PCR.

Definiciones

Como se usa en el presente documento, el término "genoma" pretende significar el complemento completo de ADN cromosómico que se encuentra dentro del núcleo de una célula eucariota. El término también se puede usar para referirse a todo el complemento genético de un procarionte, un virus, una mitocondria o un cloroplasto, o al complemento genético nuclear haploide de una especie eucariota.

Como se usa en el presente documento, la expresión "ADN genómico" o "ADNg" pretende significar una o más moléculas desoxirribonucleótidas poliméricas cromosómicas de origen natural del núcleo de una célula eucariota o un procarionte, un virus, una mitocondria o un cloroplasto y que contiene secuencias que son transcritas de forma natural en ARN, así como secuencias que no son transcritas de forma natural en ARN por la célula. Un ADNg de una célula eucariota contiene al menos un centrómero, dos telómeros, un origen de replicación y una secuencia que no es transcrita en ARN por la célula eucariota que incluye, por ejemplo, un intrón o promotor de la transcripción. Un ADNg de una célula procarionte contiene al menos un origen de replicación y una secuencia que no es transcrita en ARN por la célula procarionte que incluye, por ejemplo, un promotor de la transcripción. Un ADN genómico eucariota se puede distinguir de ADN genómico procarionte, viral u organoide, por ejemplo, de acuerdo con la presencia de intrones en el ADN genómico eucariota y la ausencia de intrones en los otros ADNg.

Como se usa en el presente documento, el término "detectar" pretende significar cualquier método de determinación de la presencia de una determinada molécula, tal como un ácido nucleico que tenga una secuencia de nucleótidos específica. Las técnicas usadas para detectar un ácido nucleico incluyen, por ejemplo, la hibridación a la secuencia por detectar. Sin embargo, las realizaciones particulares de la presente invención no requieren la hibridación

directamente a la secuencia por detectar, sino que, más bien, la hibridación se puede producir cerca de la secuencia por detectar, o adyacente a la secuencia por detectar. El uso del término "cerca" pretende dar a entender a un intervalo de aproximadamente 150 bases de la secuencia por detectar. Otras distancias a lo largo de un ácido nucleico que están en el intervalo de aproximadamente 150 bases y, por lo tanto, cerca incluyen, por ejemplo, a

5 aproximadamente 100, 50, 40, 30, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 bases de la secuencia por detectar. La hibridación se puede dar en secuencias que disten más del locus o de la secuencia por detectar, incluyendo, por ejemplo, una distancia de aproximadamente 250 bases, 500 bases, 1 kilobase o más, hasta e incluyendo la longitud de los ácidos nucleicos diana o fragmentos de genoma que se están detectando.

10 Los ejemplos de reactivos que son útiles para la detección incluyen, pero sin limitación, sondas radiomarcadas, sondas marcadas con fluoróforos, sondas marcadas con puntos cuánticos, sondas marcadas con cromóforos, sondas marcadas con enzimas, sondas marcadas con ligandos de afinidad, sondas marcadas con espines electromagnéticos, sondas marcadas con átomos pesados, sondas marcadas con marcadores de nanopartículas de dispersión de la luz, u otras nanopartículas o cubiertas esféricas, y sondas marcadas con cualquier otro marcador de

15 generación de señales conocido por los expertos en la materia. Los ejemplos no limitantes de restos marcadores útiles para la detección en la invención incluyen, sin limitación, enzimas adecuadas tales como peroxidasa de rábano picante, fosfatasa alcalina, β -galactosidasa o acetilcolinesterasa; miembros de un par de unión que son capaces de formar complejos tales como estreptavidina/biotina, avidina/biotina o un complejo antígeno/anticuerpo incluyendo, por ejemplo, IgG de conejo e IgG anti-conejo; fluoróforos tales como umbeliferona, fluoresceína, isotiocianato de

20 fluoresceína, rodamina, tetrametilrodamina, eosina, proteína fluorescente verde, eritrosina, cumarina, metilcumarina, pireno, verde malaquita, estilbeno, amarillo lucifer, Cascade Blue®, rojo Texas, fluoresceína de diclorotriazinilamina, cloruro de dansilo, ficoeritrina, complejos lantánidos fluorescentes tales como los que incluyen europio y terbio, Cy3, Cy5, balizas moleculares y derivados fluorescentes de los mismos, así como otros conocidos en la técnica como se describe, por ejemplo, en "Principles of Fluorescence Spectroscopy", Joseph R. Lakowicz (Redactor), Plenum Pub Corp, II edición (julio de 1999) y la VI edición de "Molecular Probes Handbook" de Richard P. Hoagland; un material

25 luminiscente tal como luminol; materiales de dispersión de luz o de resonancia de plasmones tales como partículas de oro o plata, o puntos cuánticos; o materiales radiactivos que incluyen ^{14}C , ^{123}I , ^{124}I , ^{125}I , ^{131}I , Tc99m, ^{35}S o ^3H .

Como se usa en el presente documento, la expresión "locus tipificables" pretende significar lugares específicos de la

30 secuencia de un ácido nucleico. La expresión puede incluir secuencias de ácido nucleico predeterminadas o previstas que se espera que estén presentes en las moléculas de ácido nucleico aisladas. La expresión "locus tipificables" pretende abarcar polimorfismos de un solo nucleótido (SNP), mutaciones, número variable de repeticiones en tándem (VNTR) y repeticiones en tándem simples (STR), otros polimorfismos, inserciones, deleciones, variantes de corte y empalme o cualquier otros marcadores genéticos conocidos. Los ejemplos de

35 recursos que proporcionan SNP conocidos y otras variaciones genéticas incluyen, pero sin limitación, el dbSNP administrado por el NCBI y disponible en línea en ncbi.nlm.nih.gov/SNP/ y la base de datos HCVBASE descrita en Fredman *et al.*, *Nucleic Acids Research*, 30:387-91, (2002), que se encuentra disponible en línea en hgvsbase.cgb.ki.se/.

40 Como se usa en el presente documento, la expresión "amplificar de manera representativa" pretende significar la replicación de un molde de ácido nucleico para producir una copia de ácido nucleico en la que la proporción de cada secuencia de la copia con respecto a todas las demás secuencias de la copia es sustancialmente igual a las proporciones en el molde de ácido nucleico. Un molde de ácido nucleico incluido en la expresión puede ser una sola

45 molécula, tal como un cromosoma, o una pluralidad de moléculas, tal como una colección de cromosomas que componen un genoma o una parte de un genoma. De manera similar, una copia de ácido nucleico puede ser una sola molécula o una pluralidad de moléculas. Los ácidos nucleicos pueden ser ADN o ARN, o miméticos o derivados de los mismos. Un ácido nucleico de copia puede ser una pluralidad de fragmentos que sean más pequeños que el molde de ADN. Por consiguiente, la expresión puede incluir la replicación de un genoma, o parte del mismo, de

50 manera que la proporción de cada fragmento del genoma resultante con respecto al resto de fragmentos de genoma de la población sea sustancialmente igual a la proporción de su secuencia con respecto a otras secuencias de fragmentos de genoma del genoma. El ADN que se replica se puede aislar de una muestra de tejido o de sangre, de una muestra forense, de una célula fijada con formalina o de otras fuentes. Un ADN genómico usado en la invención puede estar intacto, mayormente intacto o fragmentado. Una molécula de ácido nucleico, tal como un molde o una

55 copia del mismo, puede ser cualquiera de una variedad de tamaños, incluyendo, sin limitación, como máximo aproximadamente 1 MB, 0,5 MB, 0,1 MB, 50 kb, 10 kb, 5 kb, 3 kb, 2 kb, 1 kb, 0,5 kb, 0,25; 0,1; 0,05 o 0,02 kb.

Por consiguiente, la expresión "representativa amplificada" pretende significar una copia de ácido nucleico en la que la proporción de cada secuencia en la copia con respecto a todas las demás secuencias de la copia es sustancialmente igual a las proporciones en el molde de ácido nucleico. Cuando se usa en referencia a una

60 población de fragmentos de genoma, por ejemplo, la expresión pretende significar una población de fragmentos de genoma donde la proporción de cada fragmento del genoma con respecto al resto de fragmentos de genoma de la población es sustancialmente igual a la proporción de su secuencia con respecto a las otras secuencias de fragmentos de genoma del genoma. "Similitud sustancial" entre la proporción de secuencias en una representación amplificada y un molde de ADN genómico significa que al menos un 60 % de los locus de la representación están no

65 más de 5 veces sobre-representados o infra-representados. En tales representaciones, al menos, el 70 %, 80 %, 90 %, 95 % o 99 % de los locus pueden estar, por ejemplo, no más de 5, 4, 3 o 2 veces sobre-representados o infra-

representados. Un ácido nucleico incluido en la expresión puede ser ADN, ARN o un análogo del mismo. El número de copias de cada secuencia de ácido nucleico de una población representativa amplificada puede ser, por ejemplo, al menos 2, 5, 10, 25, 50, 100, 1.000, 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 o 1×10^{10} veces superior al molde o más.

5 Las poblaciones de fragmentos de genoma ilustrativas que incluyen secuencias idénticas a una porción de un genoma incluyen, por ejemplo, representaciones de alta complejidad o representaciones de baja complejidad. Como se usa en el presente documento, la expresión "representación de alta complejidad" pretende significar una copia de ácido nucleico que tiene al menos aproximadamente un 50 % de la secuencia de su molde. Por lo tanto, una
10 representación de alta complejidad de un ADN genómico puede incluir, sin limitación, al menos aproximadamente un 60 %, 70 %, 75 %, 80 %, 85 %, 90 %, 95 % o 99 % de la secuencia del genoma del molde. Como se usa en el presente documento, la expresión "representación de baja complejidad" pretende significar una copia de ácido nucleico que tiene como máximo aproximadamente un 49 % de la secuencia de su molde. Por lo tanto, una
15 representación de baja complejidad de un ADN genómico puede incluir, sin limitación, como máximo aproximadamente un 49 %, 40 %, 30 %, 20 %, 10 %, 5 % o 1 % de la secuencia del genoma. En realizaciones particulares, una población de fragmentos de genoma de la invención puede tener una complejidad que represente al menos aproximadamente un 5 %, 10 %, 20 %, 30 % o 40 % de la secuencia del genoma.

20 Como se usa en el presente documento, la expresión "detectar directamente," cuando se usa en referencia a un ácido nucleico, pretende significar percibir o discernir una propiedad del ácido nucleico en una muestra basándose en el nivel del ácido nucleico de la muestra. La expresión puede incluir, por ejemplo, percibir o discernir una propiedad de un ácido nucleico de una muestra sin amplificar el ácido nucleico de la muestra, o la detección sin amplificación. A modo de ejemplo, una propiedad que se puede percibir o discernir incluye, sin limitación, una
25 secuencia de nucleótidos, la presencia de un determinado nucleótido tal como un polimorfismo o una mutación en un determinado sitio de una secuencia, o similares. Un ejemplo no limitante de un método de detección directa es la detección de un ácido nucleico por hibridación de una sonda marcada con el ácido nucleico y la determinación de la presencia del ácido nucleico basándose en la presencia del marcador hibridado. En el presente documento, se describen otros ejemplos de detección directa, e incluyen, por ejemplo, la extensión de una sola base (SBE) y la extensión de cebadores específicos del alelo (ASPE). Los expertos en la materia entenderán que, tras la detección,
30 es posible amplificar una muestra de ácido nucleico no amplificado, tal como una muestra de fragmentos de ADN genómico sin amplificar.

35 En realizaciones particulares, la detección directa puede incluir la generación de un complejo de ácido nucleico de doble cadena entre un locus tipificable y su secuencia complementaria, y percibir el complejo sin generar copias adicionales del locus tipificable. En algunas realizaciones, la detección directa de un locus tipificable puede implicar la formación de un solo complejo de hibridación excluyendo, de este modo, la hibridación repetida con una determinada molécula de ácido nucleico que tiene el locus tipificable.

40 Un método de detección de una posición detectable, tal como un locus tipificable o una secuencia ligada genéticamente a un locus tipificable, puede incluir, por ejemplo, la hibridación de un oligonucleótido con la posición consultada, o la hibridación por un oligonucleótido cercano o adyacente con la posición consultada, seguida de la extensión del oligonucleótido hibridado a través de la posición consultada.

45 Varios métodos de detección directa útiles en la invención y descritos en el presente documento, incluyendo, sin limitación, SBE y ASPE, emplean sondas que tanto capturan un fragmento del genoma como producen una señal indicativa de la presencia de un determinado locus de SNP en el fragmento. En particular, se puede llevar a cabo un método de la invención en condiciones en las que la detección de un SNP u otra característica de un oligonucleótido capturado, tal como un fragmento del genoma, no requiera la adición exógena de un oligonucleótido de consulta. Sin embargo, si se desea, se pueden usar oligonucleótidos de consulta añadidos de forma exógena. A continuación, se
50 exponen ejemplos de métodos que emplean oligonucleótidos de consulta añadidos exógenamente, tales como ensayo de ligación de oligonucleótidos (OLA), ligación por extensión (GoldenGate®), métodos de detección basados en círculos rodantes, hibridación de oligonucleótidos específicos del alelo (ASO) y otros.

55 Como se usa en el presente documento, la expresión "amplificar", cuando se usa en referencia a un ácido nucleico de una sola cadena, pretende significar la producción de una o más copias del ácido nucleico de una sola cadena, o una porción del mismo.

60 Como se usa en el presente documento, la expresión "fragmento de genoma" pretende significar una molécula de ácido nucleico aislada que tiene una secuencia que es sustancialmente idéntica a una porción de un cromosoma. Un cromosoma se entiende que es un cuerpo que contiene ADN lineal o, a veces, circular de un virus, un organismo procarionta, un núcleo eucariota que contiene la mayor parte o la totalidad de los genes replicados. Una población de fragmentos de genoma puede incluir secuencias idénticas a sustancialmente todo un genoma o una porción del mismo. Un fragmento del genoma puede tener, por ejemplo, una secuencia que sea sustancialmente idéntica a al menos aproximadamente 25, 50, 70, 100, 200, 300, 400, 500, 600, 700, 800, 900 o 1000 o más nucleótidos de un
65 cromosoma. Un fragmento de genoma puede ser ADN, ARN o un análogo del mismo. Los expertos en la materia entenderán que una secuencia de ARN y una secuencia cromosómica de ADN que difieren en la presencia de

uracilos en lugar de timinas son sustancialmente idénticas en cuanto a la secuencia.

Como se usa en el presente documento, el término "nativo", cuando se usa en referencia a un genoma, pretende significar producido por el aislamiento de una célula u otro huésped. El término pretende excluir a los genomas que son producidos por síntesis *in vitro*, la replicación o amplificación.

Como se usa en el presente documento, la expresión "correspondiente a", cuando se usa en referencia a un locus tipificable, pretende significar que tiene una secuencia de nucleótidos que es idéntica o complementaria a la secuencia del locus tipificable, o una porción de diagnóstico del mismo. Los ejemplos de porciones de diagnóstico incluyen, por ejemplo, secuencias de ácido nucleico adyacentes o cercanas al locus tipificable de interés.

Como se usa en el presente documento, el término "multiplexación" pretende significar la realización al mismo tiempo de una pluralidad de ensayos en una o más muestra. La multiplexación puede incluir además la realización al mismo tiempo de una pluralidad de ensayos en cada una de una pluralidad de muestras separadas. Por ejemplo, el número de mezclas de reacción analizadas se puede basar en el número de pocillos de una placa de múltiples pocillos, y el número de ensayos realizados en cada pocillo se puede basar en el número de sondas que entran en contacto con el contenido de cada pocillo. Por lo tanto, las placas de microtitulación de 96 pocillos, 384 pocillos o 1.536 pocillos utilizarán matrices compuestas que comprenden 96, 384 y 1.536 matrices individuales, aunque como será apreciado por los expertos en la materia, no es necesario que todos los pocillos de microtitulación contengan una matriz individual. Dependiendo del tamaño de la placa de microtitulación y del tamaño de la matriz individual, se puede ejecutar de forma simultánea un número muy elevado de ensayos, por ejemplo, usando matrices individuales de 2.000 y una placa de microtitulación de 96 pocillos, se pueden realizar 192.000 experimentos a la vez; las mismas matrices en una placa de microtitulación de 384 producen 768.000 experimentos simultáneos, y una placa de microtitulación de 1536 da 3.072.000 experimentos. Aunque la multiplexión se ha ejemplificado con respecto a placas de microtitulación, se entenderá que se pueden usar otros formatos para la multiplexación incluyendo, por ejemplo, los descritos en el documento US 2002/0102578 A1.

Como se usa en el presente documento, el término "polimerasa" pretende significar una enzima que produce una réplica complementaria de una molécula de ácido nucleico usando el ácido nucleico como una cadena molde. Las ADN polimerasas se unen a la cadena molde y después desplazan la cadena molde, añadiendo nucleótidos al grupo hidroxilo libre en el extremo 3' de una cadena creciente de ácido nucleico. Las ADN polimerasas sintetizan moléculas de ADN complementarias a partir de moldes de ADN o ARN, y las ARN polimerasas sintetizan moléculas de ARN a partir de moldes de ADN (transcripción). Las ADN polimerasas generalmente usan una cadena de ARN o ADN preexistente corta, denominada cebador, para iniciar el crecimiento de la cadena. Algunas ADN polimerasas solo pueden replicar moldes de una sola cadena, mientras que otras ADN polimerasas desplazan la cadena secuencia arriba del sitio donde se están añadiendo bases a una cadena. Como se usa en el presente documento, la expresión "desplazamiento de la cadena", cuando se usa en referencia a una polimerasa, pretende significar que tiene una actividad que elimina una cadena complementaria de una cadena molde que está siendo leída por la polimerasa. Los ejemplos de polimerasas que tienen actividad de desplazamiento de cadena incluyen, sin limitación, el fragmento grande de la polimerasa de Bst (*Bacillus stearothermophilus*), polimerasa de Klenow exo⁻ o la polimerasa T7 exo⁻ de grado de secuenciación.

Además, algunas ADN polimerasas degradan la cadena de en frente de ellas, reemplazándola de manera efectiva con la cadena en crecimiento de detrás. Esto se conoce como una actividad exonucleasa. Algunas ADN polimerasas de uso en el mercado o en el laboratorio se han modificado, bien por mutación o de otra manera, para reducir o eliminar la actividad exonucleasa. También se realizan otras mutaciones o modificaciones con frecuencia para mejorar la capacidad de la ADN polimerasa con el fin de usar nucleótidos no naturales como sustratos.

Como se usa en el presente documento, la expresión "capacidad de procesamiento" se refiere al número de bases, en promedio, añadido a un ácido nucleico que está siendo sintetizado por una polimerasa antes de que la polimerasa se separe del molde de ácido nucleico que se está replicando. Las polimerasas de baja capacidad de procesamiento, como media, sintetizan cadenas de ácidos nucleicos más cortos en comparación con las polimerasas de alta capacidad de procesamiento. Una polimerasa de baja capacidad de procesamiento sintetizará, como media, un ácido nucleico que tendrá una longitud inferior a aproximadamente 100 bases antes de separarse del molde de ácido nucleico que se está replicando. Otros ejemplos de longitudes medias para un ácido nucleico sintetizado por una polimerasa de baja capacidad de procesamiento antes de separarse del molde de ácido nucleico que se está replicando incluyen, sin limitación, menos de aproximadamente 80, 50, 25, 10 o 5 bases.

Como se usa en el presente documento, el término "mellado", cuando se usa en referencia a un ácido nucleico de doble cadena, pretende significar que carece de al menos un enlace covalente de la cadena principal que conecta secuencias adyacentes en una primera cadena y que tiene una segunda cadena complementaria hibridada a ambas secuencias adyacentes de la primera cadena.

Como se usa en el presente documento, la expresión "agente de mellado" pretende significar una entidad física, química o bioquímica que escinde un enlace covalente que conecta las secuencias adyacentes de una primera cadena de ácido nucleico, generando de este modo un producto en el que las secuencias adyacentes están hibridadas con la misma cadena complementaria. Los ejemplos de agentes de mellado incluyen, sin limitación,

endonucleasas de restricción de mellado de una sola cadena que reconocen una secuencia específica tal como N.BstNBI, MthI o proteína genell de bacteriófago f1; ADNsa I; reactivos químicos tales como radicales libres; o ultrasonido.

5 Como se usa en el presente documento, el término "aislado/a", cuando se usa en referencia a una sustancia biológica, pretende significar retirado/a de al menos una parte de las moléculas asociadas con o que se dan con la sustancia en su entorno nativo. Por consiguiente, el término "aislamiento", cuando se usa en referencia a una sustancia biológica, pretende significar la retirada de la sustancia de su entorno nativo o la retirada de al menos una parte de las moléculas asociadas con o que se dan con el ácido nucleico o la sustancia en su entorno nativo. Los ejemplos de sustancias que se pueden aislar incluyen, sin limitación, ácidos nucleicos, proteínas, cromosomas, células, tejidos o similares. Una sustancia biológica aislada, tal como un ácido nucleico, puede estar esencialmente libre de otras sustancias biológicas. Por ejemplo, un ácido nucleico aislado puede estar al menos aproximadamente un 90 %, 95 %, 99 % o 100 % libre de material no nucleotídico asociado de manera natural con el mismo. Un ácido nucleico aislado puede, por ejemplo, estar esencialmente libre de otros ácidos nucleicos, de manera que su secuencia se aumente hasta una fracción significativamente superior del ácido nucleico total presente en la solución de interés que en las células de las que fue tomada la secuencia. Por ejemplo, un ácido nucleico aislado puede estar presente a un nivel 2, 5, 10, 50, 100 o 1.000 veces superior que otros ácidos nucleicos *in vitro* con respecto a los niveles en las células de la que fue tomado. Esto se podría deber a la reducción preferencial en la cantidad de otro ADN o ARN presente, o al aumento preferencial en la cantidad de secuencia de ADN o ARN específica, o a una combinación de ambos parámetros.

Como se usa en el presente documento, el término "complejidad", cuando se usa en referencia a una secuencia de ácido nucleico, pretende significar la longitud total de la secuencia única de un genoma. La complejidad de un genoma puede ser equivalente o inferior a la longitud de una única copia del genoma (es decir, la secuencia haploide). Las estimaciones de la complejidad del genoma pueden ser inferiores a la longitud total si se ajustan por la presencia de secuencias repetidas. Es posible ajustar la longitud de las secuencias repetidas usadas para tales estimaciones para adaptarse a un determinado análisis. Por ejemplo, la complejidad puede ser la suma del número de palabras de la secuencia única en una secuencia del genoma haploide más la longitud de la palabra de la secuencia. Una palabra de la secuencia es una secuencia continua de una longitud definida de al menos 10 nucleótidos. El número de secuencias de repetición, y por lo tanto, la longitud de la secuencia única, en un genoma dependerá de la longitud de la palabra de la secuencia. Más específicamente, a medida que se aumenta la longitud de la palabra de secuencia hasta, por ejemplo, 15, 20, 25, 30, 50, 100 o más nucleótidos, la estimación de la complejidad aumentará, en general, aproximándose al límite superior de la longitud del genoma haploide.

35 Descripción detallada de la invención

Un objeto de la invención es proporcionar un método sensible y exacto para consultar simultáneamente una pluralidad de locus de genes de una muestra de ADN. En particular, el método de la invención se puede usar para determinar el genotipo de un individuo mediante la detección directa de una pluralidad de polimorfismos de un solo nucleótido en una muestra de ADN genómico o ADNc del individuo. Una ventaja de la invención es que se puede obtener una pequeña cantidad de ADN genómico de un individuo, y amplificarlo para obtener una población representativa amplificada de fragmentos de genoma que se pueda consultar en los métodos de la invención. Por lo tanto, los métodos son particularmente útiles para la genotipificación del ADN genómico obtenido de muestras de tejido relativamente pequeñas tales como una biopsia o una muestra archivada. En general, los métodos se usarán para amplificar un número relativamente pequeño de copias del genoma molde. En realizaciones particulares, se puede obtener una muestra de ADN genómico de una sola célula y genotipificarlo.

Una ventaja adicional de la detección directa de locus genéticos en los métodos de la invención es que no es necesario amplificar un fragmento de ADN genómico diana una vez que haya sido capturado por una sonda apropiada. Por lo tanto, los métodos pueden proporcionar la ventaja de reducir o evitar la necesidad de medios complejos y costosos para la detección tras la captura. Si hay suficiente ADN presente, la detección de locus tipificables se puede llevar a cabo mediante una técnica que no requiera la amplificación de una diana capturada tal como la extensión de una sola base (SBE) o la extensión de cebadores específicos del alelo (ASPE). Otros métodos de detección directa incluyen la ligadura, la ligadura por extensión, el ensayo Invader, la hibridación con una secuencia complementaria marcada, o similares. Tales técnicas de detección directa se pueden llevar a cabo, por ejemplo, directamente en un complejo de sonda-diana capturado como se expone a continuación. Aunque los métodos de detección basados en la amplificación de dianas no son necesarios en los métodos de la invención, los métodos son compatibles con una variedad de métodos de detección basados en la amplificación tales como Invader, basado en la PCR, o tecnologías basadas en ensayos de ligación de oligonucleótidos (basadas en OLA) que, si se desea, se pueden usar.

La invención proporciona métodos de amplificación de todo el genoma que se pueden usar para amplificar el ADN genómico antes de la evaluación genética tal como la detección de locus tipificables en el genoma. Los métodos de amplificación de genomas completos de la invención se pueden usar para aumentar la cantidad de ADN genómico sin comprometer la calidad ni la representación de ninguna secuencia dada. Por lo tanto, los métodos se pueden usar para amplificar una cantidad relativamente pequeña de ADN genómico de una manera independiente de la

secuencia para proporcionar niveles del ADN genómico que se puedan genotipificar. Sorprendentemente, es posible amplificar un genoma complejo con una polimerasa de baja capacidad de procesamiento para obtener una población de fragmentos de genoma que sea representativa del genoma, que tenga una alta complejidad y que contenga fragmentos que tengan un tamaño conveniente para la hibridación a una matriz de ácido nucleico típica.

5 Como se expone más detalladamente a continuación, se puede incubar una población representativa compleja de fragmentos de genoma con una pluralidad de sondas, y detectarse específicamente una fracción relativamente pequeña de estos fragmentos, que tenga los locus de interés, a pesar de la presencia de una cantidad sustancialmente grande de otras secuencias genómicas presentes en la población de fragmentos. Por otra parte, la
10 detección específica para tales representaciones complejas se puede producir, incluso si la hibridación de las sondas se lleva a cabo con grandes cantidades y altas concentraciones de las poblaciones de fragmentos de genoma. Por lo tanto, una ventaja de la invención es que la genotipificación del genoma completo se puede llevar a cabo en presencia de un fondo de ADN genómico de alta complejidad.

15 Además, la amplificación de ADN genómico en los métodos divulgados en el presente documento no requiere la reacción en cadena de la polimerasa. En concreto, la amplificación se puede llevar a cabo de manera que las secuencias se amplifiquen varias veces en condiciones isotérmicas. Por lo tanto, aunque se puede usar una etapa de temperatura elevada, por ejemplo, para desnaturalizar inicialmente un molde de ADN genómico, no es necesario usar ciclos de temperatura. Por consiguiente, no será necesario usar los repetidos aumentos de la temperatura, que
20 normalmente se usan para desnaturalizar híbridos, ni los repetidos retornos a las temperaturas de hibridación.

Tras la captura y la separación de los locus tipificables en una matriz, los locus tipificables individuales se pueden marcar *in positus* (en su lugar) a través de un ensayo de detección posterior tal como ASPE o SBE. Por lo tanto, se puede capturar una población de fragmentos de genoma obtenida por amplificación del genoma completo con una
25 polimerasa de baja capacidad de procesamiento mediante una matriz de sondas, y determinarse el genotipo del genoma en base a los locus tipificables detectados individualmente en cada sonda como se establece a continuación y se demuestra en los ejemplos. Un enfoque de genotipificación *in positus* tiene ventajas notables, ya que permite la multiplexación extensa del ensayo cuando se desee.

30 El uso de la tecnología de matrices de ADN de alta densidad para la detección de locus tipificables en un genoma completo o una muestra de ADN complejo, tal como una muestra de ADNc, puede verse facilitada por los métodos de amplificación de la invención, debido a que el método puede producir un número de copias de locus tipificables, o secuencias complementarias a locus tipificables para modificar a escala en una proporción relativa hasta su
35 representación en la muestra de molde. El mantenimiento de la representación relativamente uniforme es ventajoso en muchas aplicaciones, porque si algunas zonas del genoma que contiene marcadores genéticos específicos no se replica fielmente, no se detectará en un ensayo ajustado para la amplificación media. La invención puede detectar por escalado un número deseado de locus tipificables simultánea o secuencialmente, según se desee. Los métodos se pueden usar para detectar simultáneamente al menos 10 locus tipificables, al menos 100, 1.000, 1×10^4 , 1×10^5 ,
40 1×10^6 , 1×10^7 locus tipificables o más. Del mismo modo, estos números de locus tipificables se pueden determinar en un formato secuencial cuando se desee. Por lo tanto, la invención se puede usar para determinar el genotipo de los individuos a escala de un amplio genoma si se desea.

Los métodos de amplificación de genomas completos de la invención y los métodos de genotipificación de genomas completos de la invención son útiles, solos o en combinación, en una serie de aplicaciones, incluyendo, por ejemplo,
45 análisis de haplotipos de un solo espermatozoide, genotipificación de un gran número de individuos en un formato de alto rendimiento o identificación de nuevos haplotipos. Además, la invención reduce la cantidad de muestra de ADN o ARN necesaria en muchos ensayos de matriz actuales. Es más, la mejora de la sensibilidad de las matrices disponibles con la invención puede conducir a menor necesidad de muestras, una mejor puntuación del método LOD y un mayor rango dinámico.

50 La invención se puede usar para identificar nuevos marcadores o haplotipos que diagnostiquen rasgos tales como los enumerados anteriormente. Tales estudios se pueden llevar a cabo mediante la comparación de genotipos para grupos de individuos que tengan un rasgo o un conjunto de rasgos con un grupo de control que carezca del rasgo en base a la expectativa de que habrá frecuencias más altas de los componentes genéticos que contribuyen en un
55 grupo de personas con un rasgo compartido, tal como una determinada enfermedad o respuesta a un fármaco, vacuna, agente patógeno o factor ambiental, que en un grupo de personas similares sin la enfermedad o la respuesta. Por consiguiente, los métodos de la invención se pueden usar para encontrar regiones cromosómicas que tengan diferentes distribuciones de haplotipos en los dos grupos de personas, las que tienen una enfermedad o respuesta y las que no la tienen. A continuación, se puede estudiar cada región con más detalle para descubrir las
60 variantes en las que los genes de la región contribuyen a la enfermedad o la respuesta, lo que lleva a intervenciones más eficaces. Esto también puede permitir el desarrollo de pruebas para predecir qué fármacos o vacunas son eficaces en individuos con determinados genotipos de los genes que afectan el metabolismo de fármacos. Por lo tanto, la invención se puede usar para determinar el genotipo de un individuo basándose en la identificación de qué marcadores genéticos se encuentran en el genoma del individuo. El conocimiento del genotipo de un individuo se
65 puede usar para determinar una variedad de rasgos tales como la respuesta a factores ambientales, la susceptibilidad a una infección, la eficacia de determinados fármacos o vacunas, o el riesgo de tener respuestas

adversas a fármacos o vacunas.

La divulgación se ejemplifica en el presente documento con respecto a la amplificación y/o la detección de locus tipificables para un genoma completo. Los expertos en la materia reconocerán, a partir de las enseñanzas del presente documento, que los métodos también se pueden usar con otras muestras de ácidos nucleicos complejos que incluyen, por ejemplo, una fracción de un genoma, tal como un cromosoma o un subconjunto de cromosomas; una muestra que tenga múltiples genomas diferentes, tal como una muestra de biopsia que tenga ADN genómico de un huésped, así como uno o más parásitos o una muestra ecológica que tenga múltiples organismos de un entorno en particular; o incluso ADNc o una representación de ADNc amplificado. Por consiguiente, los métodos se pueden usar para caracterizar los locus tipificables que se encuentran en una fracción de un genoma o en una muestra mixta de genoma. La divulgación proporciona un método de detección de uno o varios locus tipificables contenidos en un genoma dado. El método incluye las etapas de (a) proporcionar una población representativa amplificada de fragmentos de genoma que tenga tales locus tipificables; (b) poner en contacto los fragmentos de genoma con una pluralidad de sondas de ácido nucleico que tengan secuencias correspondientes a los locus tipificables en condiciones en las que se formen híbridos de sonda-fragmento; y (c) detectar locus tipificables de los híbridos de sonda-fragmento. En realizaciones particulares, estas sondas de ácido nucleico son, como máximo, de 125 nucleótidos de longitud. La Figura 1 muestra una visión general de un procedimiento ilustrativo de detección de locus tipificables de un genoma. Como se muestra en la Figura 1, se puede obtener una población de fragmentos de genoma de un genoma, desnaturalizarla y ponerla en contacto con una matriz de sondas de ácido nucleico, cada una con una secuencia que sea complementaria a un determinado locus tipificable del genoma. Los fragmentos de genoma que tienen locus tipificables representados en las sondas se capturan como híbridos de sonda-fragmento en ubicaciones diferenciadas de la matriz, mientras que otros fragmentos que carecen de los locus de interés se mantendrán en solución a granel. Los híbridos de sonda-fragmento se pueden detectar mediante la adición mediada por enzimas de un resto de detección (que se denomina resto de señal en la Figura 1) a la sonda. En la realización ilustrativa de la Figura 1, una polimerasa añade selectivamente un nucleótido marcado con biotina a las sondas en híbridos de sonda-fragmento. Entonces, se pueden detectar las sondas biotiniladas, por ejemplo, poniendo en contacto una avidina marcada con fluorescencia con la matriz en condiciones en las que las sondas biotiniladas se unan selectivamente, y detectar las ubicaciones de la matriz que presentan fluorescencia. Basándose en las secuencias conocidas para sondas en cada lugar, se puede determinar la presencia de determinados locus tipificables.

Un método de la invención se pueden usar para amplificar el ADN genómico (ADNg) o detectar locus tipificables de un genoma de cualquier organismo. Los métodos son idealmente adecuados para la amplificación y el análisis de grandes genomas tales como los que se encuentran normalmente en los organismos unicelulares y multicelulares eucariotas. El ADNg eucariótico ilustrativo que se puede usar en un método de la invención incluye, sin limitación, el obtenido de un mamífero tal como un roedor, ratón, rata, conejo, cobaya, ungulado, caballo, oveja, cerdo, cabra, vaca, gato, perro, primate, ser humano o primate no humano; una planta tal como *Arabidopsis thaliana*, maíz (*Zea mays*), sorgo, avena (*Oryza sativa*), trigo, arroz, canola o soja; un alga tal como *Chlamydomonas reinhardtii*, un nematodo tal como *Caenorhabditis elegans*; un insecto tal como *Drosophila melanogaster*, mosquito, mosca de la fruta, abeja de la miel o una araña; un pez tal como el pez cebra (*Danio rerio*); un reptil; un anfibio tal como una rana o *Xenopus laevis*; *Dictyostelium discoideum*; un hongo tal como *Pneumocystis carinii*, *Takifugu rubripes*, levadura, *Saccharomyces cerevisiae* o *Schizosaccharomyces pombe*; o un *plasmodium falciparum*. Un método de la invención también se puede usar para detectar locus tipificables de genomas más pequeños tales como los de un procarionta tal como una bacteria, *Escherichia coli*, estafilococos o *Mycoplasma pneumoniae*; una arqueobacteria; un virus tal como el virus de la hepatitis C o el virus de la inmunodeficiencia humana; o un viroide.

Un ADN genómico usado en la invención puede tener uno o más cromosomas. Se puede usar, por ejemplo, un ADN genómico procarionta que incluya un cromosoma. Como alternativa, en un método de la invención, se puede usar un ADN genómico eucariota que incluya una pluralidad de cromosomas. Por lo tanto, los métodos se pueden usar, por ejemplo, para amplificar o detectar locus tipificables de un ADN genómico que tenga n igual a 2 o más, 4 o más, 6 o más, 8 o más, 10 o más, 15 o más, 20 o más, 23 o más, 25 o más, 30 o más, o 35 o más cromosomas, donde n es el número de cromosomas haploide y el número de cromosomas diploide es $2n$. El tamaño de un ADN genómico usado en un método de la invención también se puede medir de acuerdo con el número de pares de bases o la longitud de nucleótidos del complemento cromosómico. Las estimaciones del tamaño ilustrativas para algunos de los genomas que son útiles en la invención son de aproximadamente 3,1 Gpb (ser humano), 2,7 Gpb (ratón), 2,8 Gpb (rata), 1,7 Gpb (pez cebra), 165 Gpb (mosca de la fruta), 13,5 Gpb (*S. cerevisiae*), 390 Gpb (hongo), 278 Gpb (mosquito) o 103 Gpb (*C. elegans*). Los expertos en la materia reconocerán que, en un método de la invención, se pueden usar genomas que tienen tamaños distintos a los ejemplificados anteriormente, incluyendo, por ejemplo, genomas más pequeños o más grandes.

El ADN genómico se puede aislar de una o más células, fluidos corporales o tejidos. Se pueden usar métodos conocidos para obtener un fluido corporal tal como sangre, sudor, lágrimas, linfa, orina, saliva, semen, líquido cefalorraquídeo, heces o líquido amniótico. De igual manera, se pueden usar métodos de biopsia conocidos para obtener células o tejidos tales como hisopo bucal, enjuague bucal, extirpación quirúrgica, biopsia por aspiración o similar. El ADN genómico también se puede obtener de una o más células o tejidos en cultivo primario, en una línea de células propagadas, una muestra archivada fijada, muestra forense o muestra arqueológica.

Los ejemplos de tipos de células de los que se puede obtener ADNg en un método de la invención incluyen, sin limitación, una célula sanguínea tal como un linfocito B, linfocito T, leucocito, eritrocito, macrófago o neutrófilo; una célula muscular tal como una célula esquelética, célula del músculo liso o célula del músculo cardíaco; células germinales tales como un espermatozoide u un óvulo; célula epitelial; células del tejido conjuntivo tales como adipocito, fibroblasto u osteoblasto; neurona; astrocito; célula estromal; célula renal; célula pancreática; célula hepática; o queratinocito. Una célula de la que se obtenga ADNg puede estar en un determinado nivel de desarrollo incluyendo, por ejemplo, una célula madre hematopoyética o una célula surgida de una célula madre hematopoyética tal como un glóbulo rojo, linfocito B, linfocito T, célula asesina natural, neutrófilo, basófilo, eosinófilo, monocito, macrófago o plaqueta. Otras células incluyen una célula estromal de médula ósea (célula madre mesenquimatosas) o una célula que se desarrolla de la misma tal como una célula ósea (osteocito), células del cartílago (condrocito), célula de grasa (adipocito) u otros tipos de células del tejido conjuntivo tales como la encontrada en los tendones; célula madre neural o una célula a la que dé lugar, incluyendo, por ejemplo, una célula nerviosa (neurona), astrocito o oligodendrocito; célula madre epitelial o una célula surgida de una célula madre epitelial tal como una célula de absorción, célula calciforme, célula de Paneth o célula completoendocrina; célula madre cutánea; células madre epidérmica; o célula madre folicular. En general, se puede usar cualquier tipo de célula madre, incluyendo, sin limitación, una célula madre embrionaria, célula madre adulta o célula madre pluripotente.

Una célula de la que se obtiene una muestra de ADNg para su uso en la invención puede ser una célula normal o una célula que muestre uno o más síntomas de una determinada enfermedad o afección. Por lo tanto, un ADNg usado en un método de la invención se puede obtener de una célula cancerosa, célula neoplásica, célula necrótica, o similares. Los expertos en la materia conocerán o serán capaces de determinar fácilmente métodos para aislar ADNg de una célula, fluido o tejido usando métodos conocidos en la técnica tales como los descritos en Sambrook *et al.*, "Molecular Cloning: A Laboratory Manual", III edición, Cold Spring Harbor Laboratory, Nueva York (2001) o en Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Md. (1998).

Un método de la invención puede incluir además las etapas de aislar un determinado tipo de célula o tejido. Los ejemplos de métodos que se pueden usar en un método de la invención para aislar una determinada célula de otras células de una población incluyen, pero sin limitación, clasificación de células activadas por fluorescencia (FACS) tal como se describe, por ejemplo, en Shapiro, "Practical Flow Cytometry", III edición Wiley-Liss; (1995), centrifugación en gradiente de densidad o separación manual usando métodos de micromanipulación con ayuda del microscopio. Los dispositivos de separación de células ilustrativos que son útiles en la invención incluyen, sin limitación, un sistema de decantación por centrifugación JE-6 Beckman, clasificador de células por citómetro de flujo controlado por ordenador Beckman Coulter EPICS ALTRA, citómetro de flujo modular de Cytomation, Inc., contador y sistema canalizador Coulter, aparato de gradiente de densidad, citocentrífugador, centrifugador Beckman J-6, clasificador celular de láser dual EPICS V o citómetro de flujo EPICS PROFILE. Los tejidos o poblaciones de células también se pueden retirar mediante técnicas quirúrgicas. Por ejemplo, se pueden extraer un tumor o células de un tumor de un tejido mediante métodos quirúrgicos o, por el contrario, se pueden retirar células no cancerosas de las proximidades de un tumor. Usando métodos tales como los expuestos más detalladamente a continuación, se puede usar la invención para comparar locus tipificables de células diferentes, incluyendo, por ejemplo, células cancerosas y no cancerosas aisladas del mismo individuo o de individuos diferentes.

Un ADNg se puede preparar para su uso en un método de la invención mediante la lisis de una célula que contenga el ADN. Por lo general, las células se lisan en condiciones que preserven sustancialmente la integridad del ADNg de la célula. En particular, se puede usar la exposición de una célula a pH alcalino para lisar una célula en un método de la invención, causando relativamente poco daño al ADNg. Se puede usar cualquiera de una variedad de compuestos básicos para la lisis incluyendo, por ejemplo, hidróxido de potasio, hidróxido de sodio y similares. Además, se puede obtener ADNg relativamente no dañado de una célula lisada por una enzima que degrade la pared celular. Las células que carecen de pared celular, bien de forma natural o debido a la eliminación enzimática, también se pueden lisar mediante la exposición a estrés osmótico. Otras condiciones que se pueden usar para lisar una célula incluyen la exposición a detergentes, la disrupción mecánica, calor de sonicación, diferencial de presión tal como en un dispositivo de prensa francesa u homogeneización de Dounce. Se pueden incluir agentes de estabilización del ADNg en un lisado celular o muestra aislada de ADNg incluyendo, por ejemplo, inhibidores de nucleasa, agentes quelantes, tampones salinos y similares. Los métodos para la lisis de una célula para obtener ADNg se pueden llevar a cabo en condiciones conocidas en la técnica como se describe, por ejemplo, en Sambrook *et al.*, *supra* (2001) o en Ausubel *et al.*, *supra*, (1998).

En realizaciones particulares de la invención, es posible amplificar directamente o detectar sin mayor aislamiento del ADNg un lisado celular bruto que contenga ADNg. Como alternativa, un ADNg se puede aislar adicionalmente de otros componentes celulares antes de la amplificación o detección. Por consiguiente, un método de detección o amplificación de la invención se puede llevar a cabo en ADNg purificado o parcialmente purificado. El ADN genómico se puede aislar usando métodos conocidos, incluyendo, por ejemplo, la extracción líquida de fase, precipitación, extracción en fase sólida, cromatografía y similares. Tales métodos a menudo se denominan minipreparaciones, y se describen, por ejemplo, en Sambrook *et al.*, *supra*, (2001) o en Ausubel *et al.*, *supra*, (1998) o se encuentran disponibles en diferentes proveedores comerciales, incluyendo, por ejemplo, Qiagen (Valencia, CA) o Promega (Madison, WI).

Se puede proporcionar una población representativa amplificada de fragmentos de genoma mediante la amplificación de un genoma nativo en condiciones que reproduzcan un molde de ADN genómico (ADNg) para producir una o más copias en las que la proporción relativa de cada secuencia de copiado es sustancialmente igual a su proporción en el ADNg original. Por lo tanto, un método de la invención puede incluir una etapa de amplificar de manera representativa un genoma nativo. En la invención, se puede usar cualquiera de una variedad de métodos de replicación del ADN genómico de una manera independiente de la secuencia.

Un método de la invención se puede usar para producir una población representativa amplificada de fragmentos de genoma de un pequeño número de copias del genoma. Por consiguiente, se pueden genotipificar o evaluar muestras de tejido pequeñas u otras muestras que tengan relativamente pocas células, por ejemplo, debido a la baja abundancia, las limitaciones de la biopsia o el alto coste, a escala de todo el genoma. La invención se puede usar para producir una población representativa amplificada de fragmentos de genoma de una sola copia del genoma nativo obtenida, por ejemplo, de una sola célula. En otras realizaciones ilustrativas de la invención, se puede producir una población representativa amplificada de fragmentos de genoma a partir de un mayor número de copias de un genoma nativo, incluyendo, pero sin limitación, aproximadamente 1.000 copias (para un genoma humano, aproximadamente 3 nanogramos de ADN) o menos, 10.000 copias o menos, 1×10^5 copias (para un genoma humano, aproximadamente 300 nanogramos de ADN) o menos, 5×10^5 copias o menos, 1×10^6 copias o menos, 1×10^8 copias o menos, 1×10^{10} copias o menos o 1×10^{12} copias o menos.

Una muestra de ADN que se amplifica de manera representativa en la invención puede ser un genoma, tal como los expuestos anteriormente u otros moldes de ADN tales como ADN mitocondrial o algún subconjunto de ADN genómico. Un ejemplo no limitante de subconjunto de ADN genómico es un cromosoma en particular o una región de un cromosoma en particular. En general, un método de amplificación usado en la invención se puede llevar a cabo usando al menos un cebador de ácido nucleico que se hibride con un molde de ácido nucleico para formar un complejo de hibridación, trifosfatos de nucleótido (PNT) y una polimerasa que modifique el cebador mediante la reacción del PNT con el hidroxilo 3' del cebador, replicando de ese modo al menos una parte del molde. Por ejemplo, los métodos basados en la PCR, generalmente, utilizan un molde de ADN, dos cebadores, varios dNTP y una ADN polimerasa. Por lo tanto, en un método típico de amplificación de genomas completos de la invención, se incubaba una muestra de ADN genómico con una mezcla de reacción que incluye componentes de amplificación tales como los expuestos anteriormente, y se forma una población representativa amplificada de fragmentos de genoma.

Un cebador usado en un método de la invención puede tener cualquiera de una variedad de composiciones o tamaños, siempre que tenga la capacidad de hibridarse con un molde de ácido nucleico con especificidad de secuencia y pueda participar en la replicación del molde. Por ejemplo, un cebador puede ser un ácido nucleico que tenga una estructura nativa o un análogo de la misma. Un ácido nucleico con una estructura nativa, generalmente, tiene una cadena principal que contiene enlaces de tipo fosfodiéster y puede ser, por ejemplo, ácido desoxirribonucleico o ácido ribonucleico. Una estructura análoga puede tener una cadena principal alternativa que incluya, sin limitación, fosforamida (véase, por ejemplo, *Beaucage et al., Tetrahedron* 49(10):1925 (1993) y las referencias del mismo; *Letsinger, J. Org. Chem.*, 35:3800 (1970), *Sprinzi et al., Eur. J. Biochem.* 81:579 (1977); *Letsinger et al., Nucl. Acids Res.*, 14:3487 (1986); *Sawai et al., Chem. Lett.*, 805 (1984), *Letsinger et al., J. Am. Chem. Soc.* 110:4470 (1988); y *Pauwels et al., Chemica Scripta* 26:141 91986)), fosforotioato (véase, por ejemplo, *Mag et al., Nucleic Acids Res.* 19:1437 (1991); y la patente de EE.UU. N° 5.644.048), fosforoditioato (véase, por ejemplo, *Briu et al., J. Am. Chem. Soc.* 111:2321 (1989), enlaces de tipo O-metilfosfoamidita (véase, por ejemplo, *Eckstein, "Oligonucleotides and Analogues: A Practical Approach", Oxford University Press*) y las cadenas principales y enlaces de ácidos nucleicos peptídicos (véase, por ejemplo, *Egholm, J. Am. Chem. Soc.* 114:1895 (1992); *Meier et al., Chem. Int. Ed. Engl.* 31:1008 (1992); *Nielsen, Nature*, 365:566 (1993); *Carlsson et al., Nature* 380:207 (1996)). Otras estructuras análogas incluyen aquellas con cadenas principales positivas (véase, por ejemplo, *Denpcy et al., Proc. Natl. Acad. Ciencia.* EE.UU. 92:6097 (1995); cadenas principales no iónicas (véase, por ejemplo, la patente de EE.UU. N° 5.386.023, 5.637.684, 5.602.240, 5.216.141 y 4.469.863; *Kiedrowski et al., Angew. Chem. Intl. Ed. inglesa* 30:423 (1991); *Letsinger et al., J. Am. Chem. Soc.* 110:4470 (1988); *Letsinger et al., "Nucleoside & Nucleotide"* 13:1597 (1994); capítulos 2 y 3, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, ed. Y. S. Sanghui y P. Dan Cook; *Mesmaeker et al., Bioorganic & Medicinal Chem. Left.* 4:395 (1994); *Jeffs et al., J. Biomolecular NMR* 34:17 (1994); *Tetrahedron Lett.* 37:743 (1996)) y cadenas principales no ribosomales, incluyendo, por ejemplo, las descritas en las patentes de EE.UU. N° 5.235.033 y 5.034.506, y en los capítulos 6 y 7, *ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research"*, ed. Y. S. Sanghui y P. Dan Cook. Las estructuras análogas que contienen uno o más azúcares carbocíclicos también son útiles en los métodos y se describen, por ejemplo, en *Jenkins et al., Chem. Soc. Rev.* (1995) p.169-176. Otras varias estructuras análogas que son útiles en la invención se describen en *Rawls, C & E News*, 2 de junio 1997, página 35.

Un ejemplo adicional de un ácido nucleico con una estructura análoga que es útil en la invención es un ácido nucleico peptídico (ANP). La cadena principal de un ANP es sustancialmente no iónica en condiciones neutras, en contraste con la cadena principal del fosfodiéster altamente cargada de los ácidos nucleicos naturales. Esto proporciona dos ventajas no limitantes. En primer lugar, la cadena principal de ANP presenta una mejor cinética de hibridación. En segundo lugar, los ANP tienen cambios mayores en la temperatura de fusión (T_m) para los pares de bases desapareados frente a los pares de bases perfectamente apareados. El ADN y ARN suelen presentar una caída de T_m de 2-4 °C por un desapareamiento interno. Con la cadena principal no iónica de los ANP, la caída está

más cerca de los 7-9 °C. Esto puede proporcionar una mejor diferenciación de las secuencias. Del mismo modo, debido a su naturaleza no iónica, la hibridación de las bases unidas a estas cadenas principales es relativamente insensible a la concentración de sal.

- 5 Un ácido nucleico útil en la invención puede contener un resto de azúcar no natural en la cadena principal. Las modificaciones de azúcar ilustrativas incluyen, pero sin limitación, modificaciones en 2' tales como la adición de halógeno, alquilo, alquilo sustituido, alcarilo, aralculo, O-alcarilo u O-aralquilo, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂, heterocicloalculo, heterocicloalcarilo, aminoalquilamino, polialquilamino, sililo sustituido, y similares. También se pueden hacer modificaciones similares en otras posiciones del azúcar, en particular, en la posición 3' del azúcar en el nucleótido del extremo 3' o en oligonucleótidos ligados a 2'-5' y en la posición 5' del nucleótido del extremo 5'.

- 15 Un ácido nucleico usado en la invención también puede incluir bases nativas o no nativas. A este respecto, un ácido desoxirribonucleico nativo puede tener una o más bases seleccionadas del grupo que consiste en adenina, timina, citosina o guanina, y un ácido ribonucleico pueden tener una o más bases seleccionadas del grupo que consiste en uracilo, adenina, citosina o guanina. Las bases no nativas a modo de ejemplo que pueden estar incluidas en un ácido nucleico, si tiene una cadena principal o una estructura análoga nativa, incluyen, sin limitación, inosina, xantanina, hipoxantanina, isocitosina, isoguanina, 5-metilcitosina, 5-hidroximetil-citosina, 2-aminoadenina, 6-metiladenina, 6-metil-guanina, 2-propil-guanina, 2-propil-adenina, 2-tioliracilo, 2-tiotimina, 2-tiocitosina, 15-halouracilo, 15-halocitosina, 5-propinil-uracilo, 5-propinil-citosina, 6-azo-uracilo, 6-azo-citosina, 6-azo-timina, 5-uracilo, 4-tiouracilo, 8-halo-adenina o guanina, 8-amino-adenina o guanina, 8-tiol-adenina o guanina, 8-tio-alquil-adenina o guanina, 8-hidroxil-adenina o guanina, uracilo o citosina 5-halo sustituida, 7-metilguanina, 7-metiladenina, 8-azaguanina, 8-azaadenina, 7-desazaguanina, 7-desazaadenina, 3-desazaguanina, 3-desazaadenina o similares. Una realización particular puede utilizar isocitosina e isoguanina en un ácido nucleico con el fin de reducir la hibridación inespecífica, como se describe, en general, en la patente de EE.UU. N° 5.681.702.

- 30 Una base no nativa usada en un ácido nucleico de la invención puede tener actividad de apareamiento de bases universal, donde es capaz de aparearse con cualquier otra base de origen natural. Las bases a modo de ejemplo que tienen actividad de apareamiento de bases universal incluyen 3-nitropirrol y 5-nitroindol. Otras bases que se pueden usar incluyen las que tienen actividad de apareamiento con un subconjunto de las bases de origen natural, tales como inosina, que se aparea con citosina, adenina o uracilo.

- 35 Un ácido nucleico que tiene una estructura modificada o análoga se puede usar en la invención, por ejemplo, para facilitar la adición de marcadores, o para aumentar la estabilidad o la semivida de la molécula en condiciones de amplificación o en otras condiciones usadas de acuerdo con la invención. Como los expertos en la materia apreciarán, en la invención, se pueden usar uno o más de los ácidos nucleicos descritos anteriormente, incluyendo, por ejemplo, en forma de una mezcla que incluya moléculas con estructuras nativas o análogas. Además, un cebador de ácido nucleico usado en la invención puede tener una estructura deseada para una determinada técnica de amplificación usada en la invención tal como las expuestas más adelante.

- 40 En realizaciones particulares, un ácido nucleico útil en la invención puede incluir un resto de detección. Un resto de detección se puede usar, por ejemplo, para detectar uno o más miembros de una población representativa amplificada de fragmentos de genoma usando métodos tales como los expuestos más adelante. Un resto de detección puede ser un marcador primario que sea directamente detectable o un marcador secundario que se pueda detectar indirectamente, por ejemplo, a través de la interacción directa o indirecta con un marcador primario. Los marcadores primarios a modo de ejemplo incluyen, sin limitación, un marcador isotópico tal como un isótopo radiactivo o pesado natural no abundante; cromóforo; luminóforo; fluoróforo; agente calorimétrico; sustancia magnética, material rico en electrones tal como un metal; marcador electroquimioluminiscente tal como Ru(bpy)₃²⁺; o resto que se pueda detectar basándose en una característica magnética nuclear, paramagnética, eléctrica, relación de carga con respecto a la masa o térmica. Los fluoróforos que son útiles en la invención incluyen, por ejemplo, complejos de lantánidos fluorescentes, incluyendo los de europio y terbio, fluoresceína, rodamina, tetrametilrodamina, eosina, eritrosina, cumarina, metil-cumarinas, pireno, verde malaquita, Cy3, Cy5, estilbeno, amarillo Lucifer, Azul Cascade®, Rojo Texas, colorantes de Alexa, ficoeritina, BODIPY y otros conocidos en la técnica tales como los descritos en Haugland, "Molecular Probes Manual", (Eugene, OR), VI edición; El catálogo Synthegen (Houston, TX), Lakowicz, "Principles of Fluorescence Spectroscopy", II ed., Plenum Press, Nueva York (1999), o el documento WO 98/59066. Los marcadores también pueden incluir enzimas tales como peroxidasa de rábano picante o fosfatasa alcalina, o partículas tales como partículas magnéticas o nanopartículas codificadas ópticamente.

- 60 Los marcadores secundarios a modo de ejemplo son restos de unión. Un resto de unión puede estar unido a un ácido nucleico para permitir la detección o el aislamiento del ácido nucleico a través de afinidad específica por un receptor. La afinidad específica entre dos elementos de unión pretende significar la unión preferencial entre uno y otro en comparación con la unión de un elemento con otros componentes o contaminantes del sistema. Los pares de unión que se unen específicamente permanecen normalmente unidos en las condiciones de detección o de separación descritas en el presente documento, incluyendo las etapas de lavado para eliminar la unión inespecífica. Dependiendo de las condiciones de unión usadas en particular, las constantes de disociación del par pueden ser,

por ejemplo, inferiores a aproximadamente 10^{-4} , 10^{-5} , 10^{-6} , 10^{-7} , 10^{-8} , 10^{-9} , 10^{-10} , 10^{-11} o 10^{-12} M⁻¹.

Los ejemplos de pares de restos de unión y receptores que se pueden usar en la invención incluyen, sin limitación, antígeno e inmunoglobulina o fragmentos activos de la misma, tales como los Fab; inmunoglobulina e inmunoglobulina (o fragmentos activos, respectivamente); avidina y biotina o análogos de las mismas que tengan especificidad por la avidina tales como imino-biotina; estreptavidina y biotina, o análogos de las mismas que tengan especificidad por la estreptavidina tales como imino-biotina; hidratos de carbono y lectinas, y otras proteínas conocidas y sus ligandos. Se entenderá que cualquiera de las partes de los pares anteriormente descritos puede estar unida a un ácido nucleico, y detectarse o aislarse basándose en la unión con su respectiva pareja. Se entenderá, además, que los diversos restos que se pueden unir a un ácido nucleico pueden funcionar como marcadores primarios y secundarios en un método de la invención. Por ejemplo, streptavidina-ficoeritrina se puede detectar como un marcador primario debido a la fluorescencia del resto de ficoeritrina o se puede detectar como un marcador secundario debido a su afinidad por anticuerpos anti-estreptavidina, como se expone más detalladamente a continuación en lo que se refiere a los métodos de amplificación de señales.

En una realización particular, el marcador secundario puede ser un resto químicamente modificable. En la presente realización, se pueden incorporar marcadores que tienen grupos funcionales reactivos a un ácido nucleico. Posteriormente, se puede hacer reaccionar covalentemente el grupo funcional con un marcador primario. Los grupos funcionales adecuados incluyen, pero sin limitación, grupos amino, grupos carboxi, grupos maleimida, grupos oxo y grupos tiol. Los restos de unión pueden ser particularmente útiles cuando se unen a los cebadores usados para la amplificación de un ADN_g, porque una población representativa amplificada de fragmentos de genoma producida con tales cebadores se puede unir a una matriz a través de dichos restos de unión. Además, los restos de unión pueden ser útiles para la separación de fragmentos amplificados de otros componentes de una reacción de amplificación, para concentrar la población representativa amplificada de fragmentos de genoma o para detectar uno o más miembros de una población representativa amplificada de fragmentos de genoma cuando se une a sondas de captura en una matriz. Los métodos de separación y detección ilustrativos para los ácidos nucleicos que tienen unidos restos de unión se exponen detalladamente más adelante.

Se puede unir un resto de unión, resto de detección o cualquier otro resto útil a un ácido nucleico tal como un fragmento de genoma amplificado usando métodos conocidos en la técnica. Por ejemplo, un cebador usado para amplificar un ácido nucleico puede incluir el resto unido a una base, ribosa, fosfato o estructura análoga en un ácido nucleico o análogo del mismo. En realizaciones particulares, se puede incorporar un resto usando nucleósidos modificados que se añadan a una cadena de nucleótidos en crecimiento, por ejemplo, durante las etapas de amplificación o detección. Los nucleósidos se pueden modificar, por ejemplo, en la base o la ribosa, o estructuras análogas de un análogo de ácido nucleico. Por lo tanto, un método de la invención puede incluir una etapa de marcaje de fragmentos de genoma para producir una población representativa amplificada de fragmentos de genoma que tenga una o más de las modificaciones establecidas anteriormente. Un cebador de ácido nucleico usado para amplificar un ADN_g en un método de la invención puede incluir una secuencia complementaria que sea de cualquier longitud capaz de unirse a un molde de ADN_g con la estabilidad y la especificidad suficientes para cebar la actividad de replicación de la polimerasa. La secuencia complementaria puede incluir la totalidad o una parte de un cebador usado para la amplificación. La longitud de la secuencia complementaria de un cebador usado para la amplificación en un método de la invención será generalmente inversamente proporcional a la distancia entre los sitios de cebado de un molde de ADN_g. Por lo tanto, la amplificación se puede llevar a cabo con cebadores que tengan secuencias complementarias relativamente cortas que incluyan, por ejemplo, como máximo, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 nucleótidos de longitud.

Los expertos en la materia reconocerán que la especificidad de la hibridación se aumenta generalmente a media que aumenta la longitud del cebador de ácido nucleico. Por lo tanto, se puede usar un cebador de ácido nucleico más largo, por ejemplo, para aumentar la especificidad o la reproducibilidad de la replicación, si se desea. Por consiguiente, un ácido nucleico usado en un método de la invención puede ser de al menos 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500 o más nucleótidos de longitud. Los expertos en la materia reconocerán que una sonda de ácido nucleico usada en la invención también puede tener cualquiera de las longitudes expuestas anteriormente a modo de ejemplo.

Dos enfoques generales para la amplificación de genomas completos que se pueden usar en la invención incluyen el uso de alguna forma de amplificación cebada al azar o la creación de una representación genómica amplificable por PCR universal. Las técnicas a modo de ejemplo para la amplificación cebada al azar incluyen, sin limitación, las basadas en la PCR, tales como la PEP-PCR o DOP-PCR o las basadas en la amplificación de desplazamiento de cadena, tales como la amplificación de cebador aleatorio. Un método a modo de ejemplo de la creación de representaciones genómicas amplificables por PCR universal se describe, por ejemplo, en Lucito *et al.*, *Proc. Natl. Acad. Sci. EE.UU.* 95:4487-4492 (1998). Una aplicación de las representaciones genómicas es la creación de inserciones genómicas cortas (por ejemplo, 30-2000 bases) a través de la digestión por restricción de ADN_g, y la adición de colas de PCR universal mediante la ligación de adaptadores. Por lo general, la amplificación o detección de ADN_g se lleva a cabo con una población de ácidos nucleicos que se hibrida con diferentes porciones de un molde de ADN_g. Una población de ácidos nucleicos usada en la invención puede incluir miembros que tengan un

complemento aleatorio o semialeatorio de secuencias. Por lo tanto, una población de ácidos nucleicos puede tener miembros con una longitud de secuencia fija en la que una o más posiciones a lo largo de la secuencia se asignen al azar dentro de la población. A modo de ejemplo, una población de cebadores 12-meros puede tener una secuencia que sea idéntica, a excepción de una determinada posición, la posición 5, donde esté incorporado cualquiera de los cuatro nucleótidos de ADN nativos, produciendo de este modo una población que tenga cuatro miembros cebadores diferentes. En una realización particular, se pueden asignar aleatoriamente de forma combinatoria múltiples posiciones a lo largo de la secuencia. Por ejemplo, un cebador de ácido nucleico puede tener 2, 5, 10, 15, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100 o más posiciones asignadas al azar. Por ejemplo un cebador 12-mero asignado al azar en cada posición con 4 posibles nucleótidos de ADN nativos contendrá hasta $4^{12} = 1,7 \times 10^7$ miembros.

En realizaciones particulares, una población de ácidos nucleicos usada en la invención puede incluir miembros con secuencias que estén diseñadas en base a algoritmos o procesos racionales. Del mismo modo, una población de ácidos nucleicos puede incluir miembros que tengan cada uno al menos una parte de su secuencia diseñada en base a algoritmos o procesos racionales. Los algoritmos o procesos de diseño racional se pueden usar para dirigir la síntesis de un producto de ácido nucleico que tenga una secuencia diferenciada o para dirigir la síntesis de una mezcla de ácido nucleico que esté sesgada para contener preferentemente determinadas secuencias.

Usando métodos de diseño racional, se pueden seleccionar secuencias de ácidos nucleicos de una población, por ejemplo, en base a secuencias conocidas del ADNg que se vaya a amplificar o detectar. Las secuencias se pueden seleccionar de manera que la población incluya preferentemente secuencias que se hibriden a ADNg con una cobertura deseada. Por ejemplo, se puede diseñar una población de cebadores para que incluya preferentemente los miembros que se hibriden a un determinado cromosoma o parte de un ADNg tal como regiones codificantes o regiones no codificantes. También se pueden seleccionar otras propiedades de una población de ácidos nucleicos para conseguir la hibridación preferencial en las posiciones a lo largo de una secuencia de ADNg que se encuentran en una longitud media, mínima o máxima deseada entre sí. Por ejemplo, la longitud del cebador se puede seleccionar para hibridar y cebar al menos aproximadamente cada 64, 256, 1.000, 4.000, 16.000 o más bases entre sí a lo largo de una secuencia de ADNg.

Los ácidos nucleicos útiles en la invención también se pueden diseñar para omitir o reducir preferentemente secuencias que se hibriden a determinadas secuencias de un ADNg por amplificar o detectar tales como repeticiones conocidas o elementos repetitivos que incluyan, por ejemplo, repeticiones Alu. Por consiguiente, se puede diseñar una sola sonda o un solo cebador tal como el usado en la amplificación aleatoria para que incluya o excluya una determinada secuencia. Del mismo modo, se puede sintetizar una población de sondas o cebadores, tal como una población de cebadores usados para la amplificación de cebadores aleatorios, para excluir o incluir preferentemente determinadas secuencias tales como repeticiones Alu. También se puede sintetizar una población de cebadores aleatorios para que incluya preferentemente un mayor contenido de nucleótidos G y/o C en comparación con los nucleótidos A y T. La población de cebadores aleatorios resultante será rica en GC y, por lo tanto, tendrá una mayor probabilidad de hibridarse con las regiones de ricas en GC de un genoma, tales como regiones de codificación de genes de un genoma humano que normalmente tienen un contenido de GC más alto que las regiones de ADNg no codificantes. Por el contrario, se pueden sintetizar cebadores ricos en AT para amplificar o aparearse preferentemente con regiones ricas en AT, tales como las regiones no codificantes de un genoma humano. Otros parámetros que se pueden usar para influir en el diseño de ácidos nucleicos incluyen, por ejemplo, la eliminación preferencial de secuencias que vuelven a los cebadores auto-complementarios, propensos a la formación de dímeros de cebadores o propensos a la formación de horquillas o la selección preferencial de secuencias que tengan una T_m máxima, mínima o media deseada. Los ejemplos de métodos y algoritmos que se pueden usar en la invención para el diseño de sondas incluyen los descritos en el documento US 2003/0096986A1.

Los cebadores de una población de cebadores aleatorios pueden tener una región de secuencia idéntica tal como una cola universal. Una cola universal puede incluir un sitio de cebado universal para una posterior etapa de amplificación o un sitio que se aparee con un determinado agente de unión útil para el aislamiento o la detección de secuencias amplificadas. Los métodos para fabricar y usar una población de cebadores aleatorios con colas universales se describen, por ejemplo, en Singer *et al.*, *Nucl. Acid. Res.*, 25:781-786 (1997) o Grothues *et al.*, *Nucl. Acids Res.* 21:1321-2 (1993).

Los expertos en la materia reconocerán que cualquiera de una variedad de ácidos nucleicos usados en la invención, tales como sondas puede tener una o más de las propiedades, o se puede producir, como se ha expuesto anteriormente incluyendo en los ejemplos proporcionados con respecto a los cebadores.

Un método de la invención para la amplificación de un genoma puede incluir una etapa de poner en contacto un ADNg con una polimerasa en condiciones para amplificar de manera representativa el ADN genómico. El tipo de polimerasa y las condiciones usadas para la amplificación en un método de la invención se pueden seleccionar para obtener fragmentos de genoma que tengan una longitud deseada. En realizaciones particulares, se pueden obtener fragmentos relativamente pequeños en un método de la invención, por ejemplo, mediante la amplificación de ADNg con una polimerasa de baja capacidad de procesamiento o mediante la fragmentación de un molde de ADNg o sus productos de amplificación con un agente de escisión de ácido nucleico tal como una endonucleasa o agente químico. Por ejemplo, se puede usar un método de la invención para obtener una población representativa

amplificada de fragmentos de genoma que sea, sin limitación, como máximo, de aproximadamente 10 kb, 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0,8 kb, 0,6 kb, 0,5 kb, 0,4 kb, 0,2 kb o 0,1 kb de longitud.

5 En realizaciones alternativas, se puede usar un método de la invención para amplificar ADNg con el fin de formar fragmentos relativamente grandes de ADN genómico. De acuerdo con tales realizaciones, un método de la invención se puede usar para obtener una población representativa amplificada de fragmentos de genoma que sea de al menos aproximadamente 10 kb, 15 kb, 20 kb, 25 kb, 30 kb o más de longitud.

10 Se puede obtener una población representativa amplificada que incluya fragmentos de genoma que tengan un tamaño relativamente pequeño, por ejemplo, mediante la amplificación del ADNg con una polimerasa de baja capacidad de procesamiento. Una polimerasa de baja capacidad de procesamiento usada en un método de la invención puede sintetizar menos de 100 bases por polimerización. Se pueden obtener fragmentos más cortos si se desea usando una polimerasa que sintetice menos de 50, 40, 30, 20, 10 o 5 bases por polimerización en las condiciones de amplificación. Una ventaja no limitante del uso de una polimerasa de baja capacidad de procesamiento para la amplificación es que se obtienen fragmentos relativamente pequeños, lo que permite la hibridación eficaz a las matrices de ácidos nucleicos. Una polimerasa de baja capacidad de procesamiento puede ser particularmente útil para la amplificación de una muestra de genoma fragmentado. Como se expone a continuación, los métodos particularmente útiles de análisis individuales pueden incluir, por ejemplo, la captura de fragmentos en ubicaciones diferenciadas de una matriz de sondas.

20 En una realización particular, se puede amplificar un molde de ADN genómico de una sola cadena o desnaturalizado, usando una polimerasa de baja capacidad de procesamiento en un método de la invención. Un molde de ADNg se puede desnaturalizar, por ejemplo, por calor, enzimas tales como helicasa, agentes químicos, tales como sal o detergentes, pH o similar. Los ejemplos de polimerasas de baja capacidad de procesamiento y útiles para amplificar ADNg en la invención incluyen, sin limitación, la polimerasa Taq, polimerasa de T4, Pol III de *E. coli* "monomérico" (que carece de la subunidad beta) o de Pol I de ADN de *E. coli* o su fragmento deficiente en nucleasa 5' conocido como polimerasa de Klenow.

30 La invención proporciona además realizaciones en las que la amplificación se produce en condiciones en las que no se desnaturaliza el molde de ADNg. Una condición a modo de ejemplo es una temperatura a la que un ADN genómico aislado permanece sustancialmente de doble cadena. Las condiciones en las que no se requiere la desnaturalización a alta temperatura del ADN, por lo general, se denominan condiciones isotérmicas. El ADN genómico se puede amplificar en condiciones isotérmicas en la invención usando una polimerasa que tenga actividad de desplazamiento de cadena. En realizaciones particulares, se puede usar una polimerasa que tenga tanto baja capacidad de procesamiento como actividad de desplazamiento de cadena para obtener una población representativa amplificada de fragmentos de genoma. Los ejemplos de polimerasas que tienen baja capacidad de procesamiento y actividad de desplazamiento de cadena incluyen, sin limitación, Pol I de *E. coli*, polimerasa de Klenow exo^- o polimerasa T7 exo^- de grado de secuenciación.

40 En general, la actividad de la polimerasa, incluyendo, por ejemplo, la capacidad de procesamiento y la actividad de desplazamiento de cadena, puede verse afectada por factores tales como el pH, la temperatura, la fuerza iónica y la composición tampón. Los expertos en la materia sabrán qué tipos de polimerasas y condiciones se pueden usar para obtener fragmentos que tengan una longitud deseada en vista de lo que se conoce respecto a la actividad de las polimerasas como se describe, por ejemplo, en Eun, "Enzymology Primer for Recombinant DNA Technology", Academic Press, San Diego (1996) o serán capaces de determinar las polimerasas y condiciones apropiadas mediante pruebas sistemáticas usando ensayos conocidos, tales como electroforesis en gel o espectrometría de masas, para medir la longitud de los fragmentos amplificados.

50 Pol I de *E. coli* o su fragmento de Klenow se pueden usar para la amplificación isotérmica de un genoma para producir pequeños fragmentos de ADN genómico, por ejemplo, en una reacción baja en sal ($I = 0,085$) incubada a una temperatura de entre aproximadamente 5 °C y 37 °C. Los ejemplos de tampones y condiciones de pH que se pueden usar para amplificar ADNg con el fragmento de Klenow incluyen, por ejemplo, Tris-HCl 50 mM (pH 7,5), $MgCl_2$ 5 mM, NaCl 50 mM, 50 ug/ml de albúmina de suero bovino (ASB), 0,2 mM de cada dNTP, 2 ug (microgramos) de cebador aleatorio ($n = 6$), 10 ng de molde de ADNg y 5 unidades de Klenow exo^- incubadas a 37 °C durante 16 horas. Se pueden emplear condiciones de reacción similares omitiendo o sustituyendo uno o más componentes de reacción. Por ejemplo, el tampón se puede reemplazar por fosfato 50 mM (pH 7,4), o se pueden usar otros valores de pH en el intervalo de aproximadamente 7,0 a 7,8. El molde de ADNg por amplificar se puede proporcionar en cualquiera de una variedad de cantidades, incluyendo, sin limitación, las indicadas anteriormente en el presente documento. En una realización alternativa, las condiciones para la amplificación pueden incluir, por ejemplo, 10 ng de molde de ADN genómico, dNTP 2 mM, $MgCl_2$ 10 mM, 0,5 U/ul (microlitro) de la polimerasa, cebador aleatorio 50 uM (micromolar) ($n = 6$) e incubación isotérmica a 37 °C durante 16 horas.

65 En realizaciones particulares, se puede llevar a cabo una reacción de amplificación en dos etapas que incluyen, por ejemplo, una etapa de hibridación inicial seguida de una etapa de extensión. Por ejemplo, se pueden hibridar 10 ng de ADNg con cebador aleatorio 100 uM ($n = 6$) en 30 ul de Tris-Cl 10 mM (pH 7,5) mediante una breve incubación a 95 °C. La reacción se puede enfriar hasta la temperatura ambiente y llevarse a cabo una etapa de hibridación

mediante la adición de un volumen igual de Tris-Cl 20 mM (pH 7,5), MgCl₂ 20 mM, ditiotreitól 15 mM, los dNTP 4 mM y 1 U/ul Klenow exo⁻, y la incubación a 37 °C durante 16 horas. Aunque se ejemplifica para la amplificación basada en Klenow, los expertos en la materia reconocerán que se pueden usar etapas de hibridación y extensión independientes para reacciones de amplificación llevadas a cabo con otras polimerasas, tales como las expuestas más adelante.

En realizaciones particulares, se pueden sustituir los cebadores que tengan regiones de hibridación aleatorias de diferentes longitudes (n) en los métodos de amplificación basados en Klenow. Por ejemplo, los cebadores aleatorios n = 6 en las condiciones de ejemplo anteriores se pueden reemplazar por cebadores que tengan otras longitudes de secuencia aleatoria incluyendo, sin limitación, n = 7, 8, 9, 10, 11 o 12 nucleótidos. Una vez más, aunque la amplificación se ejemplifica basada en Klenow, los expertos en la materia reconocerán que se pueden usar cebadores aleatorios que tengan diferentes longitudes de secuencia aleatoria (n) para las reacciones de amplificación llevadas a cabo con otras polimerasas, tales como las expuestas más adelante.

Se puede usar la ADN polimerasa de T4 para la amplificación ADN_g de una sola cadena o desnaturalizado, por ejemplo, en HEPES 50 mM, pH 7,5, Tris-HCl 50 mM, pH 8,6, o glicinato 50 mM de pH 9,7. Una mezcla de reacción típica también puede contener KCl 50 mM, MgCl₂ 5 mM, ditiotreitól 5 mM (DTT), 40 ug/ml de ADN_g, 0,2 mM de cada dNTP, 50 ug/ml de ASB, cebador aleatorio 100 uM (n = 6) y 10 unidades de polimerasa de T4 incubada a 37 °C durante al menos una hora. Se pueden usar ciclos de temperatura para desplazar cadenas idénticas para múltiples series de amplificación.

Por lo general, la polimerasa T7 tiene una alta capacidad de procesamiento permitiendo la polimerización de miles de nucleótidos antes de disociarse de un molde de ADN. Las condiciones de reacción típicas en las que la polimerasa T7 tiene una alta capacidad de procesamiento son Tris-HCl 40 mM, pH 7,5, MgCl₂ 15 mM, NaCl 25 mM, DTT 5 mM, 0,25 mM de cada dNTP, 50 ug/ml ADN_g de una sola cadena, cebador aleatorio 100 uM (n = 6) y de 0,5 a 1 unidad de polimerasa T7. Sin embargo, a temperaturas inferiores a 37 °C, se reduce en gran medida la capacidad de procesamiento de la polimerasa T7. La capacidad de procesamiento de la polimerasa T7 también se puede reducir a altas fuerzas iónicas, por ejemplo por encima de NaCl 100 mM. La forma II de la polimerasa T7, normalmente, no es capaz de amplificar ADN de doble cadena. Sin embargo, la forma I de polimerasa T7 y la polimerasa T7 modificada (SEQUENASE® Versión 2.0, que carece de la región de 28 aminoácidos Lys118 a Arg 145) puede catalizar la replicación de desplazamiento de cadena. Por consiguiente, es posible amplificar pequeños fragmentos de genoma en un método de la invención usando una polimerasa T7 modificada o condiciones modificadas tales como las expuestas anteriormente. En realizaciones particulares, se puede usar SEQUENASE® en presencia de la proteína de unión de de una sola cadena (SSB) de *E. coli* para un mayor desplazamiento de cadena. Si se desea, la SSB también se puede usar para aumentar la capacidad de procesamiento de SEQUENASE®.

La polimerasa Taq tiene una alta capacidad de procesamiento a temperaturas de aproximadamente 70 °C cuando se hace reaccionar con un exceso molar 10 veces superior de molde y de cebador aleatorio (n = 6). Una reacción de amplificación realizada en estas condiciones puede incluir además un tampón tal como Tris-HCl a aproximadamente 20 mM, pH de aproximadamente 7, MgCl₂ aproximadamente 1 a 2 mM, y 0,2 mM de cada dNTP. Además, se puede añadir un agente estabilizador tal como glicerol, gelatina, ASB o un detergente no iónico. La polimerasa Taq tiene una baja capacidad de procesamiento a temperaturas inferiores a 70 °C. Por consiguiente, se pueden obtener pequeños fragmentos de ADN_g usando la polimerasa Taq a una temperatura baja en un método de la invención, o en otra condición en la que la Taq tenga una baja capacidad de procesamiento. En otra realización, el fragmento Stoffel, que carece de los restos de 289 aminoácidos N-terminales de la polimerasa Taq y tiene una baja capacidad de procesamiento a 70 °C, se puede usar para generar fragmentos de ADN_g relativamente pequeños en un método de la invención. Taq se puede usar para amplificar moldes de ADN de una sola cadena o desnaturalizado en un método de la invención. Se pueden usar ciclos de temperatura para desplazar cadenas idénticas para múltiples series de amplificación.

Los expertos en la materia reconocerán que las condiciones para la amplificación con las diversas polimerasas como se han establecido anteriormente son a modo de ejemplo. Por lo tanto, se pueden hacer cambios menores que no alteren sustancialmente la actividad. Además, se pueden cambiar las condiciones sustancialmente para lograr una actividad de amplificación deseada o para adaptarse a una determinada aplicación de la invención.

La invención también se puede llevar a cabo con variantes de las polimerasas descritas anteriormente, siempre que conserven la actividad de la polimerasa. Los ejemplos de variantes incluyen, sin limitación, las que tienen una menor actividad exonucleasa, una mayor fidelidad, una mayor estabilidad o una mayor afinidad por análogos de nucleósidos. Los ejemplos de variantes, así como otras polimerasas que son útiles en un método de la invención incluyen, sin limitación, ADN polimerasa del bacteriófago phi29 (patentes de EE.UU. N° 5.198.543 y 5.001.050), ADN polimerasa Bca exo (-) (Walker y Linn, *Clinical Chemistry* 42:1604-1608 (1996)), ADN polimerasa de fago M2 (Matsumoto *et al.*, *Gene* 84:247 (1989)), ADN polimerasa de fago phiPRD 1 (Jung *et al.*, *Proc. Natl. Acad. Science.* EE.UU. 84:8287 (1987)), ADN polimerasa VENTT® exo (-) (Kong *et al.*, *J. Biol. Chem.* 268:1965-1975 (1993)), ADN polimerasa T5 (Chatterjee *et al.*, *Gene* 97:13-19 (1991)) y ADN polimerasa PRD1 (Zhu *et al.*, *Biochim. Biophys. Acta.* 1219:267-276 (1994)).

Otra variante más de la polimerasa que es útil en un método de la invención es una polimerasa modificada que, en comparación con su versión no modificada de tipo silvestre, tiene una capacidad reducida o eliminada para añadir nucleótidos de no molde dirigidos al extremo 3' de un ácido nucleico. Los ejemplos de variantes incluyen aquellas que modifican la actividad de la polimerasa hacia la adición de todos los tipos de nucleótidos, o uno o más tipos de nucleótidos tales como nucleótidos de pirimidina, nucleótidos de purina, A, C, T, U o G. Las modificaciones pueden incluir la modificación química de grupos de aminoácidos en la polimerasa o mutaciones de secuencias.

Un agente de formación de muescas usado en un método de la invención puede ser cualquier entidad física, química o bioquímica que escinda un enlace covalente que conecte las secuencias adyacentes en una primera cadena de ácido nucleico para producir un producto en el que las secuencias adyacentes se hibriden a la misma cadena complementaria. Los ejemplos de agentes de formación de muescas incluyen, sin limitación, enzimas de formación de muescas de una sola cadena tales como ADNasa I, N.BstNBI, MutH o proteína genell de bacteriófago f1; reactivos químicos tales como radicales libres; o ultrasonido.

Un agente de formación de muescas se puede poner con contacto con un ADNg de doble cadena mezclando el agente y el ADNg en una solución. Los expertos en la materia conocerán o serán capaces de determinar las condiciones apropiadas para mellar ADNg basándose en lo que se conoce en la técnica con respecto a la actividad del agente de formación de muescas disponible como, por ejemplo, de varios proveedores comerciales tales como Promega Corp. (Madison, Wisconsin) o Roche Applied Sciences (Indianapolis, IN). Un agente químico o biológico de formación de muescas puede ser uno que sea exógeno al ADN genómico, procediendo de una fuente que sea diferente de la del ADN. Como alternativa, en un método de la invención, se puede poner en contacto un agente de formación de muescas que se encuentre normalmente con el ADN genómico en su entorno nativo con el ADNg. Tal agente de formación de muescas endógeno se puede activar para aumentar su actividad de formación de muescas o se puede asilar del ADN genómico y, posteriormente, mezclarlo con el ADNg, por ejemplo, a una concentración más alta en comparación con su entorno nativo con el ADNg. El agente de formación de muescas, bien endógeno o exógeno a un ADNg, se puede aislar antes de ponerlo en contacto con el ADNg en un método de la invención.

Los expertos en la materia entenderán que se puede proporcionar una población representativa amplificada de fragmentos de genoma de una muestra recién aislada o de una que se haya almacenado en condiciones apropiadas para la conservación de la integridad de la muestra. Por lo tanto, una muestra proporcionada en un método de la invención puede incluir agentes que estabilicen los fragmentos, siempre y cuando los agentes no interfieran en las etapas de hibridación y detección ni en otras etapas usadas en las diversas realizaciones expuestas en el presente documento. En los casos en los que se incluye en una muestra un agente estabilizante que interfiere con los métodos, los fragmentos se pueden separar del agente mediante métodos de separación y purificación conocidos. Los expertos en la materia conocerán o serán capaces de determinar fácilmente las condiciones apropiadas para el almacenamiento de una población representativa de fragmentos de genoma basándose en condiciones conocidas en la técnica para el almacenamiento de ácidos nucleicos como se describe, por ejemplo, en Sambrook *et al.*, *Supra*, (2001) y en Ausubel *et al.*, *supra*, (1998). En realizaciones particulares, se puede amplificar un ADNg mediante un método que utilice una reacción en cadena de la polimerasa (PCR) cebada con oligonucleótidos aleatorios o degenerados con moldes de ADNg desnaturizados por calor. Se conoce un método de ejemplo como etapa previa de amplificación por extensión de cebadores (PEP). Esta técnica usa 15-meros aleatorios en combinación con ADN polimerasa Taq para iniciar copias en todo el genoma. Esta técnica se puede usar para amplificar el ADN genómico a partir de tan poco como una sola célula usando, por ejemplo, las condiciones descritas en Zhang *et al.*, *Proc. Natl. Acad. Science. EE.UU.*, 89:5847-51 (1992); Snabes *et al.*, *Proc. Natl. Acad. Science. EE.UU.*, 91:6181-85 (1994); o Barrett *et al.*, *Nucleic Acids Res.*, 23:3488-92 (1995).

Otro método de amplificación de ADNg que es útil en la invención es la PCR marcada, que usa una población de cebadores de dos dominios que tienen una región 5' constante seguida de un región 3' aleatoria como se describe, por ejemplo, en Grothues *et al.* *Nucleic Acids Res.* 21 (5):1321-2 (1993). Las primeras series de amplificación se llevan a cabo para permitir una multitud de iniciaciones en ADN desnaturizado por calor basadas en la hibridación individual de la región 3' sintetizada aleatoriamente. Debido a la naturaleza de la región 3', los sitios de iniciación serán aleatorios en todo el genoma. Tras ello, se pueden retirar los cebadores no unidos y se puede realizar una posterior replicación usando cebadores complementarios a la región 5' constante.

Un enfoque adicional que se puede usar para amplificar ADNg en un método de la invención es la reacción en cadena de la polimerasa cebada con oligonucleótidos degenerados (DOP-PCR) en las condiciones descritas, por ejemplo, por Cheung *et al.*, *Proc. Natl. Acad. Science. EE.UU.*, 93:14676-79 (1996) o la patente de EE.UU. Nº 5.043.272. Se pueden amplificar bajas cantidades de ADNg, por ejemplo, 15 pg de ADNg humano, a niveles que se detectan convenientemente en los métodos de la invención. Se pueden seleccionar las condiciones de reacción usadas en los métodos de Cheung *et al.* para la producción de una población representativa amplificada de fragmentos de genoma que tenga una cobertura casi completa del genoma humano. Además, se pueden usar versiones modificadas de DOP-PCR, tales como las descritas por Kittler *et al.* en un protocolo conocido como LL-DOP-PCR (productos largos procedentes de DOP-PCR de bajas cantidades de ADN) para amplificar ADNg de acuerdo con la invención (Kittler *et al.*, *Anal. Biochem.* 300:237-44 (2002)).

La reacción en cadena de la polimerasa de preamplificación por extensión de cebadores (PEP-PCR) también se puede usar en un método de la invención con el fin de amplificar ADNg. Las condiciones útiles para la amplificación de ADNg usando PEP-PCR incluyen, por ejemplo, las descritas en Casas *et al.*, *Biotechniques* 20:219-25 (1996).

5 La amplificación de ADNg en un método de la invención también se puede llevar a cabo en un molde de ADNg que no haya sido desnaturalizado. Por consiguiente, la invención puede incluir una etapa de producción de una población representativa amplificada de fragmentos de genoma a partir de un molde de ADNg en condiciones isotérmicas. Los métodos de amplificación isotérmica a modo de ejemplo que se pueden usar en un método de la invención incluyen, pero sin limitación, amplificación por desplazamiento múltiple (MDA) en condiciones tales como
 10 las descritas en Dean *et al.*, *Proc Natl. Acad. Sci. EE.UU.* 99:5261-66 (2002) o la amplificación isotérmica de ácido nucleico por desplazamiento de cadena como se describe en la patente de EE.UU. N° 6.214.587. Otros métodos no basados en la PCR que se pueden usar en la invención incluyen, por ejemplo, la amplificación por desplazamiento de cadena (SDA) que se describe en Walker *et al.*, "Molecular Methods for Virus Detection", Academic Press, Inc., 1995; Patentes de EE.UU. N° 5.455.166 y 5.130.238, y Walker *et al.*, *Nucl. Acids Res.* 20:1691-96 (1992) o la
 15 amplificación por desplazamiento de cadena hiperramificadas que se describe en Lage *et al.*, "Genome Research" 13:294-307 (2003). Los métodos de amplificación isotérmica se pueden usar con la polimerasa ϕ 29 de desplazamiento de cadena o fragmento grande de ADN polimerasa Bst, 5'→3' exo⁻ para la amplificación de cebador aleatorio de ADN genómico. El uso de estas polimerasas se aprovecha de su alta capacidad de procesamiento y actividad de desplazamiento de cadena. La alta capacidad de procesamiento permite a las polimerasas producir
 20 fragmentos que son de 10-20 kb de longitud. Como se ha expuesto anteriormente, se pueden producir fragmentos más pequeños en condiciones isotérmicas usando polimerasas que tengan una baja capacidad de procesamiento y actividad de desplazamiento de cadena tales como la polimerasa de Klenow.

En realizaciones particulares de la invención, un ADN genómico o población de fragmentos de ADNg amplificados
 25 se puede transcribir *in vitro* a fragmentos de ARN genómico (ARNg). La creación de ARNg en un método de la invención ofrece varias ventajas no limitantes para la detección de locus tipificables en ensayos de extensión de cebadores, tales como los ensayos de extensión de cebadores basados en matrices de ADN. La extensión de cebadores basada en matrices incluye, por lo general, una etapa tal como eliminaciones, adiciones o reemplazos de aminoácidos. Los ejemplos de polimerasas modificadas que tienen una capacidad reducida o eliminada para añadir
 30 nucleótidos dirigidos de no molde al extremo 3' de un ácido nucleico se describen, por ejemplo, en el documento US 6.306.588 o Yang *et al.*, *Nucl. Acids Res.* 30: 4314-4320 (2002). En una realización particular, tal variante de polimerasa se puede usar en un método de detección SBE o ASPE descrito en el presente documento.

Un ADN genómico de doble cadena que vaya a ser amplificado por una polimerasa de desplazamiento de cadena se
 35 puede hacer reaccionar con un agente de formación de muescas para producir roturas de una sola cadena en la estructura covalente del molde de ADN genómico. La introducción de roturas de una sola cadena en un molde de ADNg se puede usar, por ejemplo, para mejorar la eficacia de la amplificación o la capacidad de reproducción de la amplificación isotérmica. La formación de muescas se puede usar, por ejemplo, en una reacción de amplificación con cebador aleatorio o una reacción de amplificación cebada aleatoriamente. Una ventaja no limitante de introducir
 40 roturas de una sola cadena en una reacción de amplificación es que se puede usar en lugar de la desnaturalización por calor. La desnaturalización por calor es perjudicial para ciertas reacciones de amplificación cebadas aleatoriamente según lo descrito, por ejemplo, en Lage *et al.*, *Genome Res.* 13:294-307 (2003). A este respecto, las ubicaciones en las que hay un molde de ADNg mellado pueden proporcionar sitios de cebado para la actividad polimerasa. Por lo tanto, el contacto de un ADNg con un agente de formación de muescas puede aumentar el
 45 número de sitios de cebado del molde de ADNg, mejorando así la eficacia de la amplificación. El número de muescas o la ubicación de las muescas o ambos parámetros pueden verse influenciados por el uso de determinadas condiciones que favorezcan un nivel de actividad de formación de muescas deseado o un uso de un agente de formación de muescas que sea específico de la secuencia. Así pues, el uso de un agente de formación de muescas puede mejorar la capacidad de reproducción de la amplificación.

50 Por consiguiente, la divulgación proporciona además un método de amplificación de ADN genómico que incluye las etapas de: (a) proporcionar ADN genómico de doble cadena aislado; (b) poner en contacto el ADN genómico de doble cadena con un agente de formación de muescas, produciendo así ADN genómico de doble cadena mellado; y (c) poner en contacto el ADN genómico de doble cadena mellado con una polimerasa de desplazamiento de cadena
 55 y una pluralidad de cebadores, en la que se amplifica el ADN genómico. Como se ha expuesto anteriormente, la pluralidad de cebadores puede ser una población de cebadores aleatorios, por ejemplo, una reacción de amplificación con cebadores aleatorios de hibridación de un ADN diana a una sonda de ADN inmovilizada y la posterior modificación o extensión del híbrido de sonda-diana con una ADN polimerasa. Estos ensayos a menudo se pueden ver comprometidos por los artefactos derivados de la formación no deseada de híbridos de sonda-sonda, debido a su proximidad física en la superficie de la matriz, y la posterior extensión ectópica de estos híbridos de sonda-sonda. En realizaciones de la invención donde el ADNg se convierte en ARNg, se pueden evitar tales artefactos, porque la ADN polimerasa se sustituye con transcriptasa inversa (RT), que no modifica de manera eficaz ni extiende los híbridos de sonda-sonda, ya que son híbridos de ADN-ADN y la transcriptasa inversa es selectiva de los híbridos que tienen un molde de ARN. Además, el uso de ARNg y la transcriptasa inversa para la detección de
 60 híbridos de diana-sonda minimiza la extensión ectópica en un ensayo de hibridación directa/extensión de cebadores basado en matriz. En una reacción de extensión de cebadores basada en matriz, la autoextensión tanto inter- como

intra-sonda (extensión ectópica) puede conducir a altos fondos. El uso de RT y ARNg evita artefactos debido a la extensión ectópica porque, aunque la RT puede extender fácilmente una sonda de ADN hibridada a una diana de ARN, no extenderá de manera eficaz los complejos de ADN-ADN.

5 Por consiguiente, la divulgación proporciona un método para la detección de locus tipificables de un genoma. El método incluye las etapas de (a) transcribir *in vitro* una población de fragmentos de ADN_g amplificados, obteniéndose de este modo fragmentos de ARN genómico (ARN_g); (b) hibridar los fragmentos de ARNg con una pluralidad de sondas de ácido nucleico que tengan secuencias correspondientes a los locus tipificables; y (c) detectar locus tipificables de los fragmentos de ARNg que se hibridan con las sondas.

10 En la Figura 8, se muestra un ejemplo esquemático de un método para amplificar ADN_g con el fin de producir fragmentos de ARNg. Como se muestra en el panel 8A, se puede amplificar ADN_g con ADN polimerasa y una población de cebadores de ADN aleatorios para producir una población representativa de fragmentos de genoma antes de una etapa de transcripción *in vitro*. En el ejemplo mostrado, el ADN_g es marcado mediante cebado aleatorio (RPL) usando una población de cebadores que incluye una región aleatoria de 9 nucleótidos y una región fija que tiene una secuencia de cebado universal de (U1) y una secuencia de promotor T7 (T7). En el ejemplo mostrado en la Figura 8, la secuencia aleatoria es de 9 nucleótidos de longitud. Sin embargo, se entenderá que se puede usar cualquiera de una variedad de longitudes de secuencia aleatorias para adaptarse a una aplicación particular de la invención incluyendo, por ejemplo, una secuencia aleatoria que sea de 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15 o más nucleótidos de longitud. Además, una secuencia aleatoria de un cebador usado en un método de la invención puede incluir posiciones intercaladas que tengan un nucleótido fijo o regiones que tengan una secuencia fija de dos o más nucleótidos, si se desea.

25 Como se muestra en el panel B, la población representativa de fragmentos de genoma marcados con promotor T7 se puede transcribir *in vitro* en forma de ARNg usando una ARN polimerasa T7 y un cebador T7 complementario (cT7). La transcripción de ADN_g en fragmentos de ARNg también se puede llevar a cabo con otros promotores tales como T3 o SP6 y sus respectivas polimerasas como se expone más detalladamente a continuación.

30 Una población representativa basada en ARNg de fragmentos de genoma producidos por la transcripción *in vitro* se puede manipular y detectar en cualquiera de una variedad de maneras como se establece en el presente documento. Por ejemplo, los fragmentos de genoma basados en ARNg producidos mediante los métodos ejemplificados en la Figura 8B tendrán colas marcadas con U1. Estas colas se pueden usar, por ejemplo, para aislar los fragmentos de ARNg de ADN_g y otros componentes de la reacción de amplificación usando una secuencia de captura complementaria unida a una fase sólida. Los fragmentos de ARN genómicas se pueden detectar o copiar en ADN usando una transcriptasa inversa. La población representativa basada en ARNg de fragmentos de genoma se puede detectar directamente usando métodos tales como los indicados a continuación o, de forma alternativa, se pueden copiar en el ADN antes de la detección. Como se muestra en la etapa de amplificación a modo de ejemplo de la Figura 8C, la población de fragmentos de ARNg se puede replicar usando cebadores específicos del locus, que tienen opcionalmente una segunda secuencia universal (U2) y una transcriptasa inversa. A esta etapa le puede seguir la amplificación usando PCR universal con cebadores U1 y U2. Por lo tanto, los fragmentos de ARNg se pueden replicar para producir una población representativa amplificada de fragmentos de genoma específica del locus. Como se expone a continuación con más detalle, la replicación dirigida por transcriptasa inversa del ARNg con cebadores específicos del locus puede proporcionar una reducción de la complejidad y, si se desea, puede añadir un sitio de cebado universal U2. En realizaciones en las que la secuencia de U2 está presente, cada población de fragmentos de genoma producida por la replicación con cebadores específicos del locus tendrá secuencias flanqueantes U1 y U2 que son útiles para detectar o amplificar la población. Por lo tanto, los productos completamente extendidos se pueden amplificar en una reacción de PCR universal cebada en los sitios de los cebadores U1 y U2.

50 Por otra parte, como se muestra en la Figura 8D, un "dímero cebador" no puede extenderse en la etapa de detección, porque la transcriptasa inversa no puede extender una molde de ADN de manera muy eficaz. Por el contrario, una ADN polimerasa puede extender el dímero cebador L1-L2, pudiendo conducir a artefactos de detección. Por lo tanto, el uso de poblaciones representativas basadas en ARNg de fragmentos de genoma puede proporcionar la ventaja no limitante de evitar artefactos en algunos métodos de detección de multiplexado. Así pues, el uso de ARNg puede proporcionar la ventaja de una mayor eficacia para la detección de multiplexado de un gran número de locus tipificables.

60 Un cebador de ácido nucleico usado en un método de la invención para transcribir ADN_g en una población representativa basada en ARNg de fragmentos de genoma o para transcribir inversamente ARNg puede tener una longitud, composición u otras propiedades como se expone en el presente documento en lo que se refiere a los cebadores usados con otras polimerasas y moldes. Los expertos en la materia conocerán o serán capaces de determinar las propiedades pertinentes de un cebador de ácido nucleico para su uso en una etapa de transcripción *in vitro* o de transcriptasa inversa de la invención basándose en la orientación y las enseñanzas establecidas en el presente documento y en lo que se sabe acerca de las transcriptasas inversas o ARN polimerasas tal como se expone a continuación y se describe, por ejemplo, en Eun *et al.*, *supra* (1996).

65

Además, aunque las poblaciones de cebadores ejemplificadas anteriormente en relación con la realización de la Figura 8 tienen una sola secuencia de U1 y una sola secuencia de U2, se entenderá que una población de cebadores útiles en la invención puede incluir más de una región de secuencia constante. Por lo tanto, una pluralidad de subpoblaciones de cebadores aleatorios, cada una con diferentes regiones de secuencias constantes, puede estar presente en una población más grande usada para la hibridación o amplificación en un método de la invención. En un método de la invención, se puede usar cualquier ARN polimerasa que sea capaz de sintetizar un ARN complementario a partir de un molde de ADN. Una ARN polimerasa de ejemplo útil en la invención es la ARN polimerasa T7. Las condiciones que se pueden usar en un método de la invención para la transcripción *in vitro* con ARN polimerasa T7 incluyen, sin limitación, Tris-HCl 40 mM, pH 8,0 (37 °C), MgCl₂ 6 mM, DTT 5 mM, espermidina 1 mM, 50 ug/ml de ASB, 40 ug/ml de fragmentos de ADNg incluyendo un promotor fago, PNT 0,5-8,5 mM y de 200 a 300 unidades de ARN polimerasa T7 en 50 microlitros. Otra ARN polimerasa que se puede usar en un método de la invención es la ARN polimerasa SP6. Las condiciones a modo de ejemplo para su uso incluyen, sin limitación, Tris-HCl 40 mM, pH 8,0 (25 °C), MgCl₂ 6 mM, DTT 10 mM, espermidina 2 mM, 50 ug/ml de fragmentos de ASB, 50 ug/ml de ADNg que contienen un promotor de SP6, 0,5 mM de cada NTP y 10 unidades de ARN polimerasa SP6 en 50 microlitros.

También se puede usar ARN polimerasa T3 en un método de la invención para la transcripción *in vitro*, por ejemplo, en condiciones que incluyen Tris-HCl 50 mM, pH 7,8 (37 °C), NaCl 25 mM, MgCl₂ 8 mM, DTT 5 mM, espermidina 2 mM, 50 ug/ml de ASB, 50 ug/ml de fragmentos de ADNg que contienen un promotor de T3, 0,5 mM de cada NTP, y ARN polimerasa T3 en 50 microlitros.

En un método de la invención, se puede usar cualquier transcriptasa inversa (RT) que catalice la síntesis de ADN complementario a partir de un molde de ARN. Las RT a modo de ejemplo que se pueden usar en un método de la invención incluyen, pero sin limitación, las de retrovirus tales como RT del virus de la mioblastosis aviar (AMV), RT del virus de la leucemia murina de Moloney (MoLV), RT del VIH-1 o RT del virus del sarcoma de Rouse (RSV). En general, una reacción de transcripción inversa usada en un método de la invención incluirá un molde de ARN, uno o más dNTP y un cebador de ácido nucleico con un grupo OH en 3'. Si se desea, se pueden añadir inhibidores de ARNs para inhibir la degradación del producto transcrito. Se pueden usar determinadas condiciones de reacción para adaptarse a una RT en particular o una aplicación particular de la invención.

Las condiciones útiles para la modificación o la elongación con RT del AMV incluyen, por ejemplo, Tris-HCl 50 mM (pH 8,3 a 42 °C), NaCl 150 mM (o KCl 100 mM), MgCl₂ 6 a 10 mM, DTT 1 mM, 50 ug/ml de ASB, 50 unidades de ARNs, HCl de espermidina 0,5 mM, Na-PP_i 4 mM, 0,2 mM de cada dNTP, 1-5 ug de ARNg, 0,5 a 2,5 ug de cebador y 10 unidades RT del AMV en 50 microlitros. Sin embargo, también es posible llevar a cabo la reacción a un pH de 8,1 a 25 °C con el resto de condiciones similares. Se pueden usar otras condiciones para la actividad de la RT del AMV y, en particular, para inhibir la síntesis de ADN dependiente del ADN descrita, por ejemplo, en Likhava *et al.*, *FEBS Lett.* 274: 156-158 (1990) o Likhava *et al.*, *Mol. Biol.* (URSS) 24:396-407 (1990).

En realizaciones en las que se usa RT de MoLV, las condiciones a modo de ejemplo para la modificación o la elongación incluyen, sin limitación, Tris-HCl 50 mM (pH 8,1 a 25 °C), KCl 75 mM, MgCl₂ 3 mM, DTT 10 mM, 100 ug/ml de ASB, 20 unidades de ARNs, 50 ug/ml de actinomicina D, 0,5 mM de cada dNTP, 5-10 ug de ARNg, 0,5 a 4 ug de cebador y 200 unidades de RT del MoLV en 50 microlitros.

Las RT usadas en un método de la invención también pueden tener un origen no-retroviral que incluye, por ejemplo, virus de ADN tales como el virus de la hepatitis B o caulimovirus, bacterias tales como *Myxococcus xanthus* o algunas cepas de *E. coli*, levaduras tales como las que portan el retrotransposón Ty, hongos, invertebrados tales como los que portan el elemento de tipo copia como de *Drosophila*, o plantas. Por otra parte, la transcripción inversa deseada se puede llevar a cabo en un método de la invención usando una ADN polimerasa que tenga actividad de RT, tal como Pol I de ADN de *E. coli*. Sin embargo, por las razones expuestas anteriormente, se puede desear llevar a cabo la transcripción inversa en condiciones en las que la actividad hacia las moldes de ADN se inhiba o esté sustancialmente ausente, por ejemplo, usando una RT que no sea capaz de realizar la síntesis de ADN dependiente del ADN o usando condiciones tales como pH, fuerza iónica o concentración de Mg²⁺ que inhiba la síntesis de ADN dependiente del ADN. Además, si se desea, se puede añadir un inhibidor de la síntesis de ADN dependiente del ADN tal como actinomicina D o pirofosfato (Na-PP_i).

Una ADN polimerasa de ejemplo que es capaz de tener actividad de RT es *Tth* pol cuando se usa en presencia de Mn²⁺. Las condiciones a modo de ejemplo para la transcripción inversa de ARNg con RT *Tth* pol incluyen, sin limitación, Tris-Cl 50 mM (pH 8,8), NH₄SO₄ 16 mM, MnCl₂ 1 mM, dNTP 200 μM, 0,25 U/μl de *Tth* pol, 100 fmol/μl de molde de ARN a 70 °C durante 20 min.

La amplificación de ADNg en un método de la invención se puede llevar a cabo de manera que se produzca una población representativa amplificada de fragmentos de genoma que tenga una complejidad deseada. Por ejemplo, una población representativa amplificada de fragmentos de genoma que tenga una complejidad deseada se puede producir mediante la especificación de la frecuencia o la diversidad de los eventos de cebado o de fragmentación que ocurren durante una reacción de amplificación. Por consiguiente, la invención se puede usar para producir una población representativa amplificada de fragmentos de genoma que tenga complejidad alta o baja dependiendo del

uso deseado de la población de fragmentos. Varias de las condiciones de amplificación establecidas anteriormente y en los siguientes ejemplos proporcionan representaciones de alta complejidad. Un método de la invención puede incluir una etapa de reducción de la complejidad o se puede llevar a cabo con un método de amplificación que produzca una representación de baja complejidad, si se desea.

5 Un método a modo de ejemplo para producir una representación de baja complejidad es la PCR con adaptadores y enlazadores que requiere una digestión aleatoria inicial de ADN con una endonucleasa de restricción, la unión de los fragmentos digeridos a un oligonucleótido adaptador y la amplificación por PCR de fragmentos derivatizados de adaptadores desnaturalizados por calor como se describe, por ejemplo, en Lucito *et al.*, *Genome Res.*10:1726-36
10 (2000). La modificación de las condiciones de la digestión de ADN_g en el método se puede usar para influir en la complejidad de la población amplificada representativa de fragmentos de genoma producida. En particular, se puede conseguir una representación de baja complejidad usando una endonucleasa de corte infrecuente que tenga, por ejemplo, motivo de reconocimiento más largo o de 6 bases. Por consiguiente, se puede usar un cortador frecuente para obtener una representación de alta complejidad. Por ejemplo, Dpn II, que reconoce el sitio de cuatro nucleótidos GATC y, por lo tanto, restringe el ADN_g con relativa frecuencia, puede producir una población representativa de fragmentos de genoma humano que contenga aproximadamente un 70 % del genoma. Por el contrario, se puede usar un cortador relativamente poco frecuente para producir una representación de baja complejidad. Por ejemplo, Bgl II, que reconoce el sitio de seis nucleótidos AGATCT y, por lo tanto, restringe el ADN_g relativamente con poca frecuencia, se puede usar para producir una población representativa de fragmentos de genoma humano que contenga solo aproximadamente el 2,5 % de un genoma. Además, se puede fragmentar un ADN_g en una longitud media que sea inferior a la capacidad de procesamiento de la polimerasa usada para la amplificación, reduciendo así la complejidad de la población representativa amplificada de fragmentos de genoma producida.

25 Un método adicional para producir una representación de baja complejidad es el uso de dos o más adaptadores para la PCR de adaptadores y enlazadores anclados. En realizaciones particulares, se puede conseguir reducir la complejidad mediante la fragmentación de una muestra de ADN_g usando al menos dos enzimas de restricción; la unión de los adaptadores a los fragmentos resultantes; y la amplificación selectiva de los fragmentos que fueron cortados en un extremo por una enzima de restricción y en el otro extremo por una enzima de restricción diferente.
30 Si una enzima es un cortador de 6 y la otra es un cortador de 4, la representación se anclará sobre los sitios cortadores de 6 con un tamaño medio determinado por la frecuencia de la digestión del cortador de 4 (aproximadamente cada 256 bases). Este es un tamaño útil para la amplificación basada en PCR. La complejidad de la muestra resultante se puede regular mediante la elección de enzimas que corten con una determinada frecuencia. La amplificación selectiva también se puede realizar mediante el diseño de un adaptador que tenga un saliente en 5' y el segundo adaptador que tenga un saliente en 3', donde los salientes tienen los sitios de hibridación para cebadores de la amplificación usados para replicar los fragmentos. En el documento US 2003/0096235 A1, se describen condiciones a modo de ejemplo para el uso de múltiples adaptadores para reducir la complejidad.

40 La reducción de la complejidad también se puede llevar a cabo de una manera específica del locus. Por consiguiente, la invención proporciona además un método de producción de una población de fragmentos de genoma representativa amplificada, específica del locus, de complejidad reducida. El método incluye las etapas de (a) replicar un genoma nativo con una pluralidad de cebadores aleatorios, produciendo de este modo una población representativa amplificada de fragmentos de genoma; (b) replicar una subpoblación de la población representativa amplificada de fragmentos de genoma con una pluralidad de diferentes cebadores específicos del locus, produciendo de este modo una población representativa amplificada de fragmentos de genoma específica del locus; y (c) aislar la subpoblación, produciendo así una población de fragmentos de genoma representativa amplificada, específica del locus, de complejidad reducida.

50 Un método a modo de ejemplo que se puede usar para la reducción de la complejidad es la amplificación para producir fragmentos de ARN_g como se muestra en la Figura 8 y se ha descrito anteriormente. En la Figura 9, se muestra un ejemplo esquemático de un método para producir una población de fragmentos de genoma representativa amplificada, específica del locus, de complejidad reducida. Como se muestra en la Figura 9a, se puede amplificar una muestra de ADN_g una técnica de marcaje cebado aleatoriamente (RPL) empleando una población de cebadores de ácido nucleico, cada uno con una secuencia 3' aleatoria para la hibridación al ADN_g y una cola de cebado universal 5' (secuencia U1). Por lo tanto, una reacción de marcaje cebado aleatoriamente puede producir una población representativa amplificada de fragmentos de genoma, flanqueada por un sitio de cebado universal. En el ejemplo mostrado en la Figura 9, la secuencia aleatoria tiene 9 nucleótidos. Sin embargo, se entenderá que se puede usar cualquiera de una variedad de longitudes de secuencia aleatorias o composiciones para adaptarse a una aplicación particular de la invención incluyendo, por ejemplo, las establecidas anteriormente en el presente documento. En general, a medida que se reduce la longitud de la porción de hibridación aleatoria de una población de cebadores aleatorios, se aumentará el número de posibles sitios de hibridación en un genoma, aumentándose así la complejidad de la representación amplificada.

65 Como se muestra en la Figura 9B, se puede aislar una población representativa amplificada de fragmentos de genoma de ADN genómico, por ejemplo, mediante inmovilización sobre perlas en fase sólida. En el ejemplo de la Figura 9A, se puede facilitar la inmovilización de los fragmentos amplificados con una biotina unida al cebador N₉-

U1. El producto de amplificación biotinilado puede ser capturado por una fase sólida derivatizada con avidina o estreptavidina y, si se desea, posteriormente aislado del molde de ADNg. Anteriormente, se han expuesto otros restos de captura a modo de ejemplo y sus receptores inmovilizados que se pueden usar en un cebador para la amplificación de cebador aleatorio. Por lo tanto, un método de amplificación de ADNg puede incluir además una
 5 etapa de captura o aislamiento de una población representativa amplificada de fragmentos de genoma. Los ejemplos de sustratos que se pueden usar para capturar o aislar una población representativa amplificada de fragmentos de genoma incluyen, por ejemplo, los indicados a continuación con respecto a la separación de ácidos nucleicos de una sola cadena de híbridos de ácidos nucleicos.

10 Los expertos en la materia reconocerán que los fragmentos amplificados de genoma se pueden separar de otros componentes de la reacción en un método de la invención usando un sustrato en fase sólida según se ha ejemplificado anteriormente. De igual manera, los fragmentos de genoma amplificados se pueden separar en base a otras propiedades de los fragmentos tales como su tamaño. Por lo tanto, se pueden usar métodos de filtración o de cromatografía tales como la cromatografía de exclusión por tamaño para separar fragmentos de genoma de otros
 15 componentes de reacción tales como las sondas que no estén hibridadas.

Un método de la invención puede incluir una etapa de replicar una subpoblación de la población representativa amplificada de fragmentos de genoma con una pluralidad de diferentes cebadores específicos del locus, cada uno con una región de la secuencia específica del locus 3' y una región de secuencia constante 5'. Continuando con el
 20 ejemplo de la Figura 9B, el producto amplificado con cebadores aleatorios inmovilizados se puede hibridar con una población de diferentes cebadores que tengan diferentes secuencias 3' específicas del locus identificadas como L1, L2 o L3, y una segunda cola universal 5' (U2). En este punto, se puede incluir una etapa de lavado, si se desea, para eliminar los cebadores no hibridados y en exceso. Las condiciones para el lavado pueden incluir cualquiera que elimine los ácidos nucleicos unidos inespecíficamente, manteniendo los híbridos específicos. A continuación, se
 25 puede usar la extensión de los cebadores para replicar una subpoblación de la población representativa amplificada de fragmentos de genoma que tenga secuencias complementarias a los cebadores específicos del locus. Esta subpoblación tendrá menor complejidad que el ADNg original y la población amplificada de fragmentos de genoma que se produjo con el cebador N₉-U1. Además, la reducción de la complejidad será específica del locus debido a la selección con los cebadores específicos del locus en la segunda etapa de amplificación. Es posible modificar el
 30 número de diferentes cebadores específicos del locus y la longitud de las secuencias específicas del locus para aumentar o disminuir la complejidad de una representación obtenida en un método de la invención.

La extensión de los cebadores que contienen U2 por toda la longitud de los fragmentos capturados en el ejemplo
 35 mostrado en la Figura 9B producirá una población representativa amplificada de fragmentos de genoma específica del locus marcada con la primera región constante (U1) y la segunda región constante (U2). Por lo tanto, los productos completamente extendidos se pueden amplificar en una reacción de PCR universal cebada en los sitios de los cebadores U1 y U2. Por consiguiente, un método de la invención puede incluir una etapa de replicación de una población representativa amplificada de fragmentos de genoma específica del locus de complejidad reducida, con cebadores complementarios a la primera y segunda región constante flanqueante. Además, se puede hacer la
 40 detección de los fragmentos basándose en la presencia ambas secuencias U1 y U2, por ejemplo, usando técnicas descritas a continuación en lo que se refiere a la detección de sondas OLA modificadas.

La reducción de la complejidad también se puede llevar a cabo mediante la eliminación de determinadas secuencias
 45 de una población de fragmentos de genoma. En una realización, es posible inhibir la hibridación de secuencias abundantes o de alto número de copias de una muestra de fragmentos de genoma con sondas de detección o de captura. Se puede usar, por ejemplo, el análisis de Cot, en el que las especies abundantes son impulsadas cinéticamente a rehibridarse dejando las especies de una sola copia en un estado de una sola cadena capaz de hibridarse a sondas. Así pues, en realizaciones particulares, se puede tratar previamente una muestra de
 50 fragmentos de genoma con oligonucleótidos de Cot que sean complementarias a determinadas secuencias repetidas, o a otras secuencias que se desee valorar fuera de la muestra, antes de la exposición de la muestra a una matriz de sondas. En otro ejemplo, se puede enfriar una muestra de fragmentos de genoma hasta una temperatura y durante un breve período de tiempo que sean suficientes para que una fracción sustancial de las secuencias sobre-expresadas se vuelva a hibridar, pero insuficientes para la rehibridación sustancial de las secuencias presentes en
 55 bajos números de copias. La muestra resultante tendrá una cantidad reducida de secuencias repetidas disponibles para la posterior interacción con una matriz de sondas.

Es posible separar los fragmentos no deseados que forman especies de doble cadena, por ejemplo, en el análisis de Cot o la rehibridación de fragmentos de genoma, de las especies de una sola cadena basándose en las diferentes
 60 propiedades de los ácidos nucleicos de una sola cadena y de cadena doble. En una realización particular, se pueden usar enzimas que escindan preferentemente el ADN de doble cadena. Por ejemplo, la ADNsa I puede escindir ADN de doble cadena de 100 a 500 veces más rápido que el ADN de una sola cadena en condiciones conocidas. Por consiguiente, es posible eliminar los fragmentos no deseados mediante el tratamiento con oligonucleótidos de Cot o mediante la rehibridación de los fragmentos, y el tratamiento con ADNsa I en condiciones en las que los fragmentos
 65 no deseados formen preferentemente especies de doble cadena y se escindan. Además, se pueden usar otras enzimas que preferentemente modifiquen, escindan o se unan a especies de doble cadena en comparación con las especies de una sola cadena para separar las especies en un método de la invención, tales como endonucleasas de

restricción específicas de la secuencia o endonucleasas específicas de dúplex de cangrejo de Kamchatka.

También se puede usar la PCR de cebadores aleatorios para amplificar un ADN genómico en un método de la invención. La PCR de cebadores aleatorios se puede llevar a cabo mediante la replicación de una muestra de ADNg con un cebador en condiciones no rigurosas, de manera que el cebador se hibride aleatoriamente a varios locus del ADNg. Las posteriores etapas de PCR se pueden llevar a cabo en condiciones más rigurosas para amplificar los fragmentos generados debido al cebado aleatorio de la etapa anterior. La longitud, la secuencia o ambos parámetros de un cebador aleatorio se pueden seleccionar de acuerdo con la probabilidad de cebado en determinados intervalos a lo largo del ADNg. En este sentido, a medida que aumenta la longitud del cebador, aumentará el intervalo promedio entre las ubicaciones cebadas aleatoriamente, suponiendo que haya cambios en el resto de condiciones de amplificación. Del mismo modo, un cebador que tiene una secuencia complementaria o similar a una secuencia repetida se cebará con más frecuencia, produciendo intervalos más cortos entre los fragmentos amplificados que un cebador que carezca de secuencias que sean similares a las secuencias repetidas en un genoma por amplificar. La amplificación con cebadores aleatorios se puede llevar a cabo en condiciones similares a las descritas, por ejemplo, en Bassam *et al.*, *Australas Biotechnol.* 4:232-6 (1994). De acuerdo con la invención, la amplificación se puede llevar a cabo en condiciones isotérmicas usando un cebador aleatorio, condiciones de hibridación de baja rigurosidad y una polimerasa de desplazamiento de cadena.

Otro método que se puede usar para amplificar un genoma en la invención es la PCR inter-Alu. En este método, los cebadores están diseñados para hibridarse a secuencias Alu que se repitan a lo largo del genoma. La amplificación por PCR con estos cebadores producirá fragmentos flanqueados por repeticiones Alu. Los expertos en la materia reconocerán que se pueden llevar a cabo métodos similares con cebadores que se hibriden con otras secuencias repetidas en un genoma de interés, tales como regiones reguladoras de la transcripción, sitios de corte y empalme o similares. Además, los cebadores a secuencias repetidas se pueden usar en los métodos de amplificación isotérmica tales como los expuestos en el presente documento.

La complejidad y el grado de representación resultante de la amplificación con un determinado conjunto de cebadores se pueden ajustar usando diferentes condiciones de hibridación de los cebadores. En la presente invención, se puede usar una variedad de condiciones de hibridación tales como condiciones de alta, moderada o baja rigurosidad, incluyendo, pero sin limitación, las descritas en Sambrook *et al.*, *supra*, (2001) o en Ausubel *et al.*, *supra*, (1998). Las condiciones rigurosas favorecen la hibridación específica dependiente de la secuencia. En general, las secuencias más largas y las temperaturas más elevadas favorecen la hibridación específica dependiente de la secuencia. En Tijssen, *Techniques in Biochemistry and Molecular Biology—Hybridization with Nucleic Acid Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993), se encuentra una guía útil sobre la hibridación de ácidos nucleicos.

Las etapas de amplificación y detección usadas en la invención, generalmente, se llevan a cabo en condiciones rigurosas que permiten selectivamente la formación de un complejo de hibridación en presencia de secuencias complementarias. La rigurosidad se puede controlar mediante la modificación de un parámetro de la etapa que sea una variable termodinámica, incluyendo, pero sin limitación, la temperatura, concentración de formamida, concentración de sal, concentración de sal caotrópica, pH, concentración de disolvente orgánico, o similares. Estos parámetros también se pueden usar para controlar la unión inespecífica, como se describe en general en la patente de EE.UU. Nº 5.681.697. Por lo tanto, si se desea, se pueden realizar ciertas etapas en condiciones de una rigurosidad relativamente alta para reducir la unión inespecífica.

En general, las condiciones de alta rigurosidad incluyen temperaturas que son aproximadamente 5-10 °C inferiores al punto de fusión térmica (T_m) para las secuencias de hibridación a una determinada fuerza iónica y pH. Las condiciones de alta rigurosidad incluyen aquellas que permiten la unión de un primer ácido nucleico con un ácido nucleico complementario que tenga al menos aproximadamente un 90 % de pares de bases complementarios en su longitud y pueden incluir, por ejemplo, secuencias que sean al menos aproximadamente un 95 %, 98 %, 99 % o 100 % complementarias. Las condiciones rigurosas pueden incluir además, por ejemplo, aquellas en las que la concentración de sal es inferior a aproximadamente 1,0 M de iones de sodio (u otras sales), normalmente, una concentración aproximadamente 0,01 a 1,0 M a pH de 7,0 a 8,3 y la temperatura es al menos aproximadamente de 30 °C para las secuencias cortas de hibridación (por ejemplo, de 10 a 50 nucleótidos) y de al menos aproximadamente 60 °C para las secuencias largas de hibridación (por ejemplo, de más de 50 nucleótidos). Las condiciones de alta rigurosidad también se pueden lograr con la adición de agentes desestabilizantes de hélice, tales como la formamida. Las condiciones de alta rigurosidad pueden incluir, por ejemplo, condiciones equivalentes a la hibridación en formamida al 50 %, 5 x solución de Denhart, 5 x SSPE, SDS al 0,2 % a 42 °C, seguidas de lavado en 0,1 x SSPE, y SDS al 0,1 % a 65 °C. Los híbridos de ácido nucleico se pueden estabilizar aún más mediante la modificación covalente con uno o más agentes de reticulación.

Las condiciones moderadamente rigurosas incluyen aquellas que permiten la unión de un primer ácido nucleico con un ácido nucleico complementario que tenga al menos aproximadamente un 60 % de pares de bases complementarios a lo largo de del primer ácido nucleico. Dependiendo de las condiciones particulares de rigurosidad moderada usadas, se puede formar un híbrido entre secuencias que tenga la complementariedad de al menos aproximadamente un 75 %, 85 % o 90 % de los pares de bases a lo largo de la región hibridada. Las condiciones

moderadamente rigurosas incluyen, por ejemplo, condiciones equivalentes a la hibridación en formamida al 50 %, 5 x solución de Denhart, 5 x SSPE, SDS al 0,2 % a 42 °C, seguidas de lavado en 0,2 x SSPE, SDS al 0,2 % a 65 °C.

5 La hibridación de baja rigurosidad incluye, por ejemplo, condiciones equivalentes a la hibridación en formamida al 10 %, 5 x solución de Denhart, 6 x SSPE, SDS al 0,2 % a 42 °C, seguidas del lavado en 1 x SSPE, SDS al 0,2 % a 50° C. La solución de Denhart y SSPE son bien conocidos por los expertos en la materia como lo son otros tampones de hibridación adecuados (véase, por ejemplo, Sambrook *et al.*, *supra* (2001) o Ausubel *et al.*, *supra* (1998)).

10 En realizaciones de la invención en las que se modificará un híbrido, por ejemplo, mediante una polimerasa, las condiciones se pueden seleccionar además para adaptarse a la reacción de modificación en particular. Por ejemplo, cuando la modificación consiste en la replicación o amplificación, se pueden usar condiciones tales como las expuestas anteriormente con respecto a determinadas polimerasas. Se entenderá que el agente modificador tal como una polimerasa se puede añadir en cualquier momento durante una etapa de amplificación o detección, 15 incluyendo, por ejemplo, antes, durante o después de la adición de los componentes de ácido nucleico de la reacción de modificación.

Los métodos de la invención se pueden usar para amplificar un genoma nativo en una sola etapa de reacción o en un solo recipiente de reacción con el fin de producir una población representativa amplificada de fragmentos de 20 genoma que tenga una elevada complejidad. La capacidad de usar una sola etapa o un único recipiente de reacción proporciona la ventaja no limitante de aumentar la eficacia de la amplificación en comparación con los métodos que requieren múltiples etapas o recipientes de reacción. Además, en realizaciones particulares, se puede conseguir una población representativa amplificada de fragmentos de genoma alta complejidad en condiciones que no requieran la combinación de productos de múltiples reacciones de amplificación. Por lo tanto, los fragmentos de una población 25 representativa amplificada de fragmentos de genoma se pueden obtener en paralelo en vez de secuencialmente en diversas realizaciones de la invención. Sin embargo, es posible usar los métodos en realizaciones en las que se lleven a cabo diferentes etapas de reacción en recipientes separados, de forma secuencial, o en las que se combinen los productos de múltiples reacciones, por ejemplo, para adaptarse a aplicaciones particulares.

30 En la patente de EE.UU. N° 6.355.431, se puede encontrar una mayor descripción de métodos a modo de ejemplo que se pueden usar en la invención para amplificar ácidos nucleicos tales como genomas nativos o fragmentos de los mismos, e incluye la amplificación mediante la reacción en cadena de polimerasa (PCR), la PCR cebada aleatoriamente, la PCR cebada arbitrariamente, la amplificación de desplazamiento de cadena, la amplificación basada en secuencias de ácido nucleico y la amplificación mediada por la transcripción.

35 Tras la replicación de un genoma o población de fragmentos de genoma, los ácidos nucleicos que contienen una modificación deseada se pueden separar de los ácidos nucleicos no modificados, tales como cebadores que no hayan reaccionado o el molde. Por ejemplo, puede ser deseable eliminar cebadores no extendidos o sin reaccionar, porque los cebadores no extendidos pueden competir con los cebadores extendidos o marcados en una variedad de 40 métodos de detección que se usan en la invención, disminuyendo así la señal. Por consiguiente, se puede usar una serie de diferentes técnicas para facilitar la eliminación de los cebadores no extendidos. Aunque la siguiente descripción se dirige a reacciones de amplificación para mayor claridad, se entenderá que estas técnicas también se pueden usar para separar ácidos nucleicos modificados y no modificados en una etapa de detección.

45 La separación de los ácidos nucleicos puede estar mediada por la incorporación selectiva de un marcador, incluyendo, por ejemplo, uno o más de los marcadores primarios o secundarios descritos anteriormente en el presente documento. Los ácidos nucleicos que tienen un marcador secundario incorporado se pueden separar de los que carecen del marcador basándose, por ejemplo, en la unión a un receptor que tenga especificidad por el marcador. El receptor puede estar unido, por ejemplo, a un sustrato en fase sólida como se ha expuesto 50 anteriormente en relación con la realización ejemplificada en la Figura 9. Se pueden usar marcadores primarios para separar ácidos nucleicos en un método de clasificación, tal como la clasificación de células activadas por fluorescencia. Del mismo modo, los ácidos nucleicos que tienen un marcador secundario incorporado se pueden separar de los que carecen del marcador en un método de clasificación basado en la detección de un receptor que proporcione una marcador primario al complejo de ácido nucleico-receptor. La separación también se puede realizar 55 usando resinas de exclusión por tamaño convencionales, tales como la resina G-50, ultrafiltración tal como con columnas Amicon o Centricon, o métodos de precipitación de tipo etanol.

Un ácido nucleico se puede marcar convenientemente en un método de la invención con un resto introducido durante una reacción de amplificación o modificación a través de un cebador marcado, precursor de nucleótido 60 marcado o ambos. En realizaciones particulares, uno o más PNT usados para replicar un ácido nucleico pueden incluir un marcador detectable secundario que se puede usar para separar cebadores modificados de cebadores no modificados que carecen del marcador. Los marcadores secundarios encuentran un uso particular en las técnicas de detección, que incluyen etapas para la separación sondas marcadas y sin marcar, tales como SBE, OLA o escisión invasiva. Los marcadores particularmente útiles incluyen, pero sin limitación, uno de un par de parejas de unión; 65 restos químicamente modificables; o inhibidores de nucleasas.

A modo de ejemplo, un marcador secundario puede ser un hapteno o un antígeno que tenga afinidad por una inmunoglobulina, o fragmento funcional de la misma, unido a un soporte sólido. Los ácidos nucleicos marcados que están unidos a la inmunoglobulina se pueden separar de los ácidos nucleicos no marcados mediante la separación física del soporte sólido y la fracción soluble. Además, los sistemas de avidina/biotina, incluyendo, por ejemplo, los que utilizan estreptavidina, miméticos de biotina o ambos, se pueden usar para separar ácidos nucleicos modificados de los que no se han modificado. Normalmente, el más pequeño de las dos parejas de unión se une a un ácido nucleico. Sin embargo, la unión de la pareja mayor también puede ser útil. Por ejemplo, la adición de estreptavidina a un ácido nucleico aumenta su tamaño y cambia sus propiedades físicas, que se pueden aprovechar para la separación. Por consiguiente, un ácido nucleico marcado con estreptavidina se puede separar de los ácidos nucleicos no marcados de una mezcla usando una técnica tal como cromatografía de exclusión por tamaño, cromatografía de afinidad, filtración o precipitación diferencial.

En realizaciones que incluyen la unión de una pareja de unión a un soporte sólido, el soporte sólido se puede seleccionar, por ejemplo, de los descritos en el presente documento con respecto a las matrices de detección. Los sustratos particularmente útiles incluyen, por ejemplo, perlas magnéticas que se pueden introducir fácilmente en la muestra de ácido nucleico y retirarse fácilmente con un imán. También se pueden usar otros sustratos de cromatografía de afinidad conocidos. Se pueden usar métodos conocidos para unir una pareja de unión a un soporte sólido.

Por lo general, un método de detección de locus tipificables de un genoma se lleva a cabo en una población representativa amplificada de fragmentos de genoma obtenida, por ejemplo, mediante un método establecido anteriormente. Como alternativa, los locus tipificables se pueden determinar para una población representativa de fragmentos de genoma derivada de un genoma mediante por un método diferente de un método de amplificación. En una realización, se puede obtener una población representativa de fragmentos de genoma mediante la fragmentación de un genoma nativo. Los ejemplos de métodos que se pueden usar para fragmentar un genoma se exponen más adelante. Los expertos en la materia reconocerán que los métodos de fragmentación se pueden usar como una alternativa a los métodos de amplificación descritos en el presente documento o, si se desea en combinación con una técnica de amplificación.

Un genoma nativo aislado puede ser fragmentado por cualquier entidad física, química o bioquímica que cree roturas de doble cadena en el ADN. En realizaciones particulares, un genoma nativo puede ser digerido con una endonucleasa. Las endonucleasas útiles en los métodos de la invención incluyen aquellas que se escinden en una secuencia de reconocimiento específica o las que escinden de manera inespecífica ADN, tales como ADNsel. Las endonucleasas están disponibles en la técnica y se pueden obtener, por ejemplo, de fuentes comerciales tales como New England Biolabs (Beverly, Mass.) o Life Technologies Inc. (Rockville, Md.), entre otras. Se pueden usar endonucleasas específicas para generar fragmentos de polinucleótidos de un determinado tamaño medio, según la frecuencia con la que se espera que la enzima corte una secuencia aleatoria. Por ejemplo, se esperaría que una endonucleasa que tiene una secuencia de reconocimiento de seis nucleótidos produjera, como media, fragmentos de 4.096 pares de bases de longitud. La longitud media de los fragmentos se puede estimar mediante el tratamiento del ADN como una secuencia aleatoria y estimando la frecuencia de un sitio de reconocimiento en la secuencia aleatoria de acuerdo con la relación $4^n = s$, donde n es el número de bases reconocidas por la endonucleasa y s es el tamaño medio de los fragmentos producidos. Las condiciones de incubación también se pueden modificar, como se describe a continuación, para alterar la eficacia enzimática de la endonucleasa, alterando así el tamaño medio de los fragmentos producidos. Usando el ejemplo de una endonucleasa que tiene un sitio de reconocimiento de 6 pares de bases, una disminución en la eficacia enzimática puede producir fragmentos que sean, como media, de más de 4.096 pares de bases de longitud.

También se pueden usar endonucleasas inespecíficas para producir fragmentos de genoma de un tamaño medio deseado. Debido a que la reacción de las endonucleasas es bimolecular, se puede manipular la velocidad de fragmentación mediante la modificación de condiciones tales como las concentraciones de la endonucleasa, ADN o ambas. En concreto, se puede usar una reducción de la concentración de cualquiera de entre la endonucleasa, el ADN o ambos para reducir la velocidad de reacción, produciendo tamaños medios mayores de los fragmentos. El aumento de las concentraciones de cualquiera de entre la endonucleasa, la secuencia de reconocimiento de ADN o ambas permitirá una mayor eficacia, acercándose a la velocidad máxima (V_{max}) para la enzima en particular que conduce a la reducción de los tamaños medios de los fragmentos. También se pueden aplicar cambios similares en la condiciones a las endonucleasas específicas de sitio porque, sus reacciones con el ADN también son bimoleculares. Hay otras condiciones de reacción que también pueden afectar a la velocidad de escisión, incluyendo, por ejemplo, temperatura, concentración de sal y el tiempo de reacción. Los métodos para modificar las velocidades de reacción de las nucleasas con el fin de producir fragmentos de polinucleótidos de un determinado tamaño medio se describen, por ejemplo, en Sambrook *et al.*, *supra*, (2001) o en Ausubel *et al.*, *supra*, (1998).

Otros métodos que se pueden usar para producir fragmentos de genoma incluyen, por ejemplo, el tratamiento con agentes químicos que perturban la cadena principal de fosfodiéster del ADN, tales como los que escinden enlaces mediante un mecanismo de radicales libres, luz UV, ruptura mecánica o similar. Estos y los métodos expuestos anteriormente se pueden usar para producir fragmentos genómicos de un genoma nativo, escindir además fragmentos de genoma o escindir otros ácidos nucleicos usados en la invención. Los métodos de interrupción mecánica a modo de ejemplo adicionales que se pueden usar para producir fragmentos de genoma incluyen el

tratamiento con ultrasonidos y el cizallamiento.

La amplificación de genomas completos con cebadores aleatorios normalmente produce mayores rendimientos de amplificación y el aumento de la representación cuando se usa ADN genómico intacto como molde en comparación con moldes fragmentados. En aplicaciones de la invención donde se desee la amplificación de ADN genómico fragmentado, es posible ligar los fragmentos juntos para producir ADN concatenado. El ADN concatenada luego se puede usar en un método de amplificación de todo el genoma, tal como los expuestos anteriormente en el presente documento. Los ejemplos de condiciones que se pueden usar en una reacción de concatenación de fragmentos de genoma se describen, por ejemplo, en el documento WO 03/033724 A1.

En realizaciones en las que no se desee la fragmentación de una muestra de ácido nucleico diana, los fragmentos se pueden modificar para su uso en un método de la invención. Por ejemplo, un ADN genómico se puede modificar para facilitar la amplificación. Una modificación a modo de ejemplo que puede facilitar la amplificación es la concatenación de fragmentos de genoma para formar moldes extendidos que se puedan amplificar de manera eficaz, por ejemplo, mediante la amplificación con cebadores aleatorios. La concatenación se puede llevar a cabo, por ejemplo, mediante el tratamiento de una población de fragmentos de genoma con ARN ligasa T4 en condiciones conocidas en la técnica tales como las descritas en McCoy *et al.*, *Biochem.* 19:635-642 (1980). La concatenación también se puede llevar a cabo usando una mezcla de endonucleasa, polimerasa y ligasa de AP. El ADN dañado se puede reparar usando enzimas apropiadas tales como la mezcla de polimerasa Restorase® disponible EN Sigma-Aldrich (R1028). Otra modificación que se puede usar es la adición de colas universales a fragmentos de genoma. Los ejemplos de métodos de incorporación de colas universales incluyen, sin limitación, el tratamiento de los fragmentos con la transferasa de desoxinucleótidos terminales en los extremos 3' de la cola con un mononucleótido tal como dGTP. Por consiguiente, se puede añadir una cola poly G como cola universal a fragmentos de genoma. También se pueden añadir colas Poly C, T, U, A, u otras colas de nucleótidos. Las colas universales también se pueden añadir mediante el tratamiento de fragmentos de genoma con ARN ligasa T4 y oligonucleótidos que tengan un adaptador dúplex 4-mero aleatorio y una secuencia de cola universal en condiciones en las que se añada la secuencia de cola universal a uno o ambos extremos de los fragmentos de genoma.

El ejemplo X describe métodos para la amplificación de fragmentos producidos mediante el tratamiento con bisulfito del ADN metilado. Los expertos en la materia reconocerán que los métodos de amplificación descritos en el Ejemplo X se pueden usar para las muestras de fragmentos de ácido nucleico de cualquiera de una variedad de composiciones y producidos por cualquiera de una variedad de mecanismos. Otros ejemplos de fragmentos de ADN útiles en la invención incluyen, sin limitación, ADNc o ADN genómico degradado, por ejemplo, de tejidos o células almacenados tales como los que se almacenan fijadas con formalina, fijados en formaldehído, embebidos en parafina, embebidos en polímero, embebidos en etanol o por alguna combinación de los mismos. Lo ADN fragmentados también se pueden obtener de muestras forenses, muestras arqueológicas, muestras paleontológicas, muestras momificadas, muestras petrificadas y otras muestras que se han desintegrado debido a un período prolongado de tiempo entre la muerte de la célula o del tejido y el análisis de su ADN genómico. Un método de detección de locus tipificables de un genoma puede incluir además la etapa de poner en contacto los fragmentos de genoma con una pluralidad de sondas de ácido nucleico que tienen secuencias correspondientes a los locus tipificables en condiciones en las que se formen híbridos de sonda-fragmento. Una sonda usada en un método de la invención puede tener cualquiera de una variedad de composiciones o tamaños, siempre que tenga la capacidad de unirse a un ácido nucleico diana con especificidad de secuencia. Por lo general, una sonda usada en los métodos es un ácido nucleico que incluye, por ejemplo, uno que tiene una estructura nativa o un análogo del mismo. Los ejemplos de sondas de ácido nucleico que se pueden usar en un método de la invención incluyen, sin limitación, las expuestas anteriormente en relación con los cebadores y otros ácidos nucleicos útiles en la invención. Se entenderá, además, que también se pueden usar otras sondas específicas de secuencia en un método de la invención incluyendo, por ejemplo, péptidos, proteínas u otros compuestos poliméricos.

Las sondas de la presente invención pueden ser complementarias a locus tipificables u otras posiciones de detección que sean indicativas de la presencia de los locus tipificables en una población representativa de fragmentos de genoma. Por lo tanto, una etapa de detección de un locus tipificable de fragmentos de genoma puede incluir, por ejemplo, la detección del propio locus o la detección de otra secuencia que esté vinculada o asociada genéticamente. No es necesario que esta complementariedad sea perfecta. Por ejemplo, puede haber cualquier número de apareamientos erróneos de pares de bases dentro de un complejo de ácido nucleico hibridado, siempre que los apareamientos erróneos no impidan la formación de un complejo de hibridación suficientemente estable para su detección en las condiciones en las que se esté usando.

Además, las sondas de ácido nucleico usadas en un método de la invención pueden incluir regiones de secuencia que no sean complementarias a secuencias diana u otras secuencias presentes en una determinada población de fragmentos de genoma. Estas regiones de secuencias de complementariedad no con la diana pueden incluir, por ejemplo, secuencias de unión para la fijación de las sondas a un sustrato, sitios de hibridación para otros ácidos nucleicos tales como un cebador u otras secuencias deseadas. Una región de secuencia de complementariedad con la diana de una sonda de ácido nucleico puede tener una longitud que sea, por ejemplo, de al menos 10 nucleótidos. Las regiones de complementariedad con la diana más largas también pueden ser útiles, incluyendo, sin limitación, aquellas que sean de al menos aproximadamente 15, 20, 25, 35, 50, 70, 100, 500, 1.000 o 5.000 nucleótidos de

- longitud o más largas. Como se ha expuesto anteriormente, determinadas realizaciones de la invención proporcionan la capacidad de amplificar un genoma nativo para producir una población representativa de fragmentos de genoma relativamente pequeños. Una ventaja no limitante de la detección de locus tipificables de un genoma en fragmentos de genoma pequeños es que los locus que están relativamente cerca se pueden separar para la
- 5 detección individual. Por consiguiente, en realizaciones particulares tales como la detección de secuencias diana pequeñas, una región de complementariedad con la diana de una sonda de ácido nucleico puede ser como máximo de aproximadamente 100, 90, 80, 70, 60, 50, 40, 35, 30, 25, 20 o 10 nucleótidos de longitud.
- Las secuencias de complementariedad con la diana a modo de ejemplo que son útiles en la invención se exponen a
- 10 continuación en el contexto de diversas técnicas de detección. Los expertos en la materia entenderán que no es necesario limitar las sondas a su uso en la técnica de detección particular ejemplificada, sino que más bien se pueden usar en cualquiera de una variedad de diferentes técnicas de detección según se desee para una aplicación particular de la invención.
- 15 Una sonda usada en un método de la invención puede tener, además, una modificación, por ejemplo, para soportar un determinado método de detección. Por ejemplo, en realizaciones en las que no se desea la amplificación ni la modificación de una sonda en particular, la sonda puede tener una estructura que sea resistente a la modificación. Como ejemplos concretos, una sonda puede carecer de un grupo OH en 3' o un resto caperuza en 3', siendo por lo tanto inerte a la modificación con una polimerasa. En realizaciones particulares, una sonda puede incluir un
- 20 marcador detectable, incluyendo, sin limitación, uno o más de los marcadores de ácido nucleico primarios o secundarios establecidos anteriormente. Como alternativa, la detección se puede basar en una característica intrínseca de la sonda, fragmento o híbrido de manera que no se requiere marcaje. Los ejemplos de características intrínsecas que se pueden detectar incluyen, pero sin limitación, la masa, la conductividad eléctrica, la absorbancia energética, la fluorescencia o similares.
- 25 Se puede usar cualquiera de una variedad de condiciones para hibridar sondas con fragmentos de genoma incluyendo, sin limitación, las indicadas anteriormente en lo que se refiere a la hibridación de cebadores con dianas. En realizaciones particulares, las condiciones de hibridación pueden soportar la modificación o la replicación de la sonda, fragmento de genoma o ambos. Sin embargo, dependiendo del método de detección en el que se aplique la sonda, las condiciones de hibridación no necesitan soportar la modificación de un híbrido de sonda-fragmento. Por consiguiente, la presencia de un determinado fragmento se puede determinar basándose en una propiedad detectable del fragmento de genoma, la sonda o ambos. Las condiciones de hibridación adicionales a modo de ejemplo se exponen a continuación en lo que respecta a métodos de detección particulares.
- 30 Una pluralidad de fragmentos de genoma que está en contacto con sondas en un método de la invención puede representar toda o parte de una secuencia de genoma. Por consiguiente, la complejidad de la pluralidad de fragmentos de genoma puede ser equivalente al tamaño del genoma del que se amplificó o produjo de otro modo. Por ejemplo, una pluralidad de fragmentos de genoma humano que se pone en contacto con sondas puede tener una complejidad de aproximadamente 3,1 gigabases, que es aproximadamente equivalente a la del genoma de longitud completa. También se pueden usar representaciones de menor complejidad. De nuevo, usando el genoma humano como ejemplo no limitante, una pluralidad de fragmentos de genoma que se pone en contacto con sondas puede tener una complejidad de al menos aproximadamente 2 gigabases, que es una representación de aproximadamente el 60 % del genoma humano, o una complejidad de al menos aproximadamente 1 gigabase, que es una representación de al menos aproximadamente el 30 % del genoma humano. La complejidad de una pluralidad de sondas en contacto con sondas en un método de la invención puede ser, por ejemplo, al menos aproximadamente de 0,1, gigabases, 0,2 gigabases, 0,5 gigabases, 0,8 gigabases, 1 gigabase, 1,5 gigabases, 2 gigabases, 2,5 gigabases, 3 gigabases, 3,5 gigabases, 4 gigabases, 4,5 gigabases, 5 gigabases o superior.
- 35 Como, en un método de la invención, se usan pluralidades de fragmentos de genoma de mayor complejidad, normalmente, es deseable usar cantidades superiores de ADN. Por consiguiente, la cantidad de ADN de una pluralidad de fragmentos de genoma que se pone en contacto con sondas en un método descrito en el presente documento puede ser de al menos aproximadamente 1 ug, 10 ug, 50 ug, 100 ug, 150 ug, 200 ug, 300 ug, 400 ug, 500 ug, 1000 ug o superior (ug en el presente documento se refiere a un microgramo). En una muestra de fluido, puede haber una pluralidad de fragmentos de genoma a cualquier concentración que dé los resultados deseados, tal
- 40 como un nivel deseado de hibridación específica de la secuencia entre las sondas y los fragmentos o de cantidad de locus detectados. Por ejemplo, la concentración de una pluralidad de fragmentos de genoma en contacto con sondas en un método de la invención puede ser de al menos aproximadamente 0,1 ug/ul, 0,2 ug/ul, 0,5 ug/ul, 0,8 ug/ul, 1 ug/ul, 1,5 ug/ul, 2 ug/ul, 5 ug/ul, 10 ug/ul (ul en el presente documento se refiere a un microlitro).
- 50 El número de sondas puestas en contacto con una pluralidad de fragmentos de genoma es puede seleccionar en base a una aplicación deseada de los métodos. Las poblaciones de sonda y las matrices a modo de ejemplo que se pueden usar incluyen las conocidos en la técnica y/o expuestas en el presente documento. El número de diferentes sondas que forman híbridos específicos de la secuencia con fragmentos de genoma puede ser, por ejemplo, de al menos aproximadamente 100, 500, 1.000, 5.000, 1×10^4 , 5×10^4 , 1×10^5 , 5×10^5 , 1×10^6 , 5×10^6 , o superior,
- 65 incluyendo un número de sondas de una población o una matriz conocida en la técnica y/o expuesta en el presente documento.

Tras la hibridación, si se desea, se pueden separar los ácidos nucleicos no hibridados de los híbridos. Los ácidos nucleicos de una sola cadena y los ácidos nucleicos híbridos se pueden separar en base a propiedades que son diferentes para las dos especies, incluyendo, por ejemplo, tamaño, masa, absorbancia energética, fluorescencia, conductividad eléctrica, carga o afinidad por sustratos particulares. Los ejemplos de métodos que se pueden usar para separar ácidos nucleicos de una sola cadena y ácidos nucleicos híbridos en base a las propiedades que son diferentes para las dos especies incluyen, pero sin limitación, cromatografía de exclusión de tamaño, filtración a través de una membrana con un determinado corte de tamaño, cromatografía de afinidad, electroforesis en gel, electroforesis capilar, clasificación de células activadas por fluorescencia (FACS), y similares.

En una realización particular, la separación de ácidos nucleicos de una sola cadena, tales como sondas, dianas o ambos, de los ácidos nucleicos híbridos se puede facilitar mediante la unión de la sonda o diana a un sustrato. Un método a modo de ejemplo que incluye la separación de ácidos nucleicos usando un sustrato de fase sólida se muestra en la Figura 9 y se ha descrito anteriormente. Los híbridos formados en el ácido nucleico unido a un sustrato se pueden separar de los ácidos nucleicos no hibridados mediante la separación física del sustrato de la mezcla de reacción. Los sustratos a modo de ejemplo que se pueden usar para tal separación incluyen, sin limitación, partículas tales como perlas magnéticas, Sephadex®, vidrio de poro controlado, agarosa o similares; o superficies tales como superficies de vidrio, plástico, cerámica y similares. Los ácidos nucleicos se pueden unir a sustratos a través de enlazadores y ligandos conocidos tales como los expuestos anteriormente con respecto a los marcadores secundarios de ácido nucleico y usando métodos conocidos en la técnica. Los sustratos se pueden separar físicamente de una solución mediante cualquiera de una variedad de métodos que incluyen, por ejemplo, atracción magnética, sedimentación por gravedad, sedimentación centrífuga, filtración, FACS, atracción eléctrica, o similares. La separación también se puede llevar a cabo mediante el movimiento manual del sustrato, por ejemplo, usando las manos o un dispositivo robótico.

Un método de la invención puede incluir además una etapa de detección de locus tipificables de híbridos de sonda-fragmentos de genoma. Dependiendo de la aplicación particular de la invención, los híbridos de sonda-fragmentos de genoma se pueden detectar usando una técnica de detección directa o, como alternativa, una técnica basada en la amplificación. Las técnicas de detección directa incluyen aquellas en las que el nivel de ácidos nucleicos de los híbridos de sonda-fragmento proporciona la señal detectada. Por ejemplo, en el caso de un híbrido formado en un determinado lugar de la matriz, se puede detectar la señal procedente del lugar que surge del híbrido capturado o sus componentes de ácido nucleico sin la amplificación del híbrido ni de sus componentes de ácido nucleico. Como alternativa, la detección puede incluir la amplificación de la sonda o del fragmento de genoma o ambos para aumentar el nivel de ácido nucleico que se detecta. Como se expone a continuación en el contexto de diversas técnicas de detección a modo de ejemplo, se pueden marcar una sonda de ácido nucleico, fragmento de genoma o ambos. Además, los ácidos nucleicos de un híbrido de sonda-fragmento se pueden marcar antes, durante o después de la formación de los híbridos y la detección de locus tipificables en base a la detección de tales marcadores

Por consiguiente, un método de detección de locus tipificables de un genoma puede incluir las etapas de (a) proporcionar una población representativa amplificada de fragmentos de genoma que tenga tales locus tipificables; (b) poner en contacto los fragmentos de genoma de una pluralidad de sondas de ácido nucleico que tengan secuencias correspondientes a los locus tipificables en condiciones donde se formen híbridos de sonda-fragmento; y (c) detectar directamente locus tipificables de los híbridos de sonda-fragmento.

En general, la detección, bien directa o basada en una técnica de amplificación, se puede lograr mediante métodos que perciban propiedades que sean intrínsecas a los ácidos nucleicos o sus marcadores asociados. Las propiedades útiles incluyen, por ejemplo, las que se pueden usar para distinguir ácidos nucleicos que tienen locus tipificables de los que carecen de locus. Tales propiedades detectadas se pueden usar para distinguir diferentes ácidos nucleicos solos o en combinación con otros métodos tales como la unión a locus diferenciados de una matriz de detección. Los ejemplos de propiedades en las que se puede basar la detección incluyen, pero sin limitación, la masa, conductividad eléctrica, absorbancia de energía, fluorescencia o similares.

La detección de fluorescencia se puede llevar a cabo mediante la irradiación de un ácido nucleico o su marcador con una longitud de onda de excitación de la radiación y la detección de la radiación emitida por un fluoróforo en el mismo mediante métodos conocidos en la técnica y descritos por ejemplo, en Lakowicz, "Principles of Fluorescence Spectroscopy", II ed., Plenum Press, Nueva York (1999). Un fluoróforo se puede detectar en base a cualquiera de una variedad de fenómenos de fluorescencia incluyendo, por ejemplo, la longitud de onda de emisión, la longitud de onda de excitación, la intensidad de la transferencia de energía de resonancia de fluorescencia (FRET), la inactivación, anisotropía o tiempo de duración. La FRET se puede usar para identificar la hibridación entre un primer polinucleótido unido a un fluoróforo donante y un segundo polinucleótido unido a un fluoróforo aceptor debido a la transferencia de energía desde el donante excitado hacia el aceptor. Por lo tanto, la hibridación se puede detectar como un cambio en la longitud de onda causado por la reducción de la emisión del donante y la aparición de emisión del aceptor para el híbrido. Además, se puede usar la recuperación de la fluorescencia tras un fotoblanqueamiento (FRAP) para identificar la hibridación de acuerdo con el aumento de la fluorescencia producido en un lugar de la matriz previamente fotoblanqueada debido a la unión de un polinucleótido diana marcado con fluorescencia.

65

Otras técnicas de detección que se pueden usar para percibir o identificar ácidos nucleicos que tienen locus tipificables incluyen, por ejemplo, espectrometría de masas, que se puede usar para percibir un ácido nucleico en base a su masa; resonancia de plasmones superficiales, que se pueden usar para percibir un ácido nucleico en base a la unión a una secuencia complementaria inmovilizada en una superficie; espectroscopia de absorción, que se puede usar para percibir un ácido nucleico en base a la longitud de onda de la energía que absorbe; calorimetría, que se puede usar para percibir un ácido nucleico en base a cambios en la temperatura de su entorno debido a la unión con una secuencia complementaria; conductancia o impedancia eléctrica, que se puede usar para percibir un ácido nucleico en base a cambios en sus propiedades eléctricas o en las propiedades eléctricas de su entorno; resonancia magnética, que se puede usar para percibir un ácido nucleico en base a la presencia de núcleos magnéticos; u otras técnicas espectroscópicas o cromatográficas analíticas conocidas.

En realizaciones particulares, los locus tipificables de los híbridos de sonda-fragmento se pueden detectar en base a la presencia de la sonda, del fragmento o de ambos en el híbrido, sin la posterior modificación de las especies híbridas. Por ejemplo, se puede identificar un fragmento previamente marcado que tenga un determinado locus tipificable en base a la presencia del marcador en un lugar determinado de la matriz donde reside un complemento de ácido nucleico del locus.

La divulgación proporciona además un método para la detección de locus tipificables de un genoma que incluye las etapas de (a) proporcionar una población representativa amplificada de fragmentos de genoma que tenga locus tipificables; (b) poner en contacto los fragmentos de genoma con una pluralidad de sondas de ácido nucleico inmovilizadas que tengan secuencias correspondientes a los locus tipificables en condiciones en las que se formen híbridos de sonda inmovilizada-fragmento; (c) modificar los híbridos de sonda inmovilizada-fragmento; y (d) detectar una sonda o un fragmento que se haya modificado, detectando de ese modo los locus tipificables del genoma.

En una realización particular, se pueden modificar sondas de ácido nucleico en matriz mientras están hibridadas con fragmentos de genoma para la detección. Tales realizaciones incluyen, por ejemplo, las que utilizan ASPE, SBE, la amplificación por ligación de oligonucleótidos (OLA), la ligación por extensión (GoldenGate®), tecnología Invader, la escisión de sondas o la pirosecuenciación como se describe en la patente de EE.UU. N° 6.355.431 B1, el documento US N° de serie 10/177.727 y/o más adelante. Por lo tanto, la invención se puede llevar a cabo en un modo donde se modifica una sonda inmovilizada en lugar de un fragmento de genoma capturado por una sonda. Como alternativa, la detección puede incluir la modificación de los fragmentos de genoma mientras están hibridados con las sondas. Las modificaciones a modo de ejemplo incluyen las que son catalizadas por una enzima tal como una polimerasa. Una modificación útil puede ser la incorporación de uno o más nucleótidos o análogos de nucleótido a un cebador hibridado a una cadena de molde, donde el cebador puede ser bien la sonda o un fragmento de genoma de un híbrido de sonda-fragmento de genoma. Tal modificación puede incluir la replicación de todo o parte de un molde cebado. Se entenderá que la modificación que conduce a la replicación de solo una parte de una sonda de molde o fragmento de genoma es la detección sin amplificación del molde, pues el molde no se replica a lo largo de toda su longitud.

Los ensayos de extensión son útiles para la detección de locus tipificables. Los ensayos de extensión se llevan a cabo generalmente mediante la modificación del extremo 3' de un cebador de ácido nucleico cuando está hibridado a un segundo ácido nucleico. El segundo ácido nucleico puede actuar como molde para dirigir el tipo de modificación, por ejemplo, mediante interacciones de apareamiento de bases que se producen durante la extensión basada en la polimerasa del primer ácido nucleico para incorporar uno o más nucleótidos. Los ensayos de extensión basada en la polimerasa son particularmente útiles, por ejemplo, debido a la alta fidelidad relativa de las polimerasas y su relativa facilidad de aplicación. Los ensayos de extensión se pueden llevar a cabo para modificar sondas de ácidos nucleico que tengan extremos 3' libres, por ejemplo, cuando están unidas a un sustrato tal como una matriz. Los ejemplos de enfoques que se pueden usar incluyen, por ejemplo, la extensión de cebadores específicos del alelo (ASPE), la extensión de una sola base (SBE) o la pirosecuenciación.

En realizaciones particulares, la extensión de una sola base (SBE) se puede usar para la detección de locus tipificables. En la Figura 2, se muestra una representación esquemática ilustrativa de la SBE. En síntesis, la SBE utiliza una sonda de extensión que se hibrida con un fragmento de genoma diana en un lugar que es próximo o adyacente a una posición de detección, siendo la posición de detección indicativa de un determinado locus tipificable. Se puede usar una polimerasa para extender el extremo 3' de la sonda con un análogo de nucleótido marcado con un marcador de detección tal como los descritos anteriormente en el presente documento. En base a la fidelidad de la enzima, solo se incorpora un nucleótido a la sonda de extensión si es complementario a la posición de detección del fragmento de genoma diana. Si se desea, el nucleótido puede ser derivatizado de modo que no se puedan producir nuevas extensiones y, por lo tanto, se añada solo un único nucleótido. Es posible detectar la presencia del nucleótido marcado en la sonda extendida, por ejemplo, en un determinado lugar de una matriz, e identificar el nucleótido añadido para determinar la identidad del locus tipificable. La SBE se puede llevar a cabo en condiciones conocidas, tales como las descritas en la solicitud de patente de EE.UU. N° 09/425.633. Un nucleótido marcado se puede detectar usando métodos tales como los expuestos anteriormente o descritos en otra parte tales como Syvanen *et al*, *Genomics* 8:684-692 (1990); Syvanen *et al*, *Human Mutation* 3:172-179 (1994); las patentes de EE.UU. N° 5.846.710 y 5.888.819; Pastinen *et al*, *Genomics Res.* 7 (6): 606-614 (1997).

Un análogo de nucleótido útil para la detección por SBE puede incluir un dideoxinucleosidotrifosfato (también denominado desoxinucleótidos o ddNTP, es decir, ddATP, ddTTP, ddCTP y ddGTP), u otros análogos de nucleótidos que sean derivatizados para ser terminaciones de cadena. El uso de los nucleótidos de terminación de cadena marcados es útil, por ejemplo, en las reacciones que tienen más de un tipo de dNTP presente con el fin de evitar falsos positivos debido a la extensión más allá de la posición de detección. Los ejemplos de análogos son nucleótidos dideoxi-trifosfato (ddNTP) o terminadores Acyclo (Perkin Elmer, Foster City, CA). Generalmente, se puede usar un conjunto de nucleótidos que comprende ddATP, ddCTP, ddGTP y ddTTP, incluyendo al menos uno de ellos un marcador. Si se desea para una aplicación particular, se puede usar un conjunto de nucleótidos en el que los cuatro estén marcados. Los marcadores pueden ser iguales o, como alternativa, diferentes tipos de nucleótidos pueden tener diferentes marcadores. Como los expertos en la materia apreciarán, se puede añadir cualquier número de nucleótidos o análogos de los mismos a un cebador, siempre que una enzima polimerasa tenga un determinado nucleótido de interés incorporado en una posición de consulta que indique un locus tipificable.

Un nucleótido usado en un método de detección de SBE puede incluir además, por ejemplo, un marcador detectable, que puede ser un marcador detectable bien primario o secundario. Se puede usar cualquiera de una variedad de los marcadores de ácido nucleico indicados anteriormente en el presente documento en un método de detección de SBE. El uso de marcadores secundarios también puede facilitar la eliminación de sondas no extendidas en realizaciones particulares.

La disolución para SBE puede incluir también una enzima de extensión, tal como una ADN polimerasa. Las ADN polimerasas adecuadas incluyen, pero no se limitan a, el fragmento Klenow de la ADN polimerasa I SEQUENASE™ 1.0 y SEQUENASE™ 2.0 (U.S. Biochemical), la ADN polimerasa T5, la ADN polimerasa Phi29, Thermosequenase™ (Taq con la mutación Tabor-Richardson) y otras conocidas en la técnica o descritas en el presente documento. Si el nucleótido es complementario de la base en la posición de detección de la secuencia diana, la enzima de extensión se añadirá al cebador de extensión. De esta manera, el cebador de extensión se modifica, es decir, se extiende, para formar un cebador modificado.

En las realizaciones donde la cantidad de cebador sin extender excede en mucho al cebador marcado extendido resultante y el exceso de cebador no extendido compite con la detección del cebador marcado, se pueden eliminar los cebadores no extendidos. Se pueden eliminar, por ejemplo, los cebadores no extendidos de las reacciones de SBE que se desarrollan con cantidades pequeñas de ADN diana. En el presente documento se definen métodos útiles para eliminar los cebadores sin extender. Además, se pueden eliminar preferentemente las sondas monocatenarias de una matriz de sondas, dejando híbridos de sonda-diana bicatenarios utilizando los métodos que se muestran en detalle a continuación tales como el tratamiento con exonucleasa. Dichos métodos pueden proporcionar un aumento de la sensibilidad y de la detección selectiva del ensayo, eliminando, por ejemplo, el fondo que surge del marcado de la sonda dirigida sin molde.

Como apreciarán los expertos en la materia, la configuración de una reacción SBE puede tomar alguna de diversas formas. En realizaciones concretas, la reacción puede llevarse a cabo en disolución, y a continuación, las hebras sintetizadas recientemente se pueden detectar con las marcas detectables específicas de bases. Por ejemplo, se pueden hibridar directamente para capturar sondas que sean complementarias de los cebadores de extensión y, a continuación, se puede detectar la presencia de la marca. Dicha configuración es útil, por ejemplo, cuando los fragmentos de genoma se disponen en matrices como sondas de captura. De forma alternativa, se puede producir la reacción SBE sobre una superficie. Se puede capturar, por ejemplo un fragmento de genoma utilizando una primera sonda de captura, que se hibrida con un primer dominio diana del fragmento, y la reacción puede continuar de tal manera que la sonda queda modificada tal como se muestra en la Figura 2A.

La determinación de la base en la posición de detección se puede llevar a cabo en cualquiera de varias formas. En una realización concreta, se puede desarrollar una reacción mixta con dos, tres o cuatro nucleótidos diferentes, cada uno de los cuales incorpora una marca diferente. En esta realización, la marca de la sonda se puede distinguir de las marcas no incorporadas para determinar qué nucleótido se ha incorporado a la sonda. De forma alternativa, se pueden desarrollar reacciones discretas con un nucleótido marcado de forma diferente. Se puede llevar a cabo esto utilizando tanto una sonda unida a un único sustrato y reacciones secuenciales, o bien exponiendo la misma reacción a sondas unidas a múltiples sustratos, mostrándose el último caso en la Figura 2A. Por ejemplo, se puede añadir dATP a un híbrido sonda-fragmento, y evaluarse la señal generada; se puede eliminar el dATP y añadirse dTTP. De forma alternativa, se pueden utilizar cuatro matrices; la primera se hace reaccionar con dATP, la segunda con dTTP, etc., y evaluarse la presencia o ausencia de la señal en cada matriz.

De forma alternativa, se puede llevar a cabo un análisis ratiométrico, por ejemplo, se pueden detectar dos marcas, "A" y "B", sobre dos sustratos (por ejemplo, dos matrices). En esta realización, se llevaron a cabo dos conjuntos de reacciones de extensión del cebador, cada una en dos matrices, conteniendo cada reacción un conjunto completo de cuatro NTP de terminación de cadena. La primera reacción contiene dos nucleótidos marcados con "A" y dos nucleótidos marcados con "B" (por ejemplo, A y C pueden estar marcados con "A", y G y T pueden estar marcados con "B"). La segunda reacción también contiene las dos marcas, pero intercambiadas; por ejemplo, A y G están marcados con "A" y T y C están marcados con "B". Esta composición de reacción permite que un marcador bialélico se puntúe ratiométricamente; esto es, se compara la intensidad de las dos marcas en dos canales de "color"

diferentes para un único sustrato, utilizando datos procedentes de un conjunto de dos matrices hibridadas. Por ejemplo, si el marcador es A/G, se utiliza entonces la primera reacción de la primera matriz para calcular una puntuación de genotipación ratiométrica; si el marcador es A/C, entonces se utiliza para el cálculo la segunda reacción en la segunda matriz, si el marcador es G/T, entonces se utiliza la segunda matriz, etc. Este concepto se puede aplicar a todas las posibles combinaciones de marcadores bialélicos. De esta manera, puntuar un genotipo utilizando una puntuación ratiométrica de una única fibra puede permitir una genotipación más sólida utilizando una comparación de intensidades absolutas o normalizadas entre dos matrices diferentes.

La reacción SBE ilustrada en la Figura 2 demuestra una realización en la cual se llevan a cabo cuatro reacciones separadas sobre cuatro matrices independientes utilizando una única marca. Las realizaciones adicionales pueden incluir el uso de más de un tipo de marca en combinación con menos de cuatro poblaciones o matrices de sondas. Por ejemplo puede llevarse a cabo la SBE en un modo de dos colores utilizando una única reacción y una única población de sondas. En este modo, los cuatro nucleótidos de terminación de cadena pueden estar presentes soportando dos de los nucleótidos un primer tipo de marca y soportando los otros dos un segundo tipo de marca. La primera marca se puede usar para A y C, mientras que la segunda marca se utiliza para G y T (o G y U). Este esquema de marcado a modo de ejemplo permite la detección de casi un 80% de los SNP humanos que se producen naturalmente ya que los SNP humanos más abundantes son los polimorfismos A/G y C/T. Los expertos en la materia reconocerán que se pueden utilizar otros esquemas de marcado si se desea, por ejemplo, para adaptarlos a la abundancia de polimorfismos que se va a detectar en una aplicación concreta. El uso de SBE con múltiples tipos de marcas puede proporcionar la ventaja no limitante de reducir el número de matrices y reacciones requeridas para obtener los datos de genotipación.

La secuenciación de bases únicas (SBS) es un ensayo de extensión que se puede llevar a cabo tal como se ha definido anteriormente para SBE con la excepción que uno o más nucleótidos que no terminan cadenas se incluyen en la reacción de extensión. De esta manera, de acuerdo con la invención, se pueden incluir uno o más nucleótidos que no terminan cadena en una reacción SBE que incluye, por ejemplo, los definidos anteriormente.

Una realización a modo de ejemplo de SBS es llevar a cabo dos reacciones separadas en dos poblaciones de sondas independientes. Las dos reacciones separadas se llevan a cabo de forma ventajosa utilizando una única marca, sin embargo, si se desea, se puede utilizar más de un tipo de marca. La primera reacción puede incluir dos nucleótidos marcados diferentes que se pueden extender y que son capaces de hibridarse a 2 de los 4 nucleótidos que se producen naturalmente en el ADN genómico. La segunda reacción puede incluir 2 nucleótidos diferentes, estando los nucleótidos marcados y siendo capaces de hibridarse a los otros 2 nucleótidos que se encuentran naturalmente en el ADN genómico. Cada una de las dos reacciones puede estar desprovista de los nucleótidos que se encuentran en la otra reacción o puede incluir análogos de terminación de cadena de los nucleótidos que se encuentran en la otra reacción. A modo de ejemplo, la primera reacción (reacción AC en caliente) puede incluir dATP-biotina y dCTP-biotina. Esta primera reacción puede carecer de GTP, UTP y TTP. De forma alternativa, la primera reacción puede incluir didesoxiGTP y didesoxiUTP (o didesoxiGTP y didesoxiTTP). Continuando con el ejemplo, la segunda reacción (reacción GU en caliente) puede incluir dGTO-biotina y dUTP-biotina (o dGTP-biotina y dTTP-biotina). Esta segunda reacción puede carecer de CTP o ATP. De forma alternativa, la segunda reacción puede incluir didesoxiCTP y didesoxiATP. Este esquema de marcado a modo de ejemplo permite la detección de casi un 80% de los SNP humanos que se producen naturalmente ya que los SNP humanos más abundantes son polimorfismos A/G y C/T.

ASPE es un ensayo de extensión que utiliza sondas de extensión que difieren en la composición del nucleótido en su extremo 3'. Se muestra una representación en forma de diagrama a modo de ejemplo de ASPE en la Figura 2B. De manera breve, se puede llevar a cabo ASPE hibridando un fragmento de genoma diana con una sonda de extensión que tiene una porción de secuencia 3' que es complementaria de una posición de detección y una porción 5' que es complementaria de una secuencia que es adyacente a la posición de detección. La modificación dirigida al molde de la porción 3' de la sonda, por ejemplo, mediante la adición de un nucleótido marcado por una polimerasa da como resultado un producto de extensión marcado, pero solo si el molde incluye la secuencia diana. La presencia de dicho producto de extensión del cebador marcado se puede detectar a continuación, por ejemplo, basándose en su localización en una matriz para indicar la presencia de un locus tipable particular.

En realizaciones concretas, ASPE puede llevarse a cabo con múltiples sondas de extensión que tienen extremos 5' similares tales como los que se hibridan adyacentes a la misma posición de detección en un fragmento de genoma diana pero con extremos 3' diferentes, de tal manera que solo las sondas que tienen un extremo 3' que complementa la posición de detección están modificadas por una polimerasa. Tal como se muestra en la Figura 2B, una sonda que tiene una base 3' terminal que es complementaria de una posición de detección concreta se denomina una sonda de emparejamiento perfecto (PM) para la posición, mientras que las sondas que tienen una base 3' terminal con un emparejamiento incorrecto y no son capaces de extenderse en una reacción ASPE son sondas con un emparejamiento incorrecto (MM) para la posición. Se puede detectar la presencia del nucleótido marcado en la sonda PM y determinarse la secuencia 3' de la sonda para identificar un locus tipable concreto. Una reacción ASPE puede incluir 1, 2, o 3 sondas MM diferentes, por ejemplo, en localizaciones de matriz discretas, escogiéndose el número dependiendo de la diversidad que se produce en el locus concreto que se está ensayando. Por ejemplo, se pueden utilizar dos sondas para determinar qué 2 alelos de un locus concreto están presentes en

una muestra, mientras que se pueden tres sondas diferentes para distinguir los alelos de un locus de 3 alelos.

En realizaciones concretas, una reacción ASPE puede incluir un análogo de nucleótido que se derivatiza para que sea un terminador de cadena. De esta manera, una sonda PM en un híbrido sonda-fragmento se puede modificar para incorporar un análogo de nucleótido único sin extensión adicional. Los análogos de nucleótidos terminadores de cadena a modo de ejemplo incluyen, sin limitación, los anteriormente definidos con respecto a la reacción SBE. Además, uno o más nucleótidos utilizados en una reacción ASPE, tanto si son como si no son terminadores de cadena, pueden incluir una marca de detección tal como las descritas anteriormente en el presente documento. Por ejemplo, una reacción ASPE puede incluir un único dNTP marcado con biotina tal como se ilustra en el Ejemplo III. Si se desea, se puede marcar más de un nucleótido en la reacción ASPE. Por ejemplo, se pueden modificar las condiciones de reacción tales como las descritas en el Ejemplo II para incluir dCTP biotinilado así como dGTP biotinilado y dTTP biotinilado. Se puede llevar a cabo una reacción ASPE en presencia de los cuatro nucleótidos A, C, T, y G o en presencia de un subconjunto de estos nucleótidos que incluye, por ejemplo, un subconjunto que carece de cantidades sustanciales de uno o más de A, C, T o G.

La pirosecuenciación es un ensayo de extensión que se puede utilizar para añadir uno o más nucleótidos en una(s) posición(es) de detección; es similar a SBE excepto en que la identificación de los loci tipables se basa en la detección de un producto de reacción, el pirofosfato (PPi), producido durante la adición de un dNTP a una sonda extendida, en lugar de una marca unida al nucleótido. Se produce una molécula de PPi por cada dNTP añadido al cebador de extensión. Esto es, desarrollando reacciones secuenciales con cada uno de los nucleótidos, y controlando los productos de reacción, se determina la identidad de la base añadida. Se puede utilizar la pirosecuenciación en la invención utilizando condiciones tales como las descritas en el documento US 2002/0001801.

En realizaciones concretas, la modificación de los híbridos sonda-fragmento inmovilizados puede incluir la escisión o degradación de los híbridos que tienen una o más parejas de bases con emparejamiento incorrecto. Como con las otras modificaciones definidas en el presente documento, se pueden emplear condiciones que dan como resultado la modificación selectiva de híbridos que tienen uno o más emparejamientos incorrectos en comparación con los híbridos emparejados perfectamente. Por ejemplo, en un método de detección basado en ASPE, los híbridos sonda-fragmento emparejados de manera incorrecta se pueden escindir o degradar selectivamente en comparación con los híbridos sonda-fragmento emparejados de forma perfecta. Por ejemplo, se puede poner en contacto un híbrido con un agente que sea capaz de reconocer un emparejamiento incorrecto de una pareja de bases y modificar el híbrido emparejado de manera incorrecta tal como mediante la escisión del enlace. Los agentes a modo de ejemplo incluyen enzimas que reconocen y escinden híbridos que tienen parejas de bases emparejadas de manera incorrecta tales como la ADN glicosilasa, Cel I, T4 endonucleasa VII, T7 endonucleasa I, endonucleasa de frijol mungo o Mut-y u otras, tales como las descritas en Bradley y col., Nucl. Acids Res. 32:2632-2641 (2004). Los productos de escisión producidos a partir de híbridos emparejados de manera incorrecta se pueden eliminar, por ejemplo, mediante lavado.

De acuerdo con esto, un método de la invención puede incluir modificar híbridos sonda-fragmento inmovilizados utilizando ASPE junto con la escisión de híbridos sonda-fragmento emparejados de manera incorrecta. Una ventaja de utilizar ambas etapas de modificación en combinación es que se puede aumentar la especificidad en comparación con el uso de una sola de las etapas. Por ejemplo, en los casos donde se utiliza la detección de ASPE se obtiene un primer nivel de especificidad debido a la diferenciación de cebadores emparejados de manera correcta e incorrecta mediante la extensión de la polimerasa. En los casos en los que se produce la extensión no deseada del cebador emparejado de manera incorrecta, la escisión de los híbridos emparejados de manera incorrecta puede actuar para evitar la señal de artefacto debida a las sondas emparejadas de manera incorrecta, de este modo la especificidad y la sensibilidad del ensayo. De forma similar, se pueden aumentar la especificidad y la sensibilidad eliminando la señal de artefactos que surge debido a los híbridos emparejados de manera incorrecta formados en otros métodos de detección definidos en el presente documento tales como ensayos basados en ligadura. Se pueden eliminar los híbridos emparejados de manera incorrecta a partir de los híbridos inmovilizados en fase de disolución o en fase sólida de acuerdo con los métodos descritos en el presente documento.

En una realización concreta, se puede llevar a cabo una reacción ASPE en condiciones en la que la extensión de los híbridos sonda-fragmento emparejados de forma perfecta se impulsa hasta la finalización y se extienden también cantidades sustanciales de híbridos sonda-fragmento emparejados de manera incorrecta. Por ejemplo, en el caso de un locus que tiene un alelo A y B, se puede diseñar la sonda de emparejamiento perfecto frente al alelo A homocigótico formando un híbrido perfecto con un individuo AA y se puede diseñar la sonda emparejada de manera incorrecta frente al alelo B homocigótico, formando un híbrido perfecto con un individuo BB. De acuerdo con esto, el papel del emparejamiento perfecto y la sonda emparejada de manera incorrecta se puede invertir dependiendo de la muestra bajo observación. El producto de una extensión emparejada de manera incorrecta tendrá un par de bases emparejadas de manera incorrecta en el producto extendido y el emparejamiento perfecto no contendrá un emparejamiento incorrecto. La eliminación específica de la señal generada por la sonda emparejada de manera incorrecta, manteniendo a la vez la señal procedente de la extensión emparejada de manera perfecta intacta, puede añadir una segunda etapa de discriminación para crear una distinción más grande entre el emparejamiento perfecto y el emparejamiento incorrecto, creando un ensayo de genotipación más específico en comparación con la detección

basada únicamente en la modificación basada en la polimerasa de las sondas emparejadas de manera perfecta.

Si se desea, una sonda inmovilizada que no es parte de un híbrido de sonda-fragmento se puede modificar de manera selectiva en comparación con un híbrido de sonda-fragmento. Se puede usar la modificación selectiva de sondas no hibridadas para aumentar la especificidad y la sensibilidad del ensayo, por ejemplo, eliminando sondas que están marcadas de una manera independiente al molde durante un ensayo de extensión de la polimerasa. Una modificación selectiva particularmente útil es la degradación o la escisión de sondas monocatenarias que están presentes en una población o matriz de sondas tras el contacto con fragmentos diana en condiciones de hibridación. Las enzimas a modo de ejemplo que degradan ácidos nucleicos monocatenarios incluyen, sin limitación, Exonucleasa 1 o Exonucleasa lambda.

En realizaciones que utilizan sondas con hidroxilos reactivos en sus extremos 3' y extensión de la polimerasa, una exonucleasa útil es una que digiere preferentemente ADN monocatenario en la detección de 3' a 5'. De esta manera, los híbridos sonda-diana bicatenarios que se forman bajo condiciones de ensayo particulares están protegidos preferentemente de la degradación como está el saliente 3' de la diana que sirve como molde para la extensión de la polimerasa de la sonda. Sin embargo, las sondas monocatenarias que no se hibridan a la diana en las condiciones de ensayo se degradan preferentemente. Además, dicho tratamiento con la exonucleasa puede degradar preferentemente regiones monocatenarias de fragmentos de genoma u otros ácidos nucleicos en casos en los que los fragmentos o ácidos nucleicos están retenidos por una matriz debido a la interacción con porciones que interactúan sin sonda de los ácidos nucleicos diana. De esta manera, el tratamiento con la exonucleasa puede evitar artefactos que pueden surgir debido a una red con puentes de 2 o más ácidos nucleicos unidos a una sonda. La digestión con exonucleasa se lleva a cabo normalmente después de la etapa de extensión de la sonda.

En algunas realizaciones, la detección de loci tipables puede incluir la amplificación de dianas de fragmentos del genoma tras la formación de híbridos sonda-fragmento, dando como resultado un aumento significativo en el número de moléculas diana. Las técnicas de detección basadas en la amplificación de dianas pueden incluir, por ejemplo, la reacción en cadena de la polimerasa (PCR), amplificación del desplazamiento de hebra (SDA), o la amplificación basada en la secuencia de ácido nucleico (NASBA). De forma alternativa, en lugar de amplificar la diana, las técnicas alternativas pueden usar la diana como molde para replicar una sonda hibridada, permitiendo que un pequeño número de moléculas diana dé como resultado un gran número de sondas de señalización, que se pueden detectar a continuación. Las estrategias basadas en la amplificación de la sonda incluyen, por ejemplo, la reacción en cadena de la ligasa (LCR), la tecnología de ciclación de sonda (CPT), técnicas de escisión invasivas, tales como la tecnología Invader™, la tecnología de la replicasa Q-beta (QβR) o los ensayos de tipo sándwich. Dichas técnicas se pueden llevar a cabo, por ejemplo, en las condiciones descritas en el documento de los Estados Unidos con N° de serie 60/161.148, 09/553.993 y 090/556.463; y en la Patente de los Estados Unidos N° 6.355.431 B1, o tal como se muestra a continuación. Estas técnicas se ilustran a continuación, en el contexto de los fragmentos de genoma utilizados como ácidos nucleicos diana que se hibridan con sondas de ácidos nucleicos dispuestas en forma de matrices. Debe entenderse que dichas realizaciones de fragmentos de genomas pueden disponerse en forma de matrices como sondas e hibridarse con dianas de ácidos nucleicos.

La detección con amplificación de ligadura de oligonucleótidos (OLA) implica la ligadura dependiente de molde de dos sondas pequeñas en una única sonda larga, utilizando una secuencia diana de fragmento del genoma como molde. En una realización concreta, una secuencia diana monocatenaria incluye un primer dominio diana y un segundo dominio diana, que son adyacentes y contiguos. Una primera sonda OLA y una segunda sonda OLA se pueden hibridar con secuencias complementarias de los respectivos dominios diana. A continuación, las dos sondas OLA se unen de manera covalente entre sí para formar una sonda modificada. En realizaciones en las que las sondas se hibridan directamente adyacentes entre sí, puede producirse el enlace covalente mediante una ligasa. En una realización, una de las sondas de ligadura puede unirse a una superficie tal como una matriz o una partícula. En otra realización ambas sondas de ligadura pueden unirse a una superficie tal como una matriz o una partícula.

De forma alternativa, se puede usar un ensayo de ligadura de extensión (Golden-Gate™) donde las sondas que se hibridan no son contiguas y se añaden uno o más nucleótidos junto con uno o más agentes que se unen a las sondas a través de los nucleótidos añadidos. Los agentes a modo de ejemplo incluyen, por ejemplo, polimerasas y ligasas. Si se desea, los híbridos entre las sondas y dianas modificadas se pueden desnaturalizar, y repetirse el procedimiento de amplificación conduciendo a la generación de un combinado de sondas ligadas. Como anteriormente, estas sondas de ligadura de extensión pueden estar unidas, aunque no es necesario, a una superficie tal como una matriz o una partícula. Las condiciones adicionales para el ensayo de ligadura de extensión que son útiles en la invención se describen, por ejemplo, en la patente de los Estados Unidos N° 6.355.431 B1 y en la solicitud de los Estados Unidos con N° de Serie 10/177.727.

Se denomina OLA a la reacción en cadena de ligadura (LCR) cuando se usan dianas de fragmentos de genoma bicatenarios. En la LCR, la secuencia diana se puede desnaturalizar, y añadirse dos conjuntos de sondas: un conjunto tal como se ha reseñado anteriormente para una hebra de la diana, y un conjunto separado (es decir, los ácidos nucleicos de la sonda de los cebadores tercero y cuarto) para la otra hebra de la diana. Se pueden usar condiciones en las que la primera y segunda sondas se hibridan a la diana y se modifican para formar una sonda extendida. Tras la desnaturalización del híbrido de sonda modificado por la diana, se puede usar la sonda

modificada como molde, además de la segunda secuencia diana, para la unión de la tercera y cuarta sondas. De forma similar, la tercera y cuarta sondas ligadas pueden servir como molde para la unión de la primera y segunda sondas, además de la primera hebra diana. De esta manera, puede producirse una amplificación exponencial en lugar que solamente una lineal cuando se repite el procedimiento de desnaturalización y ligadura.

5 El producto de la sonda OLA modificado puede detectarse en cualquiera de una variedad de maneras. En una realización concreta, una reacción de modificación de sonda dirigida a molde se puede llevar a cabo en disolución y la sonda modificada hibridarse con una sonda de captura en una matriz. Una sonda de captura es generalmente complementaria de al menos una porción de la sonda OLA modificada. En una realización a modo de ejemplo la
10 primera sonda OLA puede incluir una marca detectable y la segunda sonda OLA puede ser sustancialmente complementaria de la sonda de captura. Una ventaja no limitante de esta realización es que los artefactos debidos a la presencia de sondas marcadas que no se modifican en el ensayo se minimizan porque las sondas no modificadas no incluyen la secuencia complementaria que se hibrida mediante la sonda de captura. Una técnica de detección OLA también puede incluir la etapa de eliminar las sondas marcadas no modificadas a partir de una mezcla de
15 reacción antes de poner en contacto la mezcla de reacción con una sonda de captura tal como se describe, por ejemplo, en la Patente de los Estados Unidos N° 6.355.431 B1.

De forma alternativa, se puede inmovilizar un fragmento de genoma sobre una superficie en fase sólida y llevarse a cabo una reacción para modificar las sondas OLA hibridadas sobre la superficie de la fase sólida. Se pueden
20 eliminar las sondas no modificadas mediante lavado en las condiciones de restricción adecuadas. A continuación se pueden eluir las sondas modificadas a partir del fragmento de genoma diana utilizando las condiciones de desnaturalización, tales como NaOH 0,1 N y detectarse tal como se describe en el presente documento. Otras condiciones en las cuales el fragmento de genoma se puede detectar como una secuencia diana en una técnica OLA incluyen, por ejemplo, las descritas en las Patentes de los Estados Unidos con números 6,355.431 B1,
25 5.185.243, 5.679.524 y 5.573.907; documentos EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; WO 97/31256; y WO 89/09835, y documentos de los Estados Unidos con N°s de serie. 60/078,102 and 60/073,011.

Se pueden detectar loci tipables en un método de la invención utilizando amplificación de sonda circular de moléculas de ácido nucleico circularizadas (RCA). En una primera reacción, se puede hibridar una única sonda con un fragmento de genoma diana de tal manera que la sonda se circulariza a la vez que se hibrida con la diana. Cada término de la sonda se hibrida de manera adyacente sobre el ácido nucleico diana y la adición de una polimerasa da como resultado la extensión de la sonda circular. Sin embargo, debido a que la sonda no tiene final, la polimerasa
30 continúa extendiendo la sonda repetidamente. Esto da como resultado la amplificación de la sonda circular. Siguiendo con la RCA, se puede detectar la sonda circular amplificada. Esto se puede llevar a cabo de una variedad de maneras; por ejemplo, el cebador se puede marcar o la polimerasa puede incorporar nucleótidos marcados y detectarse los productos marcados mediante una sonda de captura en una matriz de detección. La amplificación de sonda circular de moléculas de ácido nucleico circularizadas puede llevarse a cabo en condiciones como las que se describen generalmente en Baner y col. (1998) Nuc. Acids Res. 26: 5073-5078; Barany, F. (1991) Proc. Natl. Acad.
35 Sci. USA 88: 189-193; y Lizardi y col. (1998) Nat Genet. 19:225-232.

Además, las sondas circulares de moléculas de ADN circularizadas utilizadas en la invención pueden tener características estructurales que las convierten en incapaces de replicarse cuando no se hibridan con una diana. Por ejemplo, uno o ambos de los términos que se hibridan con la diana pueden tener una secuencia que forme una estructura de tallo intramolecular, tal como una estructura de horquilla. Se puede preparar la estructura de tallo de una secuencia que permita la apertura de la sonda circular que se va a circularizar cuando se hibrida con una secuencia diana legítima pero da como resultado la inactivación de sondas circulares abiertas no circularizadas. Esta inactivación reduce o elimina la capacidad de la sonda circular abierta de cebar la síntesis de una sonda modificada en un ensayo de detección o de servir como molde para la amplificación de sonda circular de moléculas
45 de ácido nucleico circularizadas. Las sondas a modo de ejemplo capaces de formar estructuras de tallos intramoleculares y los métodos para su uso se pueden utilizar en la invención tal como se describe en la Patente de los Estados Unidos N° 6573051.

En otra realización, la detección puede incluir OLA seguido de RCA. En esta realización, un cebador inmovilizado se puede poner en contacto con un fragmento de genoma. Las secuencias complementarias se hibridarán entre sí dando como resultado un duplete inmovilizado. Se puede poner también en contacto un segundo cebador con el ácido nucleico diana. El segundo cebador se hibrida con el ácido nucleico diana adyacente al primer cebador. Se puede llevar a cabo una reacción OLA para unirse con el primer y el segundo cebador tal como un producto de cebador modificado, por ejemplo, tal como se ha descrito anteriormente. AS continuación se puede eliminar el fragmento de genoma y el producto del cebador modificado inmovilizado, hibridarse con una sonda de RCA que es complementaria del producto del cebador modificado pero no con el cebador inmovilizado no modificado. A continuación se puede llevar a cabo una reacción de RCA.
60

En una realización concreta se puede usar una sonda de cadena tanto para OLA y el molde circular para la RCA. Cada término de la sonda de cadena puede contener una secuencia complementaria de un fragmento de genoma diana. De forma más específica, el primer extremo de la sonda de cadena puede ser sustancialmente
65

complementario de un primer dominio diana, y el segundo extremo de la sonda de RCA puede ser sustancialmente complementario de un segundo dominio diana, adyacente al primer dominio. La hibridación de la sonda de cadena con el fragmento del genoma diana da como resultado la formación de un complejo de hibridación que contiene una sonda circular que actúa como un complejo del molde de RCA. La adición de una polimerasa al complejo del molde de RCA puede permitir la formación de un ácido nucleico del producto amplificado. Tras la RCA, se puede detectar el ácido nucleico del producto amplificado, por ejemplo, mediante hibridación con una matriz tanto directa como indirectamente y detectarse una marca asociada.

Una sonda de cadena utilizada en la invención puede incluir adicionalmente otras características tales como una secuencia adaptadora, un sitio de restricción para la escisión de concatémeros, una secuencia de una marca, o un sitio de cebado para el cebado de la reacción de RCA tal como se describe, por ejemplo, en la Patente de los Estados Unidos N° 6.355.431 B1. La misma patente describe también los métodos de la sonda de cadena que se pueden usar para detectar los loci tipables de los fragmentos de genoma dianas en un método de la invención.

Una variación de la LCR que se puede usar para detectar los loci tipables en un método de la invención utiliza la ligadura química en condiciones tales como las descritas en las Patentes de los Estados Unidos con números 5.616.464 y 6.767.259. En esta realización similar a la modificación enzimática, se puede utilizar una pareja de sondas, donde la primera sonda es sustancialmente complementaria de un segundo dominio adyacente de la diana. Cada sonda puede incluir una porción que actúa como una "cadena secundaria" que forma la mitad de una estructura de tallo no covalente entre las sondas, más bien que la unión con la secuencia diana. Las realizaciones particulares utilizan ácidos nucleicos sustancialmente complementarios como cadenas secundarias. De esta manera, tras la hibridación de las sondas con la secuencia diana, las cadenas secundarias de las sondas se colocan en proximidad espacial. Al menos una de las cadenas secundarias puede incluir un agente de reticulación activable, generalmente unido de forma covalente a la cadena secundaria que, tras su activación, da como resultado una reticulación química o una ligadura química con la sonda adyacente. El grupo activable puede incluir cualquier resto que permita la reticulación de las cadenas secundarias, e incluye grupos activados química, fotónica o térmicamente, tales como grupos fotoactivables. En algunas realizaciones, un único grupo activable en una de las cadenas secundarias es suficiente para dar como resultado una reticulación a través de la interacción con un grupo funcional en la otra cadena secundaria; en realizaciones alternativas, se pueden incluir grupos activables en cada cadena secundaria. Se pueden marcar una o ambas sondas

Una vez que se forma un complejo de hibridación, y que se ha activado el agente de reticulación de tal manera que las sondas se han unido covalentemente entre sí, la reacción se puede someter a las condiciones que permitan la disociación del complejo de hibridación, liberando de esta manera la diana para servir como molde para la siguiente ligadura o reticulación. De esta manera, se puede producir la amplificación de la señal, y se pueden detectar los productos reticulados, por ejemplo, mediante hibridación con una matriz tanto directa como indirectamente y detectarse una marca asociada.

En realizaciones concretas, la detección basada en la amplificación puede conseguirse utilizando técnicas de escisión invasivas. Al utilizar dicha solución, se puede hibridar un fragmento de genoma diana con dos sondas distintas. Las dos sondas son una sonda invasora, que es sustancialmente complementaria de una primera porción de fragmento de genoma diana, y una sonda señal, que tiene un extremo 3' sustancialmente complementario de una secuencia que tiene una posición de detección y un extremo 5' no complementario que puede formar una cola monocatenaria. La cola puede incluir una secuencia de detección y normalmente contiene también al menos una marca detectable. Sin embargo, debido a que la secuencia de detección en una sonda señal puede funcionar como una secuencia diana de una sonda de captura, pueden utilizarse configuraciones de tipo sándwich que utilizan sondas de marca tal como se describe en el presente documento y la sonda señal no necesita incluir una marca detectable.

La hibridación de las sondas invasora y señal próximas o adyacentes entre sí en un fragmento de genoma diana pueden formar cualquiera de varias estructuras útiles para la detección del híbrido de fragmentos de la sonda. Por ejemplo, puede formar una estructura de escisión ahorquillada, proporcionando por tanto un sustrato para una nucleasa que escinde la secuencia de detección procedente de la sonda señal. El sitio de escisión está controlado por la distancia o solapamiento entre el extremo 3' de la sonda invasora y la horquilla en la dirección 3' de la sonda señal. Por tanto, ni el oligonucleótido se escinde cuando se hibrida de manera incorrecta ni cuando se desune a un fragmento de genoma diana.

En las realizaciones concretas, se puede usar una nucleasa termoestable que reconoce la estructura de escisión ahorquillada y cataliza la liberación de la cola, permitiendo por tanto la ciclación térmica de la reacción de escisión y amplificarse, si se desea. Las nucleasas a modo de ejemplo que se pueden usar incluyen, sin limitación, las derivadas de *Thermus aquaticus*, *Thermus flavus*, o *Thermus thermophilus*; las descritas en las Patentes de los Estados Unidos con números 5.719.028 y 5.843.669, o endonucleasas Flap (FEN) tal como se describe, por ejemplo en la patente de los Estados Unidos N°. 5.843.669 y Lyamichev y col., Nature Biotechnology 17: 292-297 (1999).

Si se desea, la porción 3' la sonda señal escindida puede extraerse, por ejemplo, uniéndose con una etiqueta de captura en fase sólida, tal como una perla unida a estreptavidina, o mediante reticulación a través de una etiqueta de

captura para producir agregados. La secuencia de detección 5' de una sonda señal, puede detectarse utilizando los métodos que se muestran a continuación tales como la hibridación con una sonda en una matriz. Puede usarse además en la invención tecnología de escisión invasiva, utilizando las condiciones y los métodos de detección descritos, por ejemplo en las Patentes de los Estados Unidos con números 6.355.431; 5.846.717; 5.614.402; 5.719.028; 5.541.311; o 5.843.669.

Una técnica de detección basada en la amplificación que se puede usar adicionalmente para detectar loci tipables es la tecnología de la sonda de ciclación (CPT). Una sonda de CPT puede incluir dos secuencias de sonda separadas por un enlace escindible. La sonda CPT es sustancialmente complementaria de una secuencia diana de un fragmento de genoma y de esta manera se hibridará con éste para formar un híbrido de fragmentos de la sonda. La sonda CPT puede hibridarse con un fragmento de genoma diana en un método de la invención. Normalmente, la temperatura y la secuencia de la sonda se seleccionan de tal manera que la sonda principal se disociará. Dependiendo de la aplicación concreta, la CPT se puede llevar a cabo en disolución, o bien la sonda diana o la sonda escindible se pueden unir a un soporte sólido. Un híbrido de fragmentos de la sonda formado en los métodos se puede someter a condiciones de escisión que dan lugar a que el enlace escindible se escinda selectivamente, sin escindir la secuencia diana, separando por tanto las secuencias de las dos sondas. A continuación, las secuencias de las dos sondas se pueden disociar de la diana. En realizaciones concretas, se puede usar la sonda en exceso y dejar que se repita la reacción algún número de veces de tal manera que se amplifique la cantidad eficaz de sonda escindida.

Cualquier enlace dentro de una sonda CPT se puede escindir de forma selectiva cuando la sonda es parte de un complejo de hibridación, esto es, cuando se forma un complejo bicatenario como un enlace escindible. Se pueden usar en la invención cualquiera de una variedad de enlaces escindibles incluyendo, por ejemplo, ARN que se puede escindir en un híbrido de ADN:ARN por diversas nucleasas bicatenarias tales como ribonucleasas. Dichas nucleasas hendirán o escindirán de forma selectiva los nucleósidos de ARN a partir de un complejo de hibridación de ARN:ADN más bien que el ADN en el mencionado híbrido o en un ADN monocatenario. Los ejemplos adicionales de enlaces escindibles y agentes de escisión que se pueden usar en la invención se describen en la Patente de los Estados Unidos N° 6.355.431 B1 y en las referencias citadas en la anterior

Tras la finalización de una reacción de escisión de CPT, las sondas escindibles sin escindir se pueden eliminar o neutralizar si se desea, antes de la detección de las sondas escindidas para evitar señales positivas falsas. Esto se puede llevar a cabo mediante cualquiera de una variedad de formas que incluyen, por ejemplo, la unión de las sondas a un soporte sólido antes de la escisión, de tal manera que, tras la reacción de la CPT, las sondas escindidas que se han liberado en la disolución, se pueden separar físicamente de las sondas sin escindir que permanecen sobre el soporte. Las sondas sin escindir y escindidas se pueden separar también en función de sus diferencias de longitud, captura de una marca de unión concreta o utilizando la secuencia, por ejemplo, los métodos descritos en la patente de los Estados Unidos N° 6.355.431.

Las sondas escindidas producidas por una reacción de CPT se pueden detectar utilizando métodos tales como la hibridación con una matriz u otros métodos que se muestran en el presente documento. Por ejemplo, una sonda escindida se puede unir con una sonda de captura, bien de forma directa como indirecta, y detectarse una marca asociada. La tecnología CPT se puede llevar a cabo en las condiciones descritas, por ejemplo, en las Patentes de los Estados Unidos con números 5.011.769; 5.403.711; 5.660.988; y 4.876.187, y en las solicitudes PCT publicadas WO 95/05480; WO 95/1416, y WO 95/00667, y en el documento de los Estados Unidos con N° de Serie 09/014.304.

En realizaciones concretas, se puede usar la CPT con una sonda que contiene un enlace escindible para detectar los emparejamientos incorrectos, como se describe de forma general en las patentes de los Estados Unidos con números 5.660.988, y WO 95/14106. En dichas realizaciones, la secuencia del enlace escindible se puede colocar en una posición en una secuencia más larga que corresponde a una secuencia concreta que se va a detectar, es decir, el área de un posible emparejamiento incorrecto. En algunas realizaciones de la detección del emparejamiento incorrecto, la velocidad de generación de los fragmentos liberados es tal que los métodos proporcionan, esencialmente, un resultado sí / no, en el que la detección de virtualmente cualquier fragmento liberado indica la presencia de un locus tipable deseado. De forma alternativa o adicional, la cantidad final de fragmentos escindidos se puede cuantificar para indicar la presencia o ausencia de un locus tipable.

Los loci tipables de híbridos sonda-fragmento también se pueden detectar en un método de la invención utilizando un ensayo de tipo sándwich. Un ensayo de tipo sándwich es una técnica basada en la amplificación en la que múltiples sondas, normalmente marcadas, se unen a un único fragmento de genoma diana. En una realización a modo de ejemplo, un fragmento de genoma diana se puede unir a un sustrato sólido a través de una sonda de captura complementaria. Normalmente, una única sonda de captura estará presente para cada secuencia de locus tipable que se va a detectar. En el caso de una matriz de perlas, cada perla puede tener una de las únicas sondas de captura. Si se desea, se pueden usar sondas extensoras de captura, que permiten que una superficie universal tenga un único tipo de sonda de captura que se puede usar para detectar múltiples secuencias diana. Las sondas extensoras de captura incluyen una primera porción que se hibridará con todo o parte de la sonda de captura, y una segunda porción que se hibridará con una primera porción de la secuencia diana que se va a detectar. De acuerdo con esto, se pueden generar sondas solubles personalizadas, que los expertos en la materia apreciarán que pueden

simplificar y reducir costes en muchas aplicaciones de la invención. En realizaciones concretas, se pueden usar dos sondas extensoras de captura. Esto puede proporcionar, una ventaja no limitante para estabilizar complejos sujetos a ensayo, por ejemplo, cuando una secuencia diana que se va a detectar es grande o cuando se usan sondas amplificadoras grandes (particularmente sondas amplificadoras ramificadas o dendríméricas).

5 Una vez que un fragmento de genoma diana se ha unido a un sustrato sólido, tal como una perla, a través de una sonda de captura, una sonda amplificadora se puede hibridar con el fragmento para formar un híbrido de fragmentos de sonda. Las sondas amplificadoras a modo de ejemplo que se pueden usar en un método de la invención y las condiciones para su uso en ensayos de tipo sándwich se describen en la Patente de los Estados Unidos nº 10 6.355.431. De forma breve, una sonda amplificadora es un ácido nucleico que tiene al menos una secuencia de sonda, y al menos una secuencia de amplificación. Se puede usar una primera secuencia de sonda de una sonda amplificadora, ya sea de forma directa como indirecta, para hibridarse a una secuencia diana de un fragmento de genoma. Una secuencia de amplificación de una sonda amplificadora puede ser cualquiera de una variedad de 15 secuencias que se usan, bien de manera directa o de manera indirecta, para unirse a una primera porción de una sonda de marca. Normalmente, una sonda amplificadora incluirá una pluralidad de secuencias de amplificación. Las secuencias de amplificación se pueden unir entre sí de una variedad de maneras incluyendo, por ejemplo, unirse covalentemente de manera directa entre sí, o para intervenir secuencias de restos químicos.

20 Las sondas de marca que comprenden marcas detectables pueden hibridarse con fragmentos de genoma formando de este modo híbridos sonda-fragmento y las marcas se pueden detectar para determinar la presencia de loci tipables. Se pueden usar las secuencias de amplificación de la sonda amplificadora, tanto directa como indirectamente, para unirse a una sonda de marca para permitir la detección. Se puede llevar a cabo la detección de las reacciones de amplificación de la invención, que incluyen la detección directa de los productos de la 25 amplificación y la detección indirecta utilizando sondas de marca (es decir, ensayos de tipo sándwich), detectando los complejos sujetos a ensayo que tienen marcas. Los métodos a modo de ejemplo para usar un ensayo de tipo sándwich y los ácidos nucleicos asociados que se pueden utilizar en la presente invención se describen adicionalmente en el documento de los Estados Unidos con Nº de Serie 60/073.011 y en las Patentes de los Estados Unidos con números 6.355.431; 5.681.702; 5.597.909; 5.545.730; 5.594.117; 5.591.584; 5.571.670; 5.580.731; 5.571.670; 5.591.584; 5.624.802; 5.635.352; 5.594.118; 5.359.100; 5.124.246 y 5.681.697.

30 Dependiendo de la aplicación concreta de los métodos de la invención, se pueden usar las técnicas de detección que se muestran anteriormente para detectar los fragmentos de genoma principales diana o para detectar las dianas en una población representativa amplificada de fragmentos de genoma.

35 En realizaciones concretas, puede ser deseable eliminar los ácidos nucleico nucleicos sin extender o sin reaccionar procedentes de una mezcla de reacción antes de la detección debido a que los cebadores sin extender o sin reaccionar pueden a menudo competir con las sondas modificadas durante la detección, disminuyendo por tanto la señal. La concentración de las sondas sin modificar con respecto a las sondas modificadas puede ser a menudo relativamente alta, por ejemplo, en realizaciones en las que se utiliza un gran exceso de sonda. De acuerdo con 40 esto, se pueden utilizar numerosas técnicas diferentes para facilitar la eliminación de los cebadores sin extender. Los métodos a modo de ejemplo que se pueden usar para eliminar cebadores sin extender incluyen, por ejemplo, los descritos en la Patente de los Estados Unidos Nº 6.355.431.

45 Tal como se muestra anteriormente, la invención se puede usar para detectar uno o más loci tipables. En particular, la invención será muy adecuada para la detección de una pluralidad de loci tipables debido a los métodos que permiten distinguir a los loci individuales en pluralidades grandes y complejas. Los loci tipables individuales se pueden distinguir en la invención basándose en la separación de los loci en los fragmentos de genoma individuales, la formación de híbridos sonda-fragmento y la detección de híbridos sonda-fragmento separados físicamente. Se puede conseguir la separación físicas de los híbridos sonda-fragmento en la invención uniendo los híbridos o sus 50 componentes con uno o más sustratos. En realizaciones concretas, se puede distinguir un híbrido de fragmentos de sonda de otras sondas y fragmentos en una pluralidad basándose en la localización física del híbrido sobre la superficie de un sustrato tal como una matriz. Un híbrido de fragmentos de sonda puede unirse también a una partícula. Las partículas se pueden detectar de forma discreta en su localización y distinguirse de otras sondas y fragmentos de acuerdo con la detección discreta de la partícula sobre una superficie tal como una matriz de perlas o 55 en una muestra de fluido tal como una corriente de fluido en un citómetro de flujo: Los formatos a modo de ejemplo para distinguir los híbridos sonda-fragmento de loci tipables individuales se muestran adicionalmente en detalle a continuación.

60 La detección de loci tipables en una población representativa amplificada de fragmentos de genoma puede emplear matrices. En realizaciones en las que se van a detectar números relativamente grandes de loci, las matrices son preferiblemente matrices de alta densidad. Las micromatrices a modo de ejemplo que se pueden usar en la invención incluyen, sin limitación, las descritas en Butte, Nature Reviews Drug Discov. 1:951-60 (2002) o en las Patentes de los Estados Unidos con números 5.429.807; 5.436.327; 5.561.071; 5.583.211; 5.658.734; 5.837.858; 5.874.219; 5.919.523; 6.136.269; 6.287.768; 6.288.220; 6.297.006; 6.291.193; 6.346.413; 6.416.949; 65 6.482.591; 6.514.751 y 6.610.482; y en los documentos WO 93/17126; WO 95/11995; WO 95/35505; EP 742 287; y EP 799 897. Los ejemplos adicionales de formatos de matriz que son útiles en la invención se describen en la

Patente de los Estados Unidos N° 6.355.431 B1, el documento de los Estados Unidos 2002/0102578 y la Publicación PCT N° WO 00/63437. Los formatos a modo de ejemplo que se pueden usar en la invención para distinguir en una muestra de fluido utilizando dispositivos microfluidicos se describen, por ejemplo, en la Patente de los Estados Unidos N° 6.524.793. Los formatos de fluidos comercialmente disponibles para distinguir perlas incluyen, por ejemplo, los utilizados en las tecnologías xMAP™ de Luminex o los métodos MPSS™ de Lynx Therapeutics. Se pueden usar diversas técnicas y tecnologías para sintetizar matrices de materiales biológicos sobre o en un sustrato o soporte para formar micromatrices. Por ejemplo, se pueden sintetizar matrices Affymetrix® GeneChip® de acuerdo con técnicas denominadas algunas veces como tecnologías VLSIPS™ (Síntesis de Polímeros Inmovilizados a Muy Grande Escala). Algunos aspectos de VLSIPS™ y de otras micromatrices y métodos de fabricación de matrices de polímeros (que incluyen proteínas) se han descrito en la Patente de los Estados Unidos N° 09/536.841, Publicación Internacional N° WO 00/58516; Patentes de los Estados Unidos con números 5.143.854, 5.242.974, 5.252.743, 5-324.633, 5.445.934, 5.744.305, 5.384.261, 5.405.783, 5.424.186, 5.451.683, 5.482.867, 5.491.074, 5.527.681, 5.550.215, 5.571.639, 5.578.832, 5.593.839, 5.599.695, 5.624.711, 5.631.734, 5.795.716, 5.831.070, 5.837.832, 5.856.101, 5.858.659, 5.936.324, 5.968.740, 5.974.164, 5.981.185, 5.981.956, 6.025.601, 6.033.860, 6.040.193, 6.090.555, 6.136.269, 6.269.846, 6.022.963, 6.083.697, 6.291.183, 6.309.831 y 6.428.752; y en las Solicitudes PCT N° PCT/US99/00730 (Publicación Internacional N° WO 99/36760) y PCT/US01/04285.

Al utilizar VLSIPS™, se puede fabricar una matriz GeneChip haciendo reaccionar la superficie hidroxilada de una oblea de cuarzo de 5 pulgadas cuadradas (32,25 cm²) con silano. A continuación pueden unirse los enlazadores con las moléculas de silano. La distancia entre estas moléculas de silano determina la densidad de empaquetado de las sondas, permitiendo a las matrices mantener aproximadamente 500.000 localizaciones de sondas, o características, en solo 1,28 centímetros cuadrados. Se pueden sintetizar millones de moléculas de ADN idénticas a cada característica utilizando un procedimiento fotolitográfico en el que las máscaras, que transportan ventanas de 18 a 20 micrómetros cuadrados que corresponden a las dimensiones de las características individuales, se colocan sobre la oblea recubierta. Cuando la luz ultravioleta incide sobre la máscara en la primera etapa de la síntesis, los enlazadores expuestos quedan desprotegidos y están dispuestos para el acoplamiento de nucleótidos. Una vez que se han activado las características deseadas, se pueden enjuagar una disolución que contiene un único tipo de desoxinucleótido con un grupo de protección extraíble sobre la superficie de la oblea. El nucleótido se une a los enlazadores activados, iniciando el procedimiento de síntesis. Se puede usar una etapa de protección de los extremos de la cadena para truncar los enlazadores sin reaccionar (o los polinucleótidos en la etapa posterior). En la siguiente etapa de síntesis, se puede colocar otra máscara sobre la oblea para permitir el siguiente ciclo de desprotección y acoplamiento. El procedimiento se repite hasta que las sondas alcanzan su longitud completa, usualmente 25 nucleótidos. Sin embargo, las sondas que tienen otras longitudes tales como las que se muestran en otra parte en el presente documento se pueden unir también a cada característica. Una vez que la síntesis se ha completado, las obleas se pueden desproteger, trocearse, y las matrices individuales resultantes se pueden empaquetar en cartuchos de flujo continuo.

Se puede usar también una matriz moteada en un método de la invención. Una matriz moteada a modo de ejemplo es la matriz CodeLink™ disponible de Amersham Biosciences. Los portas activados con CodeLink™ están revestidos con un polímero hidrófilo de cadena larga que contiene grupos reactivos a amina. Este polímero se reticula de manera covalente consigo mismo y con la superficie del porta. Se puede llevar a cabo la unión a la sonda a través de la interacción covalente entre el extremo 5' modificado con amina de la sonda de oligonucleótidos y los grupos reactivos a amina presentes en el polímero. Las sondas se pueden unir a localizaciones discretas utilizando plumas de moteado. Las plumas útiles son plumas capilares de acero inoxidable que se cargan individualmente con un resorte. Los volúmenes de carga de la pluma pueden ser menos de aproximadamente 200 nl con un volumen de administración de aproximadamente 0,1 nl o menos. Dichas plumas se pueden usar para crear características que tienen un diámetro de la mancha de, por ejemplo, aproximadamente 140-160 μm. En una realización preferida, las sondas de ácido nucleico en cada característica moteada pueden tener 30 nucleótidos de longitud. Sin embargo, las sondas que tienen otras longitudes tales como las que se muestran en otra parte en el presente documento pueden también unirse a cada mancha.

Una matriz que es útil en la invención también puede fabricarse utilizando métodos de impresión de chorro de tinta tales como la tecnología Sureprint™ disponible de Agilent Technologies. Se pueden usar dichos métodos para sintetizar sondas de oligonucleótidos in situ o para unir sondas presintetizadas que tienen restos que son reactivos con una superficie del sustrato. Una micromatriz impresa puede contener 22.575 características sobre una superficie que tiene dimensiones de porta normalizadas (aproximadamente 1 pulgada por 3 pulgadas) (2,54 cm x 7,62 cm). Normalmente, las sondas impresas tienen 25 o 60 nucleótidos de longitud. Sin embargo, las sondas que tienen otras longitudes tales como las que se muestran en otra parte en el presente documento pueden también imprimirse en cada localización.

Para algunas de las realizaciones descritas en el presente documento, las sondas de ácidos nucleicos se unen a sustratos de tal manera que tienen un extremo 3' libre para la modificación por las enzimas u otros agentes. Los expertos en la materia reconocerán que los métodos ilustrados anteriormente con respecto a la síntesis de ácidos nucleicos en la dirección 3' a 5' pueden modificarse para producir ácidos nucleicos que tienen extremos 3' libres. Por ejemplo, los métodos sintéticos conocidos en la técnica para sintetizar ácidos nucleicos en la dirección 5' a 3' y que tienen uniones 5' a soportes sólidos, se pueden usar en una impresión de chorro de tinta o en un método

fotolitográfico. Además, puede llevarse a cabo la inversión in situ de los ácidos nucleicos unidos a sustrato en su extremo 5' y desunirse en su extremo 3'. Se puede llevar a cabo la inversión de acuerdo con los métodos conocidos en la técnica tales como los descritos en Kwiatkowski y col., Nucl. Acids Res. 27: 4710-4714 (1999).

5 Una matriz de alta densidad a modo de ejemplo es una matriz de matrices o una matriz de material compuesto que tiene una pluralidad de matrices individuales que está configurada para permitir el procesamiento de múltiples muestras. Dichas matrices permiten la detección multiplete de loci tipables. Las matrices de material compuesto a modo de ejemplo que se pueden usar en la invención, por ejemplo, en formatos de detección multiplete se describen en la Patente de los Estados Unidos N° 6.429.027 y en el documento de los Estados Unidos 2002/0102578. En las realizaciones concretas cada matriz individual puede estar presente en cada pocillo de una placa de microvaloración. De esta manera, dependiendo del tamaño de la placa de microvaloración y del tamaño de la matriz individual, se pueden desarrollar simultáneamente cantidades muy grandes de ensayos; por ejemplo, utilizando matrices individuales de 2.000 y una placa de valoración de 96 pocillos, se pueden llevar a cabo 192.000 ensayos en paralelo; el mismo número de matrices en cada pocillo de una placa de microvaloración de 384 pocillos da como resultado 768.000 ensayos simultáneos, y en una placa de microvaloración proporciona 3.072.000 ensayos.

En realizaciones concretas, los ácidos nucleicos útiles para detectar loci tipables de un genoma se pueden unir a partículas que se disponen en forma de matriz o se distinguen espacialmente de otra manera. Las partículas a modo de ejemplo incluyen microesferas o perlas. Sin embargo, las partículas utilizadas en la invención no necesitan ser esféricas. Más bien, se pueden usar partículas que tienen otras formas que incluyen, pero no se limitan a, formas de discos, placas, chips, lonchas o formas irregulares. Además, las partículas usadas en la invención pueden ser porosas, aumentando de esta manera el área superficial disponible para la unión o el ensayo de híbridos sonda-fragmentos. El tamaño de las partículas puede variar, por ejemplo, entre nanómetros, tales como aproximadamente perlas de 100 nm, a milímetros, tales como aproximadamente perlas de 1 mm con partículas de tamaño intermedio, siendo de utilidad tales como máximo de aproximadamente 0,2 micrómetros, 0,5 micrómetros, 5 micrómetros o 200 micrómetros. La composición de las perlas puede variar dependiendo, por ejemplo, de la aplicación de la invención o del método de síntesis. Las composiciones de perlas adecuadas incluyen, pero no se limitan a, las utilizadas en la síntesis de péptidos, ácidos nucleicos y restos orgánicos, tales como plásticos, materiales cerámicos, vidrio, poliestireno, metilestireno, polímeros acrílicos, materiales paramagnéticos, disoluciones de óxido de torio, grafito de carbono, dióxido de titanio, látex o dextranos reticulados, tales como Sepharose TM, celulosa, nilón, micelas reticuladas de Teflon TM. Se describen partículas útiles, por ejemplo, en Microsphere Detection Guide de Bangs Laboratories, Fishers Ind.

Algunas realizaciones de detección basada en matrices en la invención se ilustran a continuación para las perlas o las microesferas. Los expertos en la materia reconocerán que las partículas de otras formas y tamaños, tales como las que se muestran anteriormente se pueden usar en lugar de las perlas o las microesferas ilustradas para estas realizaciones.

Cada partícula utilizada para la detección de loci tipables en una población de fragmentos de genoma puede incluir una sonda de captura asociada. Sin embargo, si se desea, se pueden incluir una o más partículas en una matriz o población de partículas que no contiene una sonda de captura. Una sonda de captura puede ser cualquier molécula o material que se une de manera directa o indirecta a un ácido nucleico que tiene una secuencia diana tal como un locus tipable. Una sonda de captura puede ser, por ejemplo, un ácido nucleico que tiene una secuencia que se hibrida con un ácido nucleico complementario u otra molécula que se une a un ácido nucleico de una manera específica de secuencia.

En una realización concreta, cada perla u otra localización de matriz puede tener un único tipo de sonda de captura. Sin embargo, si se desea, una pluralidad de sondas se puede unir a cada perla. Por ejemplo, una perla u otra localización de matriz pueden tener dos o más sondas que se hibridan con diferentes porciones del mismo fragmento de genoma. Las sondas pueden hibridarse a localizaciones adyacentes o a localizaciones que están separadas entre sí del ácido nucleico diana capturado. El uso de esta realización de captura de múltiples sondas puede aumentar la especificidad de la detección en comparación con el uso de solo una de las sondas. De esta manera, en casos en los que se desean sondas más pequeñas, se puede emplear una estrategia de múltiples sondas para proporcionar una especificidad comparable a las realizaciones en las que se utilizan sondas más largas. De forma similar, se puede usar una subpoblación de más de una microesfera que contiene una sonda de captura concreta para detectar loci tipables de un genoma en la invención. De esta manera, puede desarrollarse una redundancia en el sistema de ensayo mediante el uso de subpoblaciones de microesferas para sondas concretas.

En algunas realizaciones, se pueden sintetizar sondas de polímeros tales como ácidos nucleicos o péptidos mediante la adición secuencial de unidades monoméricas directamente sobre un soporte sólido utilizado en una matriz tal como una perla o la superficie de un porta. Se pueden usar en la invención métodos conocidos en la técnica para la síntesis de una variedad de diferentes compuestos químicos sobre soportes sólidos, tales como métodos para la síntesis en fase sólida de péptidos, restos orgánicos, y ácidos nucleicos. De forma alternativa, se pueden sintetizar en primer lugar sondas, y a continuación unirse covalentemente a un soporte sólido. Se pueden unir las sondas a grupos funcionales sobre un soporte sólido. Se pueden producir, si se desea, soportes sólidos funcionalizados mediante métodos conocidos en la técnica obtenidos de cualquiera de diversos suministradores

comerciales de perlas y otros soportes que tienen químicas superficiales que facilitan la unión de una funcionalidad deseada por un usuario. Las químicas superficiales a modo de ejemplo que son útiles en la invención incluyen, pero no se limitan a, grupos amino tales como aminas alifáticas y aromáticas, ácidos carboxílicos, aldehídos, amidas, grupos clorometilo, hidracida, grupos hidroxilo, sulfonatos o sulfatos. Si se desea, se puede unir una sonda a un soporte sólido a través de un enlazador químico. Dicho enlazador puede tener características que proporcionan, por ejemplo, una unión estable, una unión reversible, suficiente flexibilidad para permitir una interacción deseada con un fragmento de genoma que tiene un locus tipable que se va a detectar, o para evitar reacciones de unión indeseables. Los métodos a modo de ejemplo adicionales que se pueden usar en la invención para unir las sondas de polímero a un soporte sólido se describen en Pease y col., Proc. Natl. Acad. Sci. USA 91(11): 5022-5026 (1994); Khrapko y col., Mol Biol (Moscú) (USSR) 25: 718-730 (1991); Stimpson y col., Proc. Natl. Acad. Sci. USA 92: 6379-6383 (1995) o Guo y col., Nucleic Acids Res. 22: 5456-5465 (1994).

Generalmente, se puede configurar una matriz de matrices en cualquiera de diversas maneras. En una realización concreta, como se describe de forma más completa a continuación, se puede usar un sistema de un componente. Esto es, se puede configurar un primer sustrato que tiene una pluralidad de localizaciones de ensayo, tal como una placa de valoración, de tal manera que cada localización del ensayo contiene una matriz individual. De esta manera, la localización del ensayo y la localización de la matriz pueden ser la misma. Por ejemplo, el material plástico de una placa de microvaloración se puede formar para contener una pluralidad de pocillos de perlas en la parte inferior de cada uno de los pocillos del ensayo. Las perlas que contienen las sondas de captura de la invención se pueden cargar a continuación en los pocillos de perlas en cada localización del ensayo como se describe de forma más completa a continuación.

De forma alternativa, se puede utilizar un sistema de dos componentes. En esta realización, se pueden formar matrices individuales sobre un segundo sustrato, que a continuación se incluyen o sumergen en el sustrato de la primera placa de valoración. Una realización concreta utiliza paquetes de fibra óptica como matrices individuales grabadas en una superficie de cada fibra individual, de tal manera que las perlas que contienen las sondas de captura están cargadas en el extremo del paquete de la fibra óptica. La matriz de material compuesto incluye de esta manera numerosas matrices individuales que se configuran para incluirse en el interior de los pocillos de una placa de microvaloración.

De acuerdo con esto, la presente invención proporciona una matriz de material compuesto que tiene al menos un primer sustrato con una superficie que tiene una pluralidad de localizaciones de ensayo. Se pueden usar en la invención cualquiera de una variedad de matrices que tiene una pluralidad de agentes candidatos en un formato de matriz. El tamaño de una matriz usada en la invención puede variar dependiendo de la composición de la sonda y del uso deseado de la matriz. Se pueden preparar matrices que contienen entre aproximadamente 2 sondas diferentes a muchos millones. Generalmente, una matriz puede tener entre dos a como mucho, un billón o más de localizaciones de matriz por cm cuadrado. Una localización de matriz puede ser, por ejemplo, un área sobre una superficie en la cual una sonda o población de sondas similares están unidas o una partícula. En el caso de una partícula, su localización de matriz puede ser una coordenada fija sobre un sustrato al cual se une o asocia, o una coordenada relativa en comparación con las localizaciones de una o más de otras partículas de referencia en una muestra de fluido tales como una corriente que pasa a través de un citómetro de flujo. Las matrices de muy alta densidad son útiles en la invención que incluye, por ejemplo, aquellas que tienen entre aproximadamente 10.000.000 localizaciones de matriz/cm² a aproximadamente 2.000.000.000 localizaciones de matriz/cm² o entre aproximadamente 100.000.000 localizaciones de matriz/cm² a aproximadamente 1.000.000.000 localizaciones de matriz/cm². Se pueden usar también matrices de alta densidad incluyendo, por ejemplo, aquellas en el intervalo entre aproximadamente 100.000 localizaciones de matriz/cm² a aproximadamente 10.000.000 localizaciones de matriz/cm² o aproximadamente 1.000.000 localizaciones de matriz/cm² a aproximadamente 5.000.000 localizaciones de matriz/cm². Las matrices de densidad moderada útiles en la invención pueden variar entre aproximadamente 10.000 localizaciones de matriz/cm² a aproximadamente 100.000 localizaciones de matriz/cm², o entre aproximadamente 20.000 localizaciones de matriz/cm² a aproximadamente 50.000 localizaciones de matriz/cm². Las matrices de baja densidad tienen generalmente menos de 10.000 partículas/cm² siendo útiles entre aproximadamente 1.000 localizaciones de matriz/cm² a aproximadamente 5.000 localizaciones de matriz/cm². Son también útiles en algunas aplicaciones matrices de muy baja densidad que tienen menos de 1.000 localizaciones de matriz/cm², entre aproximadamente 10 localizaciones de matriz/cm² a aproximadamente 1000 localizaciones de matriz/cm², o entre aproximadamente 100 localizaciones de matriz/cm² a aproximadamente 500 localizaciones de matriz/cm². Los métodos de la invención no necesitan llevarse a cabo en un formato de matriz, por ejemplo, en realizaciones en las que se van a detectar uno o un pequeño número de loci. Si se desea, se pueden usar matrices que tienen múltiples sustratos, incluyendo, por ejemplo, sustratos que tienen composiciones diferentes o idénticas. De esta manera, por ejemplo, las matrices grandes pueden incluir una pluralidad de sustratos más pequeños.

Para algunas aplicaciones, el número de matrices individuales se configura mediante el tamaño de la placa de microvaloración utilizada; de esta manera, las placas de valoración de 96 pocillos, 384 pocillos y 1536 pocillos utilizan matrices de materiales compuestos que comprenden 96, 384, y 1536 matrices individuales. Como apreciarán los expertos en la materia, cada placa de valoración no necesita contener una matriz individual. Debe señalarse que las matrices de materiales compuestos pueden incluir matrices individuales que son idénticas, similares o diferentes. Por ejemplo, una matriz de material compuesto que tiene 96 matrices similares puede utilizarse en aplicaciones en

las que se desea determinar la presencia o ausencia de los mismos 2.000 loci tipables para 96 muestras diferentes. De forma alternativa, se puede usar una matriz de material compuesto que tiene 96 matrices diferentes cada una con 2.000 sondas diferentes, en aplicaciones en las que se desea determinar la presencia o ausencia de 192.000 loci tipables para una única muestra. Se pueden usar combinaciones alternativas, en las que hileras, columnas u otras porciones de una matriz de microvaloración formateada sean las mismas, por ejemplo, en los casos en los que se desea redundancia. Como apreciarán los expertos en la materia, existe una variedad de manera de configurar el sistema. Además, la naturaleza aleatoria de las matrices puede significar que se puede añadir la misma población de perlas a dos diferentes superficies, dando como resultado matrices sustancialmente similares pero quizá no idénticas.

Puede prepararse un sustrato usado en una matriz de la invención a partir de cualquier material que se pueda modificar para contener sitios individuales discretos y sea manejable para al menos un método de detección. En las realizaciones en las que se usan matrices de partículas se pueden usar como un material que es capaz de unirse o asociarse con uno o más tipos de partículas. Los sustratos útiles incluyen, pero no se limitan a, vidrio; vidrio modificado; vidrio funcionalizado, plásticos tales como acrílicos, poliestireno y copolímeros de estireno y otros materiales, polipropileno, polietileno, polibutileno, poliuretanos, Teflon, o similares; polisacáridos; nilón; nitrocelulosa, resinas, sílice, materiales basados en sílice tales como silicio o silicio modificado; carbono; metal, vidrio inorgánico; paquetes de fibra óptica o cualquiera de una variedad de diferentes polímeros. Los sustratos útiles incluyen aquellos que permiten la detección óptica, por ejemplo, siendo translúcidos a la energía de una longitud de onda de detección deseada y/o no tienen una fluorescencia apreciable por sí mismos en una longitud de onda de detección deseada.

Generalmente un sustrato usado para una matriz de la invención tiene una superficie plana o planar. Sin embargo, se pueden usar tan bien otras configuraciones de sustratos. Por ejemplo, se pueden usar configuraciones tridimensionales que incluyan una matriz, tal como una matriz de perlas en un material poroso, tal como un bloque de plástico, que permita el acceso de la muestra a las localizaciones de matriz y el uso de un microscopio confocal para la detección. De forma similar, las localizaciones del ensayo pueden colocarse en la superficie interior de un tubo, para el análisis de la muestra mediante flujo pistón. Los sustratos a modo de ejemplo que son útiles en la invención incluyen, pero no se limitan a, paquetes de fibra óptica, o sustratos planos planares tales como vidrio, poliestireno, u otros plásticos y materiales acrílicos.

La superficie de un sustrato puede incluir una pluralidad de localizaciones de matriz individuales que están físicamente separadas entre sí. Por ejemplo, la separación física se puede deber a la presencia de pocillos de ensayo, tales como en una placa de microvaloración. Otras barreras que se pueden usar para separar físicamente las localizaciones de matriz incluyen, por ejemplo, regiones hidrófobas que impedirán el flujo de los disolventes acuosos o regiones hidrófilas que impedirán el flujo de disolventes apolares o hidrófobos

Las localizaciones de la matriz que están físicamente separadas entre sí forman las localizaciones de ensayo. Una localización de ensayo puede incluir una matriz de sondas y proporciona un recipiente para mantener un fluido de tal manera que el fluido pone en contacto las sondas. Por ejemplo, un fluido que contiene fragmentos de genoma puede ponerse en contacto con sondas en condiciones de hibridación que se muestran en el presente documento o se conocen en la técnica. De forma similar, un fluido de lavado o fluido que contiene otros reactivos o analitos descritos en el presente documento se puede poner en contacto con una matriz de sondas cuando se coloca en una localización de ensayo. Si se desea, se pueden encerrar en un recinto una localización de ensayo. Los recintos a modo de ejemplo incluyen, sin limitación, un casete, un pocillo encerrado, o una superficie de porta encerrada por una junta o membrana, o ambas. Los recintos a modo de ejemplo adicionales que son útiles en la invención se describen en el documento WO 02/00336, Publicación de Solicitud de Patente de los Estados Unidos 02/0102578 o las referencias citadas anteriormente en el presente documento con respecto a los diferentes tipos de matrices.

Una localización de ensayo puede ser también el interior de una celda de flujo. Se puede colocar una matriz de sondas en la superficie interior de la celda de flujo e introducirse un fluido fluyendo en la celda. Una celda de flujo útil en la invención puede ser una celda de flujo con hueco capilar. Una celda de flujo con hueco capilar tiene una dimensión interior suficientemente estrecha y aberturas, de tal manera que un fluido se puede retener en la celda mediante la acción capilar y desplazarse posteriormente mediante la presión positiva ejercida en una abertura en un segundo fluido. Se puede proporcionar presión positiva, por ejemplo, mediante flujo por gravedad. Una celda de flujo capilar a modo de ejemplo que es útil en la invención es una formada entre la superficie de una matriz basada en porta tal como una matriz BeadChip (Illumina, Inc., San Diego CA) and una Coverplate (ThermoShandon, Inc., Pittsburgh, PA). Otra celda de flujo con hueco capilar útil es la que se utiliza en el sistema de flujo pistón GenePaint™ disponible de Tecan (Maennedorf, Suiza). De acuerdo con esto, la invención proporciona un método de modificación enzimática de ácidos nucleicos, tal como sondas unidas a sustratos, en una celda de flujo con hueco capilar. Los expertos en la materia reconocerán que se puede formar una celda con flujo capilar con cualquiera de una variedad de matrices conocidas en la técnica para conseguir capacidades de flujo fluido similares

Los sitios pueden ser un modelo tal como un diseño o configuración regular, o los sitios pueden estar en una distribución no modelizada. Una ventaja no limitante de un modelo regular de sitios es que los sitios pueden direccionarse en un plano de coordenadas X-Y. Un modelo en este sentido incluye una celda unitaria de repetición, tal como una que permite una alta densidad de perlas sobre un sustrato.

En una realización concreta, un sustrato de matriz puede ser un paquete o una matriz de fibra óptica, como se describe de forma general en el documento de los Estados Unidos con N° de Serie 08/944.850, la patente de los Estados Unidos N° 6.200.737; los documentos WO9840726, y WO9850782. Es también útil para la invención una matriz de fibra óptica unitaria preformada que tiene hebras de fibra óptica individuales discretas que se disponen y se unen coaxialmente a lo largo de sus longitudes. Una característica distintiva de una matriz de fibra óptica unitaria preformada en comparación con otros formatos de fibra óptica es que las fibras no son físicamente manipulables individualmente; esto es, no se puede separar generalmente una hebra en cualquier punto a lo largo de su longitud de otra hebra de fibra.

Los sitios de una matriz de la invención no necesitan ser sitios discretos. Por ejemplo, es posible usar una superficie uniforme de funcionalidades adhesivas o químicas, que permita, por ejemplo, la unión de partículas en cualquier posición. Estos es, la superficie de un sustrato de matriz se puede modificar para permitir la unión o la asociación de microesferas en sitios individuales, ya sean aquellos sitios contiguos o no con otros sitios. De esta manera, se puede modificar la superficie de un sustrato para formar sitios discretos de tal manera que solo una única perla se asocia con el sitio o, de forma alternativa, se puede modificar la superficie de tal manera que las perlas vayan a parar de forma aleatoria en sitios de población en cantidades diversas.

En una realización concreta, la superficie del sustrato se puede modificar para contener pocillos, o depresiones en la superficie del sustrato. Esto se puede llevar a cabo utilizando una variedad de técnicas, que incluyen, pero no se limitan a, fotolitografía, técnicas de estampación, técnicas de moldeo o técnicas de micrograbado. Como apreciarán los expertos en la materia, la técnica utilizada dependerá de la composición y de la forma del sustrato. Cuando el sustrato para una matriz de material compuesto es una placa de microvaloración, se puede utilizar una técnica de moldeo para formar pocillos de perlas en la parte inferior de los pocillos de ensayo.

En una realización concreta, se pueden llevar a cabo alteraciones físicas en una superficie de un sustrato para producir localizaciones de la matriz. Por ejemplo, cuando el sustrato es un paquete de fibras ópticas, la superficie del sustrato puede ser un extremo terminal del paquete de fibras, como se describe de forma general en las Patentes de los Estados Unidos con números 6.023.540 y 6.327.410. En esta realización se pueden preparar pocillos en un extremo terminal o distal de un paquete de fibras ópticas que tiene algunas fibras individuales. En esta realización, los núcleos de las fibras individuales se pueden grabar con respecto al revestimiento, de tal manera que se forman pocillos o depresiones pequeñas en un extremo de las fibras. Se puede alterar la profundidad de los pocillos utilizando diferentes condiciones de grabado para acomodar las partículas de un tamaño o forma concretos. Generalmente, en esta realización, las microesferas se asocian de forma no covalente en los pocillos, aunque los pocillos se pueden funcionalizar además químicamente para la unión de partículas covalentes. Tal como se muestra a continuación con detalle adicional, se pueden usar agentes de reticulación, o se puede usar una barrera física tal como una película o una membrana sobre las partículas.

En una realización concreta, la superficie de un sustrato puede estar modificada para contener sitios modificados químicamente que son útiles para unir, tanto de manera covalente como de manera no covalente, sondas o partículas que tienen sondas unidas. Los sitios químicamente modificados en este contexto incluyen, pero no se limitan a, la adición de un modelo de grupos funcionales químicos que incluye, por ejemplo, grupos amino, grupos carboxi, grupos oxo o grupos tiol. Se pueden usar dichos grupos para unir de forma covalente sondas o partículas que contienen grupos funcionales reactivos correspondientes. Otras modificaciones superficiales útiles incluyen, por ejemplo, la adición de un modelo de adhesivo que se puede usar para unir partículas; la adición de un modelo de grupos cargados para la unión electrostática de sondas o partículas; la adición de un modelo de grupos funcionales químicos que vuelve los sitios diferencialmente hidrófobos o hidrófilos de tal manera que la adición de sondas o partículas similarmente hidrófobas o hidrófilas en las condiciones adecuadas dará como resultado la asociación a los sitios sobre la base de la hidroafinidad.

Una vez que se generan las microesferas, se pueden añadir a un sustrato para formar una matriz. Se pueden preparar matrices, por ejemplo, añadiendo una disolución o suspensión de las perlas a un sustrato que contiene sitios de unión para las perlas. Una disolución portadora de las perlas puede ser un tampón de pH, un disolvente acuoso, un disolvente orgánico, o una mezcla. Tras la exposición de una suspensión de perlas a un sustrato, el disolvente se puede evaporar, y se puede eliminar el exceso de perlas. En realizaciones donde se usan métodos no covalentes para asociar las perlas a un sustrato matriz, las perlas se pueden cargar sobre el sustrato exponiendo el sustrato a una disolución de partículas, y a continuación aplicando energía, por ejemplo, agitando o haciendo vibrar la mezcla. Sin embargo, se puede usar también si se desea la carga estática. Los métodos para cargar perlas y otras partículas sobre sustratos matriz que se pueden usar en la invención se describen, por ejemplo, en la Patente de los Estados Unidos N° 6.355.431. La carga de las perlas se puede llevar a cabo antes de la modificación de las sondas en un método de detección que se muestra en el presente documento. De forma alternativa, la carga de las perlas se puede llevar a cabo tras la modificación de las sondas inmovilizadas con perlas que se hibridan con fragmentos del genoma en un método de la invención.

En algunas realizaciones, por ejemplo, cuando se lleva a cabo la unión química, las sondas o partículas con las sondas asociadas se pueden unir a un sustrato en un procedimiento ordenado no aleatorio. Por ejemplo, utilizando enlazadores de unión fotoactivables o adhesivos o máscaras fotoactivables, los sitios seleccionados sobre un

sustrato matriz se pueden activar secuencialmente para la unión, de tal manera que las poblaciones definidas de sondas o partículas se imponen a las posiciones definidas cuando se exponen al sustrato de matriz activado.

5 De forma alternativa, las sondas o partículas con sondas asociadas se pueden depositar de forma aleatoria sobre un sustrato y sus posiciones en la matriz determinarse mediante una etapa de decodificación. Esto se puede llevar a cabo antes, durante o después del uso de la matriz para detectar loci tipables utilizando métodos tales como aquellos que se muestran en el presente documento. En realizaciones en las que la colocación de la sonda es aleatoria, un sistema de codificación o de descodificación puede usarse para localizar y/o identificar las sondas en cada localización en la matriz. Esto se puede llevar a cabo en cualquiera de una variedad de maneras, tal como se describe, por ejemplo, en la Patente de los Estados Unidos N° 6.355.431.

15 En realizaciones en las que se usan partículas, se pueden incorporar firmas ópticas únicas en las partículas y se pueden usar para identificar la funcionalidad química o el ácido nucleico asociado con la partícula. Las firmas ópticas a modo de ejemplo incluyen, sin limitación, colorantes, usualmente cromóforos o fluoróforos, atrapados o unidos a perlas. Se pueden usar diferentes tipos de colorantes, diferentes relaciones de mezclas de colorantes, o diferentes concentraciones de colorantes, o una combinación de estas diferencias como firmas ópticas en la invención. Los ejemplos adicionales de partículas y de otros soportes que tienen firmas detectables que se pueden usar en la invención se describen en Cunin y col., Nature Materials 1: 39-41 (2002); Patentes de los Estados Unidos con números 6.023.540 o 6.327.410; o el documento WO9840726. De acuerdo con esta realización, la síntesis de los ácidos nucleicos se puede separar de su colocación en una matriz. De esta manera, se pueden sintetizar sondas de captura sobre perlas, y a continuación las perlas se pueden distribuir de forma aleatoria sobre una superficie modelizada. Debido a que las perlas se codifican en primer lugar con una firma óptica, esto significa que la matriz puede descodificarse con posterioridad. De esta manera, después que se prepara una matriz, se puede preparar una correlación de la localización de una localización de matriz individual en la matriz con su identidad de sonda. Esto significa que las localizaciones de la matriz se pueden distribuir de manera aleatoria sobre la matriz, un procedimiento rápido y barato en muchas aplicaciones de la invención en comparación tanto con la síntesis in situ como con las técnicas de moteado que se reseñan generalmente en los documentos de los Estados Unidos con números 98/05025, 99/14387, 08/818.199 o 09/151.877. Sin embargo, si se desea, las matrices preparadas mediante la síntesis in situ o las técnicas de moteado se pueden usar en la invención.

30 Debe señalarse que no todos los sitios de una matriz necesitan incluir una sonda o partícula. De esta manera, una matriz puede tener una o más localizaciones de matriz, sobre los sustratos que están vacíos. En algunas realizaciones, un sustrato matriz puede incluir uno o más sitios que contienen más de una perla o sonda.

35 Como apreciarán los expertos en la materia, una matriz aleatoria no necesita necesariamente descodificarse. En esta realización, las perlas o sondas se pueden unir a un sustrato matriz, y llevarse a cabo un ensayo de detección. Las localizaciones de matriz que tienen una señal positiva para la presencia de un híbrido de fragmentos de sonda con un locus tipable concreto pueden marcarse, o identificarse de otra manera para distinguirlas o separarlas de otras localizaciones de matriz. Por ejemplo, en aplicaciones en las que las perlas se marcan con un colorante fluorescente, las localizaciones de matriz para perlas positivas o negativas pueden marcarse mediante fotoblanqueamiento. Las marcas a modo de ejemplo adicionales incluyen, pero no se limitan a, precursores no fluorescentes que se convierten en la forma fluorescente mediante activación por luz o grupos fotorreticulables que pueden derivatizar una sonda o partícula con una marca o sustrato tras irradiación con la luz de una longitud de onda adecuada.

45 En una realización concreta, pueden desarrollarse algunos niveles de redundancia en una matriz usada en la invención. El desarrollo de la redundancia en una matriz puede proporcionar algunas ventajas no limitantes, que incluyen la capacidad de hacer estimaciones cuantitativas de la confianza acerca de los datos y sustanciales aumentos en la sensibilidad. Como apreciarán los expertos en la materia, existen al menos dos tipos de redundancia que pueden desarrollarse en una matriz: el uso de múltiples sondas idénticas o el uso de múltiples sondas dirigidas a la misma diana, pero que tienen diferentes funcionalidades químicas. Por ejemplo, para la detección de ácidos nucleicos, la redundancia del sensor utiliza una pluralidad de elementos sensores tales como perlas que tienen idénticos ligandos de unión tales como sondas. La redundancia de la diana utiliza elementos sensores con diferentes sondas para la misma diana: una sonda puede extender las primeras 25 bases de la diana, una segunda sonda puede extender las segundas 25 bases de la diana, etc. En el desarrollo de cualquiera o de ambos tipos de redundancia en una matriz, se puede llevar a cabo una variedad de análisis matemáticos estadísticos para el análisis de grandes conjuntos de datos. Otros métodos para la descodificación con elementos sensores redundantes y elementos diana que se pueden usar en la invención se describen, por ejemplo, en la Patente de los Estados Unidos N° 6.355.431.

60 Se pueden detectar loci tipables de híbridos sonda-fragmento en una matriz utilizando los métodos que se muestran anteriormente en el presente documento. En una realización concreta, se puede usar la redundancia de la sonda. En esta realización, una pluralidad de sondas que tienen secuencias idénticas está presente en una matriz. De esta manera, una pluralidad de subpoblaciones que tienen cada una, una pluralidad de perlas con sondas idénticas puede estar presentes en la matriz. Utilizando diversas sondas idénticas para una matriz dada, se puede combinar la señal óptica procedente de cada localización de matriz y analizarse utilizando métodos estadísticos. De esta

manera, la redundancia puede aumentar significativamente la confianza de los datos cuando se desea.

Como apreciarán los expertos en la materia, el número de sondas idénticas en una subpoblación variará con la aplicación y el uso de una matriz concreta. En general, se pueden usar cualquiera de entre 2 a miles de localizaciones de matriz idénticas, incluyendo, aproximadamente 5, 10, 20, 50 o 100 sondas o partículas idénticas.

Una vez obtenidas, las señales indicativas de los híbridos sonda-fragmento procedentes de una pluralidad de localizaciones de matriz se pueden manipular y analizarse de una variedad de formas, incluyendo el ajuste del valor inicial, el promediado, el análisis de la desviación estándar, el análisis de distribución y agrupamiento, el análisis del intervalo de confianza, el ensayo promedio, o similares. Se muestra a continuación la descripción adicional de las manipulaciones de los datos y en muchos casos se ilustra para los híbridos sonda-fragmento detectados sobre una matriz de perlas. Los expertos en la materia reconocerán que se pueden llevar a cabo manipulaciones similares para otras poblaciones de híbridos sonda-fragmento que incluyen, por ejemplo aquellas en las que otras localizaciones de matriz se tratan de forma similar a las perlas en los siguientes ejemplos.

Opcionalmente, una pluralidad de señales detectadas procedentes de una matriz o de otra mezcla de híbridos sonda-fragmento pueden ajustarse a los valores iniciales. En un procedimiento a modo de ejemplo, se pueden ajustar las señales ópticas para comenzar en un valor de 0,0 sustrayendo el entero 1,0 de todos los puntos de datos. Haciendo esto, se permite que los datos del bucle del valor inicial permanezcan a cero incluso cuando se suman juntos y se cancela la señal de ruido de la respuesta aleatoria. Cuando la muestra es un fluido, la región temporal del bucle de pulsos del fluido, sin embargo, presenta frecuentemente un cambio característico en la respuesta, tanto positiva, como negativa o neutra, antes del pulso de la muestra y requiere a menudo un ajuste del valor inicial para superar el ruido asociado con la deriva en los escasos primeros datos iniciales para cargar el desarrollo en la cámara CCD. Si no está presente la deriva, normalmente el valor inicial procedentes de los primeros puntos de datos para el mismo tipo de perla se puede sustraer de todos los datos de respuesta para el mismo tipo de perla. Si se observa deriva, el valor inicial promedio de los primeros diez puntos de datos para cada perla se pueden sustraer de todos los datos de respuesta para el mismo tipo de perla. Aplicando este ajuste del valor inicial, cuando se añaden múltiples respuestas de localización de matriz juntas, se pueden amplificar a la vez que el valor inicial permanece a cero. Debido a que todas las localizaciones de matriz responden al mismo tiempo a la muestra (por ejemplo, el pulso de la muestra), todos pueden ver el pulso en el mismo tiempo exacto y no existe registro o ajuste necesarios para superponer sus respuestas. Además, se pueden llevar a cabo otros tipos de ajuste del valor inicial que se conocen en la técnica, dependiendo de los requerimientos y de la salida del sistema utilizado.

Se puede desarrollar cualquiera de una variedad de posibles análisis estadísticos para generar parámetros estadísticos conocidos. Se conocen los análisis basados en la redundancia y se describen de forma general en textos tales como Freund y Walpole, *Mathematical Statistics*, Prentice Hall Inc., New Jersey (1980).

Si se desea, se puede llevar a cabo la suma de señales añadiendo los valores de intensidad de todas las respuestas en un punto temporal concreto. En una realización concreta, las señales se pueden sumar en varios puntos temporales, generando por tanto una respuesta temporal comprendida por la suma de respuestas de todas las perlas. Estos valores pueden ajustarse al valor inicial o ser brutos. La suma de señales se puede llevar a cabo en tiempo real o durante la reducción y el análisis de datos posterior a la adquisición de datos. En una realización, se puede llevar a cabo la suma de datos con un programa de hoja de cálculo comercial (Excel, Microsoft, Redmond, Wash.) Después que se recogen los datos de la respuesta óptica. Los métodos de suma de señales a modo de ejemplo adicionales que se pueden usar en la invención se describen en la Patente de los Estados Unidos Nº 6.355.431.

En una realización concreta, se pueden llevar a cabo análisis estadísticos para evaluar si unos puntos de datos concretos tienen validez estadística en una subpoblación utilizando técnicas que incluyen, pero no se limitan a, distribución o análisis de agrupamientos. Esto se puede llevar a cabo para descartar estadísticamente valores fuera de rango que pueden producir de otra manera un sesgo del resultado y aumentar la relación señal a ruido de cualquier experimento concreto. Los métodos útiles para determinar si los puntos de datos tienen validez estadística, se describen, por ejemplo, en la Patente de los Estados Unidos Nº 6.355.431 e incluyen, pero no se limitan a, el uso de intervalos de confianza, ensayo de promedios, o análisis de la distribución.

Una realización concreta utiliza una pluralidad de sondas de ácidos nucleicos que se dirigen a un único locus tipable pero que difieren en su secuencia real. Por ejemplo, un único fragmento de genoma diana que puede tener dos o más localizaciones de matriz teniendo cada una de ellas una sonda diferente. Esto puede añadir un nivel de confianza en aplicaciones en las que no se producen interacciones de unión no específicas con secuencias concretas. De acuerdo con esto, las sondas de ácidos nucleicos redundantes pueden tener secuencias que se superponen, son adyacentes o están separadas espacialmente.

Un método de la invención puede incluir además una etapa de poner en contacto una matriz de sondas de ácidos nucleicos con sondas de chaperona. Las sondas de chaperona son ácidos nucleicos que se hibridan a un fragmento de genoma diana en un sitio que está cerca del sitio de hibridación de una sonda usada para detectar o capturar el fragmento de genoma. Se pueden añadir sondas de chaperona antes o durante una etapa de captura o una etapa

de detección a fin de favorecer la hibridación de la detección de las sondas de captura evitando la asociación de las hebras complementarias de un fragmento de genoma de tal manera que la hebra del molde adecuada está disponible para hibridarse a las sondas de detección o captura.

- 5 Las sondas de chaperona pueden tener cualquiera de una variedad de longitudes o composiciones que incluyen, por ejemplo, las que se muestran anteriormente en el presente documento para otros ácidos nucleicos útiles en la invención. Una sonda de chaperona puede hibridarse a una secuencia diana inmediatamente adyacente hasta un sitio de hibridación de otra sonda o en un sitio que está separado del sitio de hibridación de la otra sonda. El hueco entre las sondas puede ser de 1 o más, 2 o más, 3 o más, 5 o más, 10 o más nucleótidos de longitud o más largo.
- 10 Se pueden proporcionar sondas de chaperona en cualquier concentración estequiométrica que se encuentra que se hibrida eficazmente de forma favorable con otra sonda incluyendo, por ejemplo, una relación de aproximadamente 100 moles, 10 moles, 5 moles, 2 moles, 1 mol, 0,5 moles, o 0,1 mol de sonda de chaperona por mol de fragmento de genoma diana.
- 15 Un método de la invención puede incluir además una etapa de amplificación de la señal en la que el número de marcas detectables unidas a un ácido nucleico aumenta. En una realización, una etapa de amplificación de la señal puede incluir proporcionar un ácido nucleico que está marcado con un ligando que tiene afinidad por un receptor concreto. Un primer receptor que tiene uno o más sitios capaces de unirse al ligando puede ponerse en contacto con el ácido nucleico marcado en condiciones en las que forma un complejo entre el receptor y el ácido nucleico del
- 20 ligando marcado. Además, el receptor puede ponerse en contacto con un reactivo de amplificación que tiene afinidad por el receptor. El reactivo de amplificación puede ser, por ejemplo, el ligando, o un agente mimético del ligando, o un segundo receptor que tiene afinidad por el primer receptor. El reactivo de amplificación puede, a la vez, marcarse con el ligando de tal manera que puede formar un complejo multimérico entre el receptor del ligando y el reactivo de amplificación. A continuación se puede detectar la presencia del complejo multimérico, por ejemplo,
- 25 detectando la presencia de una marca detectable sobre el receptor del reactivo de amplificación. Los componentes incluidos en una etapa de amplificación de la señal se pueden añadir en cualquier orden siempre que se forme un complejo detectable. Además se pueden usar otros restos de unión como parejas de moléculas de enlace tales como los que se muestran anteriormente en el presente documento para la amplificación de la señal.
- 30 Tal como se muestra en el esquema de amplificación de la señal a modo de ejemplo de la Figura 10, se puede llevar a cabo la amplificación de la señal utilizando un ácido nucleico marcado mediante estreptavidina-ficoeritrina (SAPE) y un anticuerpo dirigido contra SAPE biotinilado. En una realización, se puede emplear un protocolo de tres etapas en el que las sondas dispuestas en forma de matriz que se han modificado para incorporar biotina se incuban en primer lugar con estreptavidina-ficoeritrina (SAPE), seguido por incubación con un anticuerpo dirigido contra
- 35 estreptavidina biotinilado, y finalmente se incuban con SAPE de nuevo: Este procedimiento crea una amplificación en cascada de tipo sándwich debido a que la estreptavidina tiene múltiples sitios de unión y el anticuerpo tiene múltiples biotinas. Los expertos en la materia reconocerán entre las enseñanzas del presente documento que se pueden usar otros receptores tales como avidina, versiones modificadas de avidina, o anticuerpos, en un complejo de amplificación y que se pueden usar diferentes marcas tales como Cy3, Cy5 u otras que se muestran
- 40 anteriormente en el presente documento. Las técnicas de amplificación de la señal a modo de ejemplo adicionales y los componentes que se pueden usar en la invención se describen, por ejemplo, en la Patente de los Estados Unidos N° 6.203.989 B1.
- 45 Un método de la invención puede incluir además una etapa de eliminación de fragmentos del genoma a partir de híbridos sonda-fragmento seguida por la modificación de las sondas y previa a la detección de las sondas modificadas. Se pueden eliminar los fragmentos de genoma desnaturalizando los híbridos sonda-fragmento utilizando métodos conocidos en la técnica para interrumpir las interacciones de emparejamiento de bases tales como la exposición a una concentración baja de sal, disolventes orgánicos tales como formamida, calor u otros agentes desnaturalizantes. Los métodos a modo de ejemplo para desnaturalizar ácidos nucleicos híbridos que son
- 50 útiles en los métodos se describen en Sambrook y col., *más arriba* (2001) o en Ausubel y col., *más arriba*, (1998). Los fragmentos de genoma se pueden lavar tras la desnaturalización. De forma alternativa, los fragmentos de genoma pueden estar presentes en condiciones durante la detección.
- 55 Un método de la invención puede incluir adicionalmente una etapa de producir un informe que identifica que se detecta al menos un locus tipable. Un locus tipable detectado puede identificarse directamente, por ejemplo, por la secuencia, localización en un cromosoma o por un nombre reconocido del locus. De forma alternativa, el informe puede incluir los datos obtenidos a partir de un método de la invención en un formato que se puede analizar posteriormente para identificar uno o más loci detectados.
- 60 De esta manera, la invención proporciona además un informe de al menos un resultado obtenido mediante un método de la invención. Un informe de la invención puede estar en cualquiera de una variedad de formatos reconocibles que incluyen, por ejemplo, una transmisión electrónica, una memoria legible informáticamente, una salida para una interfaz de usuario gráfica, un disco compacto, un disco magnético o papel. Pueden usarse otros
- 65 formatos adecuados para la comunicación entre seres humanos, máquinas o ambos para un informe de la invención.

La invención proporciona además una matriz que incluye una población representativa inmovilizada en fase sólida de fragmentos de genoma: Se puede producir una población representativa de fragmentos de genoma e inmovilizarse utilizando métodos tales como los que se muestran en el presente documento anteriormente. Por ejemplo, se puede amplificar un genoma utilizando cebadores que tienen una marca secundaria tal como biotina o

5 grupos reticulantes reactivos y posteriormente inmovilizarse a través de la interacción con un receptor en fase sólida tal como avidina o un resto reactivo químico con el grupo reticulante. Una población representativa inmovilizada en fase sólida de fragmentos de genoma puede tener una o más de las características que se muestran anteriormente en el presente documento tales como una complejidad alta, baja o media.

10 Una población representativa inmovilizada en fase sólida de fragmentos de genoma se puede interrogar directamente utilizando los métodos de la invención. Generalmente, los ensayos y métodos de detección se han ilustrado anteriormente con respecto a las sondas inmovilizadas y fragmentos de genoma diana solubles. Los expertos en la materia reconocerán que en las realizaciones en el presente documento en las que se inmoviliza una población representativa de fragmentos de genoma, se pueden llevar a cabo de forma similar los métodos, sin

15 embargo, con los fragmentos de genoma sustituyendo las sondas en los ejemplos anteriores y las sondas tratadas como dianas en los ejemplos anteriores.

Emplear un ADN genómico diana en fase sólida puede proporcionar la ventaja de un elevado grado de multiplexado del ensayo permitiendo que cualquier cebador de detección mal hibridado o en exceso se lave antes de la posterior modificación enzimática de los cebadores, por ejemplo, en una técnica de extensión o ligadura. Las aplicaciones que se ven afectadas de forma adversa por la formación del dímero del cebador pueden mejorarse eliminando los dímeros de cebador antes de la detección. Un formato de ADN diana en fase sólida puede permitir también una rápida cinética de hibridación debido a que los cebadores se pueden hibridar a concentraciones relativamente altas, por ejemplo, mayores de aproximadamente 100 pM.

20 Los métodos que se muestran en el presente documento para amplificar el ADN genómico permiten que se amplifiquen cantidades relativamente pequeñas de ADN genómico en una gran cantidad. La inmovilización de grandes cantidades de ADN genómico en una fase sólida puede permitir que se interroguen loci tipables directamente, por ejemplo en un ensayo de extensión del cebador o basado en ligadura sin la necesidad de una posterior amplificación. La eliminación de la amplificación puede conducir a una genotipación más sólida y cuantitativa, que está a menudo disponible cuando se utiliza la detección basada en la preamplificación.

25 Otra ventaja de usar un ADN genómico diana en fase sólida es que se puede reutilizar. De esta manera, el genoma diana inmovilizado puede ser una muestra para archivo que se puede usar repetidamente con diferentes conjuntos de sondas de ácidos nucleicos. Además, en algunas aplicaciones, la contaminación arrastrada se puede reducir utilizando ADNg inmovilizado debido a que la amplificación se produce antes de la reacción de detección específica de SNP. Se entenderá que, las etapas descritas anteriormente para llevar a cabo los métodos de la invención se han definido en un orden concreto a modo de explicación. Los expertos en la materia reconocerán que las etapas se pueden llevar a cabo en cualquiera de una variedad de órdenes siempre que se consiga un resultado deseado. Por

30 ejemplo, los componentes de las reacciones que se muestran anteriormente se pueden añadir de forma simultánea, o secuencialmente, en cualquier orden que sea eficaz para producir uno o más de los resultados descritos. Además, las reacciones que se muestran en el presente documento pueden incluir una variedad de reactivos diferentes que incluyen, por ejemplo, sales, tampones, proteínas neutras, albúmina, detergentes, o similares. Se pueden añadir dichos reactivos para facilitar una hibridación y detección óptimas, reducir las interacciones de fondo no específicas, o para estabilizar los diferentes reactivos utilizados. También, los reactivos que mejoran de otra manera la eficacia de un método de la invención, tales como los inhibidores de las proteasas, inhibidores de las nucleasas, agentes antimicrobianos, o similares pueden utilizarse, dependiendo de los métodos de preparación de muestras y de la pureza de la diana. Los expertos en la materia conocerán o serán capaces de determinar los reactivos adecuados para conseguir dichos resultados.

35 Algunos de los métodos ilustrados en el presente documento con respecto a la detección de loci tipables de ADN genómico pueden aplicarse también al análisis de la expresión génica. En particular, los métodos para el marcado en la matriz de sondas de ácidos nucleicos que utilizan métodos de extensión del cebador que se pueden usar en la detección del ARN o del ADNc. Los híbridos de ADNc de la sonda se pueden detectar mediante los métodos de extensión del cebador basados en la polimerasa tal como se ha descrito en el presente documento anteriormente. De forma alternativa, para el ARN, hibridado de la matriz, se puede emplear la extensión del cebador basada en la transcriptasa inversa. Existen algunas ventajas no limitantes del marcado en la matriz para el análisis de la expresión génica. Los costes del marcado pueden disminuir drásticamente debido a que las cantidades de nucleótidos empleados son sustancialmente menores en comparación con los métodos para marcar dianas

40 capturadas. En segundo lugar, se puede reducir drásticamente la hibridación cruzada debido a que una diana debe hibridarse e incluir también la complementariedad perfecta en su extremo 3' para la incorporación de la marca en una reacción de extensión del cebador. De forma similar, se pueden usar ensayos OLA o GoldenGateTM para la detección de ADNc o ARNm hibridado. Los últimos dos métodos requieren normalmente la adición de un ácido nucleico exógeno para cada locus interrogado. Sin embargo, dichos métodos pueden ser ventajosos en aplicaciones en las que el uso de la extensión del cebador conduce a niveles inaceptables de extensión ectópica.

El marcado en la matriz anteriormente descrito con la extensión del cebador también se puede usar para controlar los sitios de corte y empalme alternativos diseñando el extremo 3' de la sonda para que coincida con una unión de corte y empalme de un ADNc o ARNm diana. El extremo se puede colocar para identificar de forma única todos los posibles sitios de corte y empalme aceptores relevantes para un gen concreto. Por ejemplo, las primeras 45 bases se pueden escoger para encontrarse completamente en el exón donante, y las últimas 5 bases en 3' se pueden encontrar en un conjunto de posibles exones aceptores de corte y empalme que se encuentran adyacentes al corte y empalme de las primeras 45 bases.

Se puede usar un ADNc o un ARNm en lugar de ADNg en un método descrito anteriormente en el presente documento para identificar loci tipables. Por ejemplo, se pueden usar un ADNc o un ARNm diana en un ensayo de genotipación. La genotipación del ADNc o del ARNm puede permitir diferencias de expresión específicas de alelos que se van a controlar, por ejemplo, a través de la "genotipación cuantitativa", o midiendo la proporción de un alelo frente al otro alelo en un marcador SNP bialélico. Las diferencias de expresión alélicas pueden ser el resultado, por ejemplo, de cambios en la velocidad de transcripción, el procesamiento del transcrito o la estabilidad del transcrito. Dicho efecto puede ser el resultado de un polimorfismo (o mutación) en una región reguladora, promotor, sitio de corte y empalme o región modificadora del sitio de corte y empalme u otras de dichas regiones. Además, los cambios epigenómicos en la cromatina tales como la metilación pueden contribuir también a las diferencias de expresión alélicas. De esta manera, se pueden usar los métodos para detectar dichos polimorfismos o mutaciones en productos expresados.

Se puede crear una representación "normalizada" a partir de un ADNc o ARNm diana mediante cualquiera de varios métodos tales como los basados en las colas de colocación de la PCR universal en una representación del ADNc (véase, por ejemplo, Brady, Yeast, 17: 211-7 (2000)) El procedimiento de normalización se puede usar para generar una representación del ADNc en la que cada locus tipable en la población está presente en relativamente el mismo número de copias. Esto puede ayudar en el procedimiento de genotipación cuantitativa de una muestra de ADNc debido a que las intensidades de la señal procedentes del ensayo de extensión del cebador basado en la matriz serán más uniformes que sin el procedimiento de normalización.

En una realización adicional, se puede usar un método de la invención para caracterizar una muestra de ARNm o de ADNc. Se puede usar una muestra de ARNm o de ADNc como una muestra diana en un método de la invención y detectarse un conjunto representativo de loci tipables. Se puede seleccionar el conjunto representativo de loci tipables para ser diagnóstico o las características de la muestra de ARNm o de ADNc. Por ejemplo, se pueden detectar los niveles de loci tipables concretos en una muestra y compararse con los niveles de referencia de estos loci, siendo indicativos los niveles de referencia de la extensión en la cual la muestra incluye las secuencias expresadas cubriendo los genes deseados. De esta manera, los métodos se pueden usar para determinar la calidad de una muestra de ARNm o de ADNc o su adecuabilidad para una aplicación concreta.

Una localización de matriz típica, tal como una perla, puede contener una población grande de sondas de ácidos nucleicos con un envasado relativamente denso. Tras la hibridación de los ácidos nucleicos diana en muchas condiciones, solo una porción de sondas en un ensayo de detección será ocupada con una diana complementaria. En dichas condiciones, es posible que las sondas con un envasado denso formarán estructuras intersonda que son susceptibles de la extensión del cebador ectópico. Además, tal como se muestra en la Figura 13A, las sondas que tienen secuencias autocomplementarias pueden tener también estructuras que son susceptibles a la extensión del cebador ectópico. La extensión ectópica se refiere a la modificación de una o ambas sondas en un híbrido inter o intrasonda durante una reacción de extensión. La expresión ectópica puede producirse sin tener en cuenta la presencia de una diana hibridada con la matriz.

De acuerdo con esto, la invención proporciona un método para inhibir la extensión ectópica de sondas un ensayo de extensión del cebador. El método incluye las etapas de (a) poner en contacto una pluralidad de sondas de ácidos nucleicos con una pluralidad de ácidos nucleicos diana en condiciones en las que se forman híbridos diana, (b) poner en contacto la pluralidad de sondas de ácidos nucleicos con un inhibidor de la extensión ectópica en condiciones en las que se forman híbridos inhibidores de la extensión ectópica de la sonda; y (c) modificar de forma selectiva las sondas en los híbridos diana de la sonda en comparación con las sondas en los híbridos inhibidores de la extensión ectópica de la sonda.

Un inhibidor de la extensión ectópica útil en la invención puede ser cualquier agente que sea capaz de unirse a una sonda de ácido nucleico monocatenario, evitando por tanto la hibridación de la sonda con una segunda sonda. Los agentes a modo de ejemplo incluyen, pero no se limitan a proteínas de unión a ácidos nucleicos monocatenarios (SSB), ácidos nucleicos tales como los que se muestran anteriormente que incluyen análogos de ácidos nucleicos, moléculas pequeñas. Dichos agentes tienen la propiedad general de unirse preferentemente con ácidos nucleicos monocatenarios sobre ácidos nucleicos bicatenarios, sin tener en cuenta la secuencia de nucleótidos. Las proteínas de unión a ácidos nucleicos monocatenarios a modo de ejemplo que se pueden usar en la invención incluyen, pero no se limitan a, Eco SSB, T4 gp32, T7 SSB, N4 SSB, Ad SSB, UP1, y similares y otras descritas, por ejemplo, en Chase y col, Ann. Rev. Biochem., 55: 103-36 (1986); Coleman y col, CRC Critical Reviews in Biochemistry, 7(3): 247-289 (1980) y en la Patente de los Estados Unidos N°. 5.773.257. La extensión ectópica en cualquiera de los ensayos de extensión del cebador que se muestran anteriormente se puede inhibir utilizando un método de la

invención. Las realizaciones a modo de ejemplo de los métodos para inhibir la extensión ectópica de las sondas en un ensayo de extensión del cebador se muestran en la Figura 13 y se describen con más detalle a continuación.

5 Tal como se muestra en la Figura 13B, la extensión ectópica puede minimizarse incubando una población de sondas con una proteína u otro agente que se une de forma selectiva a ácidos nucleicos monocatenarios, tal como SSB, T4 gen 32 o similar. El agente o la proteína se puede añadir en condiciones en las que reviste las sondas monocatenarias que no se han hibridado con un ácido nucleico diana evitando por tanto su autohibridación y posterior extensión. Un agente tal como una proteína que se une a sondas monocatenarias puede añadirse a una población de sondas antes de o durante una reacción de extensión del cebador, por ejemplo, antes de o durante una
10 etapa de hibridación.

Se puede reducir también la expresión ectópica utilizando uno o más oligonucleótidos de bloqueo (oligos). Tal como se muestra en la Fig. 13C, un oligo de bloqueo que es complementario del extremo 3' de una sonda se puede añadir en condiciones en las que se hibridará con sondas que no se han hibridado con un ácido nucleico diana. En aplicaciones en las que están presentes varias sondas, se pueden añadir una pluralidad de oligonucleótidos de bloqueo diseñados para hibridarse a los extremos 3' de las sondas. Se pueden añadir uno o más oligos de bloqueo a una población de sondas antes o durante una reacción de extensión del cebador, por ejemplo, antes o durante una
15 etapa de hibridación.

20 Tal como se muestra en la Figura 13D, se puede diseñar una sonda con porciones de secuencias complementarias capaz de formar una estructura de horquilla que no sea capaz de extenderse en las condiciones utilizadas para la etapa de extensión del cebador en un ensayo de extensión del cebador. En el ejemplo que se muestra en la Figura 13D, el extremo 3' de la sonda se hibrida con el extremo 5' de la sonda, y debido a que el extremo 5' no es adyacente a un molde legible la horquilla no se puede extender ectópicamente. Se puede diseñar una sonda para tener una región de la primera secuencia adyacente al extremo 3' de la sonda que es complementaria de una región de la segunda secuencia de la sonda de tal manera que una horquilla forma un saliente 3' que no es capaz de extenderse. Se diseña adicionalmente la estructura de horquilla de tal manera que no inhiba la hibridación con los ácidos nucleicos diana en las condiciones de la etapa de hibridación de una reacción de extensión del cebador. Por ejemplo, dos regiones de una sonda pueden tener secuencias complementarias que no se hibridan sustancialmente a las temperaturas usadas durante la hibridación diana, pero que llegan a hibridarse para formar una horquilla una vez que se reduce la temperatura durante la extensión.
25
30

Aunque se han ilustrado anteriormente los métodos para reducir la extensión ectópica con respecto a las sondas dispuestas en forma de matriz, los expertos en la materia reconocerán que los métodos se pueden aplicar de forma similar a las reacciones de extensión en otros formatos tales como reacciones en fase de disolución o perlas separadas espacialmente en una fase fluida.
35

En algunas condiciones de los ensayos de extensión, las polimerasas pueden insertar nucleótidos adicionales al final del extremo 3' de una sonda monocatenaria a la que le falta el ácido nucleico del molde hibridado. Se sabe también que dicha actividad se produce en los extremos 3' de los finales enromados de los ácidos nucleicos bicatenarios en algunas condiciones y esto se denomina como una actividad extendasa terminal (véase, por ejemplo, Hu y col., DNA and Cell Biology, 12: 763-770 (1993). De acuerdo con esto, se puede llevar a cabo una reacción de extensión usada en la invención en condiciones que inhiben la actividad extendasa terminal. Por ejemplo, se puede seleccionar una polimerasa que tenga niveles suficientemente bajos de actividad extendasa terminal en las condiciones de reacción de extensión que se van a usar o los nucleótidos que se incorporan de forma preferente mediante la actividad extendasa de una polimerasa concreta se pueden excluir de una reacción de extensión, o las sondas sin hibridar se pueden bloquear o eliminarse de una reacción de extensión.
40
45

La detección mediante hibridación directa de los ácidos nucleicos dianas puede adolecer de una disminución de la especificidad del ensayo debida a las reacciones de hibridación cruzada bajo algunas condiciones de ensayo. La detección enzimática basada en matriz de los ácidos nucleicos dianas ofrece una potente solución para aumentar la especificidad. Además del campo de la genotipación anteriormente discutido, la invención se puede aplicar a una especificidad creciente en el número de copias del ADN, agentes microbianos, expresión génica, y así sucesivamente. Esto llega a ser particularmente relevante a medida que la complejidad de la muestra de ácido nucleico aumenta hasta el nivel de una complejidad genómica humana. Por ejemplo, los experimentos con el número de copias del ADN, en los que el ADN genómico marcado se hibrida con matrices de ADN están comprometidos a menudo por problemas de especificidad. Empleando la hibridación directa en combinación con una etapa enzimática basada en matriz tal como en la extensión del cebador, u otras que se muestran anteriormente en el presente documento, se puede mejorar drásticamente la especificidad, esto es debido a que las dianas de la hibridación cruzada no se detectarán ya que el marcado por la etapa de detección enzimática se produce debido a que la complementariedad 3' es perfecta.
50
55
60

De acuerdo con otra realización de la presente invención, se proporcionan sistemas diagnósticos para llevar a cabo uno o más de los métodos descritos anteriormente en el presente documento. Se puede proporcionar un sistema diagnóstico de la invención en forma de kit, incluyendo, si se desea, un material de envasado adecuado. En una realización, por ejemplo, un sistema diagnóstico puede incluir una pluralidad de sondas de ácido nucleico, por
65

ejemplo, en un formato de matriz, y uno o más reactivos útiles para detectar un fragmento de ADN u otro ácido nucleico diana hibridado con una sonda de la matriz. De acuerdo con esto, cualquier combinación de reactivos o componentes que sea útil en un método de la invención, tales como las que se muestran en el presente documento anteriormente con respecto a métodos concretos, puede incluirse en un kit proporcionado por la invención. Por ejemplo, un kit puede incluir una o más sondas de ácido nucleico unidas a una matriz y que tienen extremos 3' libres junto con otros reactivos útiles para la reacción de detección de la extensión del cebador.

Tal como se usa en el presente documento, la frase "material de envasado" se refiere a una o más estructuras físicas usadas para alojar los contenidos del kit, tales como las sondas de ácido nucleico o los cebadores. El material de envasado puede fabricarse mediante métodos bien conocidos, preferiblemente para proporcionar un entorno estéril, exento de contaminantes. Los materiales de envasado empleados en el presente documento pueden incluir, por ejemplo, aquellos habitualmente utilizados en los sistemas de diagnóstico basados en ácidos nucleicos. Los materiales de envasado a modo de ejemplo incluyen, sin limitación, vidrio, plástico, papel, láminas, y similares, capaces de mantener en unos límites fijos un componente útil en los métodos de la invención tal como un ácido nucleico, oligonucleótido, o cebador aislado.

El material de envasado puede incluir una etiqueta que indica que los ácidos nucleicos de la invención se pueden utilizar para un método concreto. Por ejemplo, una etiqueta puede indicar que el kit es útil para detectar un conjunto concreto de loci tipables, determinando por tanto un genotipo del individuo. En otro ejemplo, una etiqueta puede indicar que el kit es útil para amplificar una muestra de ADN genómico particular.

Las instrucciones para el uso de los reactivos o componentes empaquetados se incluyen también normalmente en un kit de la invención. "Instrucciones de uso" incluye normalmente una expresión tangible que describe la concentración de reactivo o de componente o al menos un parámetro del método de ensayo, tal como las cantidades relativas de componentes del kit y la muestra que se van a premezclar, periodos de tiempo de mantenimiento de las premezclas del reactivo / muestra, temperatura, condiciones de tampón, y similares

Un método de la invención puede incluir controles para determinar el resultado deseable o indeseable de uno o más de los reactivos, componentes o etapas descritas en el presente documento. Se puede llevar a cabo la comparación de resultados de una muestra que está siendo investigada con los resultados de los controles en un método de la invención, validando por tanto los resultados, identificando las etapas que soportan la repetición o influyen la interpretación de los resultados. Si los resultados de uno o más controles quedan fuera de un intervalo deseado de resultados, un método de la invención puede incluir una etapa de modificación de un valor u otros puntos de datos obtenidos de una muestra que se está investigando. Un método de la invención puede incluir determinar los resultados de uno o más de los controles que se muestran a continuación y si los resultados quedan fuera de un intervalo deseado repetir entonces una o más de las etapas del método. De esta manera, la detección de una señal procedente de un control y la modificación de las condiciones se pueden llevar a cabo de una manera alternativa hasta que se obtiene un conjunto deseado de condiciones.

Se pueden usar controles de la amplificación en un método de la invención tal como un método que incluye una etapa de amplificar representativamente un genoma y/o de producir fragmentos de genoma. Un control de la amplificación a modo de ejemplo es un enriquecimiento de genoma extrínseco. Por ejemplo, una pequeña cantidad de ADN genómico microbiano se puede enriquecer en una reacción para la amplificación con cebadores aleatorios de un genoma humano. La cantidad de ADN genómico añadida es normalmente suficiente para competir con la potencial contaminación de otras muestras de ADN pero insuficiente para competir sustancialmente con la amplificación de la muestra de ADN genómico humano. La detección de loci que son únicos en el genoma microbiano en comparación con el genoma humano utilizando, por ejemplo, un subconjunto de sondas que se hibridan de forma selectiva con los loci microbianos y no con los loci humanos, se puede usar para determinar si una amplificación fallida es debida a los componentes de una reacción de RPA defectuosos a un ADN genómico humano de mala calidad. De forma más específica, los niveles detectables de loci microbianos resultantes de la reacción de RPA indican que el ADN genómico humano es de mala calidad y que los componentes de la reacción de RPA son funcionales y, en contraste, la ausencia de niveles detectables de loci microbianos indica un fallo de los componentes de la reacción.

Se pueden usar controles de la hibridación en un método de la invención como un método que incluye una etapa de poner en contacto fragmentos del genoma con sondas de ácido nucleico. Normalmente, los controles de la hibridación son ácidos nucleicos sintéticos que se incuban simultáneamente con ácidos nucleicos diana durante una etapa de hibridación de la sonda. Un ejemplo de un control de la hibridación útil es un conjunto de control de sondas de restricción que tienen secuencias que forman una serie de emparejamientos incorrectos con respecto a la secuencia de una diana de control de la restricción. La serie de sondas puede incluir una primera sonda que tiene una secuencia que está emparejada de manera perfecta con la secuencia de la diana control de la restricción, una segunda sonda que tiene un emparejamiento incorrecto con la secuencia de la diana control de la restricción, una tercera sonda que tiene el mismo emparejamiento incorrecto que la segunda sonda y un segundo emparejamiento incorrecto, una cuarta sonda que tiene los mismo dos emparejamientos incorrectos que la tercera sonda y un tercer emparejamiento incorrecto, etc. Es posible tener dos o más emparejamientos incorrectos por sonda en esta serie. Los emparejamientos incorrectos en la serie pueden ser adyacentes entre sí o separarse entre sí de tal manera que

uno o más nucleótidos emparejados de manera incorrecta intervienen en la secuencia. Los emparejamientos incorrectos en la serie pueden estar localizados cerca del extremo 5' de la sonda de tal manera que todas las sondas tienen un extremo 3' que está emparejado de manera perfecta con la diana control de la restricción.

- 5 El número y/o la identidad de las sondas de control de la restricción que se hibridan con la diana control de la restricción pueden estar correlacionados con el rigor de las condiciones de hibridación. En los niveles de rigor más altos, solo la primera sonda de control de la restricción de la serie anterior (la sonda de control emparejada de manera perfecta) se hibridará con la sonda de control diana. Condiciones de rigor más bajas darán como resultado más sondas de control de la restricción en la serie que se hibriden con la diana control de la restricción. De esta manera, se pueden usar las sondas de control de la restricción para identificar las condiciones que proporcionen una restricción deseada para la hibridación de los fragmentos del genoma y las sondas.

- 15 Un control adicional que se puede usar es un control de la concentración. Una diana de control de la concentración que tiene una secuencia que está emparejada de manera perfecta con la secuencia de una sonda de control de la concentración que se puede usar. Se pueden proporcionar dianas de control de la concentración a diferentes concentraciones para controlar las sondas. El límite inferior de detección para un conjunto concreto de condiciones de ensayo se puede cuantificar determinando la concentración más baja de la diana detectada. Si se desea, la diana control de la concentración puede tener uno o más emparejamientos incorrectos para la sonda de control, por ejemplo, en el extremo 3' de la secuencia diana. De acuerdo con esto, se puede realizar la evaluación de la restricción o la especificidad con las sondas de concentración.

- 25 Se pueden usar controles de modificación de la sonda en un método de la invención tal como un método que incluye una etapa de modificación de una sonda a la vez que se hibrida con un fragmento de genoma. Los ejemplos de controles de modificación de la sonda son controles de extensión que indican los niveles de extensión de la sonda mediante la polimerasa en un método de la invención. Un control de la extensión a modo de ejemplo es una sonda en horquilla o un conjunto de sondas en horquilla emparejadas de manera correcta e incorrecta. El conjunto puede incluir dos o más de las 16 posibles combinaciones de emparejamientos correctos y emparejamientos incorrectos que surgen de 4 nucleótidos (por ejemplo, un emparejamiento GC y al menos uno de GA, GT y GG). Las sondas en horquilla se unen normalmente a un sustrato en sus extremos 5' y tienen una secuencia palindrómica de tal manera que pueden formar una estructura en horquilla en sus extremos 3' en condiciones de restricción permisibles. La sonda emparejada correctamente tendrá una horquilla que termina en un emparejamiento de bases en 3', mientras que la sonda emparejada incorrectamente terminará con un emparejamiento incorrecto en 3'. La modificación de la sonda en horquilla emparejada correctamente indica que los componentes del ensayo de extensión son funcionales en las condiciones que se están empleando. Una ventaja de usar sondas de control en horquilla es que la indicación es independiente de la presencia de ácidos nucleicos diana. De esta manera, para una reacción de extensión fallida, los resultados para el control en horquilla con emparejamiento correcto se pueden usar para determinar si los problemas proceden de la muestra de ácido nucleico diana o del resto de reactivos de la reacción de extensión. La modificación de la sonda en horquilla emparejada incorrectamente se puede controlar para determinar si los reactivos de la reacción de extensión son sondas modificadoras de una manera independiente al molde. Aunque las sondas de control en horquilla se han ilustrado anteriormente con respecto a las reacciones de extensión, los expertos en la materia reconocerán que se pueden usar en otras reacciones de modificación dependientes del molde tales como una reacción de ligadura.

- 45 Otro control de modificación de la sonda útil es un control de la eficacia de la extensión. Un control de la eficacia de extensión puede incluir un conjunto de sondas de control de la eficacia de la extensión que son complementarias de las secuencias superpuestas de una diana control de la eficacia de la extensión tal como los extremos 3' de las sondas que complementan un nucleótido A, C, T o G, respectivamente, de una secuencia de 4 nucleótidos. De esta manera, la alineación de la secuencia de una diana control de la eficacia de la extensión con cuatro de dichas sondas de control de la eficacia de la extensión aparece como un conjunto escalonado de secuencias desplazadas en sus extremos 3' en un nucleótido. Un control de la eficacia de la extensión puede ser útil para determinar si las condiciones de la reacción de extensión están equilibradas o no con respecto a la incorporación de todos los nucleótidos que se están usando o si uno o más nucleótidos se están incorporando de forma selectiva.

- 55 Un método de la invención puede incluir además la evaluación de controles no polimórficos. Un control no polimórfico es un conjunto de sondas con emparejamiento perfecto y con emparejamiento incorrecto para una secuencia no polimórfica en un genoma. Las sondas con emparejamiento perfecto y con emparejamiento incorrecto son complementarias de la misma región del genoma con la excepción que sus extremos 3' son tanto complementarios como no complementarios, respectivamente, de la región de la secuencia del genoma. Se pueden usar uno o más conjuntos, por ejemplo, que tengan diferentes contenidos de GC para controlar el rigor, y/o que tengan una o más de todas las posibles combinaciones de emparejamientos correctos y emparejamientos incorrectos. Las sondas polimórficas pueden facilitar la optimización del ensayo utilizando muestras individuales únicas o mixtas cuando se comparan para agrupar datos con múltiples individuos.

- 65 Se pueden usar controles de tiras en un método de la invención tal como un método que incluye una etapa de eliminación de fragmentos de genoma procedentes de una pluralidad de sondas. Por ejemplo, una diana control de una tira marcada puede enriquecer una muestra de un fragmento de genoma antes de la hibridación con una

pluralidad de sondas de tal manera que uno de los híbridos se ha tratado para eliminar los fragmentos de genoma. Se puede detectar la presencia o ausencia de la diana marcada y correlacionarse con la eliminación insatisfactoria o satisfactoria del fragmento, respectivamente. En realizaciones concretas, se puede incorporar una marca a una diana control de una tira a la vez que se hibrida con una sonda complementaria. Por ejemplo, el extremo 3' de la diana control de la tira se puede hibridar con la sonda de tal manera que la diana se puede modificar de una manera dependiente del molde. Normalmente, la diana control de la tira y su sonda complementaria se diseñan de tal manera que la sonda no se modifica en la misma etapa que la diana. Por ejemplo, la sonda puede tener un análogo de nucleótido 3' que no es susceptible de modificación y/o el extremo 3' de la sonda puede formar un emparejamiento incorrecto con la diana. Además, la sonda que complementa la diana control de la tira puede diseñarse para que tenga una secuencia que no complementa ninguno de los fragmentos de genoma que se va a detectar en un método de la invención.

Se pueden usar controles de la detección en un método de la invención tales como un método que incluye una etapa de detección de los loci tipables de los híbridos sonda-fragmento. Por ejemplo, se puede usar un conjunto de sondas de control de la marca que tengan cantidades conocidas de una marca asociada. Las sondas de control de la marca se pueden analizar como una curva de valoración para determinar la sensibilidad o el intervalo de detección de la marca utilizada para detectar loci tipables de fragmentos de genoma. No es necesario que las sondas de control de la marca sean el mismo tipo de molécula que las sondas utilizadas para la detección de fragmentos del genoma. De acuerdo con esto, las sondas control de la marca pueden ser marcas unidas de manera directa o indirecta a una localización concreta sobre una superficie de la matriz. En el caso de una detección basada en biotina en la matriz, las zonas de control de la marca pueden ser localizaciones en la matriz que tienen cantidades conocidas de biotina unida de manera covalente.

Aunque la invención se ilustra en el presente documento con respecto a una matriz de sondas inmovilizadas, los expertos en la materia reconocerán que también se pueden emplear otros formatos de detección. Por ejemplo, los métodos que se muestran en el presente documento se pueden llevar a cabo en una fase disolución en lugar de en una fase sólida. De acuerdo con esto, las sondas en fase disolución pueden sustituir las sondas inmovilizadas en los métodos definidos anteriormente. Las sondas en fase disolución se pueden detectar de acuerdo con propiedades tales como las definidas anteriormente con respecto a las marcas de detección o a los restos de detección. Por ejemplo, las sondas pueden tener carga, masa, relación de carga a masa u otras propiedades identificables distintivas. Dichas propiedades distintivas pueden detectarse, por ejemplo, en un sistema de cromatografía tal como electroforesis capilar, gel de acrilamida, gel de agarosa o similares, o en un sistema espectroscópico tal como espectroscopía de masas. De esta manera, la divulgación proporciona además un método para detectar loci tipables de un genoma que incluye las etapas de (a) proporcionar una población representativa amplificada de fragmentos de genoma que tienen los loci tipables; (b) poner en contacto los fragmentos de genoma con una pluralidad de sondas de ácido nucleico que tienen secuencias que corresponden a los loci tipables en condiciones en las que se forman híbridos sonda-fragmento; (c) modificar los híbridos sonda-fragmento; y (d) detectar una sonda o fragmento que se ha modificado, detectando de este modo los loci tipables del genoma.

EJEMPLO I

Amplificación del genoma completo mediante el uso de la amplificación con cebadores aleatorios (Random-Primed Amplification, RPA).

Este ejemplo demuestra la producción de una población representativa amplificada de fragmentos de genoma procedentes del genoma de una levadura.

El ADN genómico de una levadura, procedente de la cepa S228C de *S. Cerevisiae*, se preparó mediante el uso de un kit de extracción Qiagen Genomic DNA y 10 ng del ADN genómico se amplificaron con la polimerasa Klenow.

Se evaluaron varios parámetros para determinar su efecto sobre el rendimiento de la reacción de amplificación con cebadores aleatorios (exo) Klenow. Las reacciones de amplificación se llevaron a cabo en condiciones similares, con la excepción de que un parámetro se modificó sistemáticamente. La Figura 3 muestra los resultados comparando las reacciones de amplificación llevadas a cabo a diferentes concentraciones de los trifosfatos de desoxinucleótido.

Después de cada reacción, el ADN amplificado se purificó en placas de ultrafiltración Montage (Millipore), cargadas en un gel de agarosa, y el ADN se cuantificó por lectura a UV₂₆₀ tal como se muestra en la Figura 3A. El rendimiento de la amplificación se determinó en función de la intensidad de la mancha en cada hilera, y los resultados se muestran en la tabla 3 (B). Como se muestra en las dos últimas columnas de la Figura 3B, se amplificaron 10 ng de molde de genoma de levadura hasta cantidades comprendidas en el intervalo de aproximadamente 6 a 80 microgramos, lo que representa aproximadamente una amplificación de 600 a 8000 veces. El tamaño promedio del fragmento en las condiciones ensayadas fue de aproximadamente 200-300 pb.

Los resultados demuestran que los rendimientos de amplificación aumentaron a mayores concentraciones del cebador o de los trifosfatos de desoxinucleótido. De esta forma, los parámetros de reacción se pueden modificar y evaluar sistemáticamente para determinar los rendimientos de amplificación deseados.

EJEMPLO II**Detección de loci de levaduras en una muestra de genoma completo de levadura hibridado con BeadArrays™**

5 Este ejemplo demuestra la detección reproducible de loci de levaduras en una muestra de genoma completo de levadura hibridado con BeadArrays™ y sondeados con un cebador de extensión específico del alelo (ASPE).

10 Seiscientos nanogramos de un ADN_g de levadura amplificado con cebadores aleatorios (RPA) se hibridó con BeadArray™ (Illumina) específico del locus. El BeadArray™ estaba compuesto por 96 pares de sonda de oligonucleótido (PM y MM, 50 bases de longitud) que interrogan diferentes loci de base génica distribuidos en la totalidad del genoma de *S. cerevisiae*. El ADN genómico amplificado de la levadura se hibridó con el BeadArray™ en las siguientes condiciones: hibridación durante la noche a 48 °C en tampón de hibridación normalizado 1X (NaCl 1 M, tampón de fosfato de potasio 100 mM (pH 7,5), Tween 20 al 0,1%, formamida al 20%). Tras la hibridación, las matrices se lavaron con tampón de hibridación 1X a 48 °C durante 5 min., seguido por un lavado con tampón de hibridación 0,1 X a temperatura ambiente durante 5 min. Finalmente, la matriz se lavó durante 5 min. con tampón de reacción ASPE para bloquear y equilibrar la matriz antes de la etapa de extensión. Tampón de reacción ASPE (tampón de extensión 10 X GG (Illumina, Inc., San Diego, CA), Tween 20 al 0,1%, BSA 100 ug/ml, y ditiotretiol 1 mM, sacarosa al 10%, betaína 500 mM).

20 Se llevó a cabo una reacción ASPE directamente en la matriz de la siguiente forma. Los BeadArrays se sumergieron en 50 uls de una mezcla de reacción ASPE que contenía el tampón de reacción ASPE descrito suplementado con dNTP 3 uM (dCTPn 1,5 uM), biotina-11-dCTP 1,5 uM, ~0,4 ul de KlenTaq (DNA Polymerase Technology, Inc, St. Louis, MO, 63104). Los BeadArrays™ se incubaron en la reacción ASPE durante 15 min. a temperatura ambiente.

25 Los BeadArrays™ se lavaron con NaOH 0,2 N reciente durante 2 min., y a continuación 2 veces en tampón de hibridación 1X durante 30 s. La marca de biotina incorporada se detectó mediante un ensayo de tipo sándwich que empleaba estreptavidina-ficoeritrina y tinción con anticuerpo biotinilado dirigido contra estreptavidina. Esto se llevó a cabo de la siguiente forma. Los BeadArrays™ se bloquearon a temperatura ambiente durante 30 min en bloqueo de caseína (Pierce, Rockford, IL). Esto fue seguido de un lavado rápido (1 min.) en tampón de hibridación 1X, antes de tinción durante 5 min. a temperatura ambiente con una disolución de estreptavidina-ficoeritrina (SAPE) (tampón de hibridación 1X, Tween 20 al 0,1%, BSA 1 mg/ml, 3 ug/ml de estreptavidina-ficoeritrina (Molecular Probes, Eugene, OR). Tras la tinción, los BeadArrays™ se lavaron rápidamente con tampón de hibridación 1X antes de realizar la contratinción con 10 ug/ml de anticuerpo biotinilado dirigido contra estreptavidina (Vector Labs, Burlingame, CA) en TBS 1X suplementado con 6 mg/ml de suero de cabra, caseína y Tween 20 al 0,1%. Esta etapa fue seguida por un lavado rápido con tampón de hibridación 1X y una segunda tinción con disolución SAPE tal como se ha descrito.

35 Tras la tinción, se llevó a cabo un lavado final con tampón de hibridación 1X.

40 El panel izquierdo de la Figura 4 muestra la imagen de una matriz tras la hibridación con la muestra de genoma completo de levadura amplificado y detección con ASPE. El diagrama del panel derecho de la Figura 4 muestra un subconjunto de intensidades de emparejamiento perfecto (PM) y emparejamiento incorrecto (MM) (48 loci de 96). Más del 88% de los loci tuvieron cocientes PM/MM superiores a 5, lo que indica la capacidad para distinguir la mayoría de los loci de los genotipos alternativos.

45 La capacidad para distinguir loci tipables en los genomas de mayor complejidad que la levadura se evaluó enriqueciendo con ADN genómico de levadura el fondo genómico de un organismo más complejo. Seiscientos nanogramos de ADN genómico de levadura (complejidad de 12 Mb) enriquecieron 150 ug de ADN genómico humano (complejidad de 3000 Mb) para imitar la presencia de loci de una sola copia en un genoma de complejidad equivalente a la del ser humano. La hibridación de esta muestra enriquecida con la matriz mostró muy pocas diferencias con el ADN de la levadura hibridado en solitario, e indicando la capacidad de la matriz para capturar específicamente las secuencias diana correctas en un fondo genómico complejo.

50 Estos resultados demuestran la detección de varios loci tipables en un genoma de levadura tras hibridación de una muestra de genoma completo con una matriz. Estos resultados demuestran además que no es necesario que la amplificación detecte una pluralidad de loci tipables en una muestra de genoma completo. Además, los resultados son reproducibles, lo que demuestra la solidez del método.

EJEMPLO III**Genotipado del genoma completo (WGG) de un ADN_g humano hibridado directamente con BeadArrays™.**

60 Este ejemplo demuestra la hibridación de una población representativa de fragmentos de genoma con una matriz, y la detección directa de varios loci tipables de los fragmentos de genoma hibridados. Este ejemplo demuestra adicionalmente la detección de loci tipables en una matriz usando cualquiera de dos ensayos de extensión de cebador diferentes.

65

Detección basada en SBE

Muestras de ADN genómico de placenta humana se obtuvieron del Coriell Inst. Camden, NJ. La muestra de ADN genómico placentario humano (150 ug) se hibridó con un BeadArray™ (Illumina) que tiene 4 paquetes independientes, conteniendo cada uno el mismo conjunto de 24 sondas no polimórficas diferentes (50-mers). El BeadArray™ estaba compuesto por 96 sondas dirigidas a loci no polimórficos humanos distribuidos aleatoriamente por la totalidad del genoma humano. Las sondas tenían 50 bases de longitud con un contenido de ~ 50% GC, y estaban diseñadas para resecuenciar bases adyacentes A (16 sondas), C (16 sondas), G (16 sondas), o T (16 sondas). Las muestras de ADN (150 ug de ADN de placenta humana) se hibridaron durante la noche a 48 °C en tampón de hibridación 1X estándar (NaCl 1 M, tampón de fosfato de potasio 100 mM (pH 7,5), Tween 20 al 0,1%, formamida al 20%) en un volumen de 15 ul.

Se llevaron a cabo cuatro reacciones SBE independientes directamente en la matriz, una en cada paquete independiente, de la siguiente forma: la reacción "A" contenía ddATP marcado con biotina y ddCTP, ddGTP, y ddTTP sin marcar. Las otras tres reacciones SBE, fueron similares, excepto en que las designaciones marcadas y sin marcar se ajustaron de forma adecuada. Las condiciones de la reacción SBE fueron las siguientes: los BeadArrays™ se sumergieron en una mezcla de reacción SBE a 50°C durante 1 min. Se proporcionaron cuatro reacciones SBE diferentes, una mezcla de resecuenciación A, C, G, o T. Por ejemplo, una mezcla de resecuenciación con 50 ul de A-SBE contenía biotina-11-ddATP 1 uM (Perkin Elmer), ddCTP 1 uM, ddGTP 1 uM, y ddUTP 1 uM, tampón Thermosequenase 1X, 0,3 U de Thermosequenase, 10 ug/ml de BSA, DTT 1 mM, y Tween 20 al 0,1 %. Las otras tres mezclas SBE fueron similares con la base adecuada marcada incluida y el resto de las bases sin marcar.

Los resultados de las reacciones SBE se muestran en la Figura 5. En la Figura 5, el conjunto de 96 sondas se dividió en cuatro grupos correspondientes a la cuatro reacciones diferentes, designadas de CA1 a CA24 para la reacción de ddATP marcado con biotina, CC1 a CC24 para la reacción de ddCTP marcado con biotina, CG1 a CG24 para la reacción de ddGTP marcado con biotina, y CT1 a CT24 para la reacción de ddTTP marcado con biotina. Tal como se muestra en la Figura 5, la mayoría de las sondas mostraron una excelente discriminación de la señal.

Detección basada en ASPE

Una muestra de ADN genómico de placenta humana preparada de manera similar (150 ug) se hibridó con un BeadArray™ que contenía 77 pares de sondas de emparejamiento perfecto (PM) y emparejamiento incorrecto (MM) que interrogaban loci no polimórficos. Las sondas ASPE estaban diseñadas para dirigirse a los sitios no polimórficos incluidos en el genoma humano. Las sondas tenían una 50 bases de longitud con un contenido de ~ 50% GC. Las sondas de emparejamiento perfecto (PM) estaban completamente emparejadas con secuencias genómicas mientras que las sondas de emparejamiento incorrecto (MM) contenían una única base emparejada de manera incorrecta con respecto a la secuencia genómica en la base 3'. El tipo de emparejamiento incorrecto estaba sesgado hacia la modelización de los polimorfismos A/G y C/T. Las condiciones de hibridación y reacción fueron análogas a las anteriormente descritas en el Ejemplo II.

Se llevó a cabo una reacción de extensión específica del alelo (ASPE) directamente en la superficie de la matriz, y la marca de biotina incorporada se detectó mediante tinción con estreptavidina-ficoeritrina. La reacción ASPE se llevó a cabo de la siguiente forma. Los BeadArrays™ se lavaron dos veces en tampón de hibridación 1X y a continuación se lavaron con tampón de reacción ASPE (sin enzima ni nucleótidos) a temperatura ambiente. La reacción ASPE se llevó a cabo sumergiendo los BeadArrays™ en una mezcla de reacción ASPE a temperatura ambiente durante 15 minutos. La mezcla ASPE contenía los siguientes componentes: dATP 3uM, dCTP 1,5 uM, biotina-11-dCTP 1,5 uM, dGTP 3 uM, dUTP 3 uM, tampón de extensión 1x GoldenGate™ (Illumina), sacarosa al 10%, betaína 500 mM, DTT 1 mM, 100 ug/ml de BSA, Tween 20 al 0,1% y 0,4 ul de KlenTaq (DNA Polymerase Inc., St. Louis, MO). La Figura 6A muestra los valores de la intensidad no tratados en los 77 pares de sondas. Las sondas PM (cuadrados) muestran intensidades muy superiores a las sondas MM para la mayoría de las sondas, permitiendo que la base interrogada se distinguiera de forma eficaz. La Figura 6B muestra una representación gráfica de los cocientes de discriminación (PM/PM+MM) para los 77 loci. Estos resultados demuestran que aproximadamente dos tercios de los loci tenían cocientes > 0,8.

Los resultados de este ejemplo demuestran que la hibridación de una población representativa de fragmentos de genoma con una matriz y la detección directa de varios loci tipables de los fragmentos de genoma hibridados proporciona una discriminación suficiente entre los locus para aplicaciones de genotipado.

EJEMPLO IV**Genotipación de fragmentos de ADN genómico amplificado**

Este ejemplo demuestra el genotipado de una población amplificada de fragmentos de genoma.

Muestras de ADN genómico de placenta humana se obtuvieron del Coriell Inst. Camden, NJ. El genoma se amplificó y se marcó con biotina usando amplificación con cebadores aleatorios en las condiciones descritas en el Ejemplo I, con la excepción de que la cantidad de genoma del molde se varió, y la longitud del cebador aleatorio se varió como se indica en la Figura 7. El resultado de la amplificación de todas las reacciones fue relativamente constante para

5

La población amplificada de fragmentos de genoma se genotipó de la siguiente forma. El genotipado se llevó a cabo en los servicios de genotipado por SNP de Illumina, usando el ensayo patentado GoldenGate™ con matrices IllumiCode™. La puntuación GenTrain es una métrica que determina el grado de agrupamiento de las intensidades de los loci SNP en una población de muestra. Una comparación de la puntuación GenTrain con el control no amplificado proporciona una estimación de la amplificación y sesgo del locus.

10

La calidad del genotipado del ADB no amplificado se comparó con la población amplificada de fragmentos de genoma tal como se muestra en la Figura 7. La cantidad de molde de genoma usado en la reacción de amplificación se muestra bajo cada barra. De los genes amplificados, las mejores puntuaciones Gen-Train se obtuvieron para la reacción de amplificación que usaban 1000 ng de genoma del molde (amplificación 40X). Las puntuaciones GenTrain para la reacción de amplificación que usan 1000 ng de molde de genoma fueron similares a las obtenidas para el ADN genómico no amplificado, lo que indica que el producto amplificado era representativo del genoma. También se obtuvieron puntuaciones GenTrain para la reacción de amplificación usando tan poco como 100 ng de molde de genoma (amplificación 400X).

15

20

Estos resultados demuestran que poblaciones amplificadas de fragmentos de genoma obtenidos de acuerdo con la invención son representativas de la secuencia de genoma en un ensayo de genotipado.

25 **EJEMPLO V**

Genotipación de genoma complete (Whole Genome Genotyping, WGG) de fragmentos amplificados de ADN genómico

Este ejemplo demuestra el genotipado del genoma completo de una población amplificada de fragmentos de genoma por hibridación directa con una matriz de ADN y una puntuación SNP con extensión de cebador basada en matriz.

30

Un conjunto de 3 X 32 muestras de ADN (1 ug cada una) se amplificaron mediante amplificación de cebadores aleatorios para producir muestras diana independientes que tienen 150 ug de fragmentos de ADN genómico. Las poblaciones amplificadas de fragmentos se hibridaron con BeadArrays™ que tienen 50-mer de sondas de captura ASPE que cubren 192 loci. Tras la hibridación, se llevó a cabo una reacción ASPE tal como se ha descrito en el Ejemplo III. Se recogieron imágenes y se analizaron las agrupaciones de genotipos mediante el programa informático patentado GenTrain (Illumina). Una imagen ilustrativa de un BeadArray™ detectado con ASPE se muestra en la Figura 11A.

35

40

La Figura 11B muestra una representación gráfica GenTrain de la comparación de theta con la intensidad para un locus. La intensidad es la intensidad total de la fluorescencia detectada para una perla concreta. Theta corresponde a la posición de la intensidad de la fluorescencia de una perla en un gráfico de dispersión de la intensidad de la fluorescencia para un alelo de un locus comparada con la intensidad de la fluorescencia para un segundo alelo del locus. En particular, la posición de la intensidad de la fluorescencia de una perla en el gráfico de dispersión corresponde a una coordenada x, y, concreta, y theta es el ángulo entre el eje x y una línea trazada desde el origen hasta dicha coordenada x, y. Tal como se muestra en la Figura 11B, se diferencian claramente dos agrupaciones homocigóticas y (B/B y A/A) una agrupación heterocigótica (A/B).

45

50

Aproximadamente el 52% de los loci proporcionaron agrupaciones bien resueltas que se consideraron loci "satisfactorios", que se analizaron posteriormente para determinar su genotipo en todas las muestras. El análisis de las llamadas del genotipo (101/192 loci) en 3 X 16 muestras cuyos genotipos de referencia eran conocidos indicó una concordancia del 99,95% (4090/4092) con un índice de llamada del 100% (Figura 12, Panel A). En las Figuras 12B y C se muestran representaciones gráficas GenCall que muestran las puntuaciones para diferentes loci de dos muestras diferentes. La puntuación GenCall de una llamada de genotipo individual es un valor comprendido entre 0 y 1 que indica la confianza en dicha llamada. Una puntuación más alta indica una mayor confianza en la llamada.

55

En las Figuras 12C y 12D se muestran representaciones gráficas GenTrain ilustrativas para dos loci diferentes. Estos datos muestran que, para la mayoría de las muestras, se diferenciaron claramente tres agrupaciones correspondientes a genotipos homocigóticos (B/B y A/A) y (A/B). Los dos puntos de color gris proceden de BeadArrays™ "sin control de diana".

60

El examen de las representaciones gráficas de las Figuras 12D y E mostró solamente dos llamadas cuestionables entre 4092 llamadas, lo que se indicó mediante flechas en las representaciones gráficas. Las llamadas se filtraron aplicando un umbral de 0,4 para la puntuación GenCall, lo que se muestra mediante la línea horizontal en las

65

Figuras 12B y C.

EJEMPLO VI

5 Inhibición de señales ectópicas

Este ejemplo demuestra el uso de la proteína de unión al ácido nucleico monocatenario (SSB) para inhibir la expresión ectópica en una reacción de extensión del cebador basada en matriz.

10 Las proteínas de unión al ácido nucleico monocatenario tales como SSB de *E. coli* y T4 gen 32 se ensayaron para determinar su capacidad para suprimir la extensión ectópica en reacciones ASPE de Klenow y Klentaq basadas en matriz. Las condiciones empleadas fueron las siguientes: la reacción ASPE de Klenow basada en matriz contenía Tris-Acetato 80 mM (pH 6,4), EDTA 0,4 mM, AcetatoMg 1,4 mM, DTT 0,5 mM, BSA 100 ug/ml, Tween-20 al 0,1%, 0,2 U/ul de exopolimerasa Klenow y dNTP 0,5 uM con nucleótidos marcados con biotina-11 en una relación 1:1 con
15 nucleótidos "fríos" de dCTP, dGTP, y dUTP. En los experimentos con SSB, la concentración fue 0,2 ug/20 ul rxn. Las condiciones para el Klentaq basado en matriz se han descrito en el Ejemplo III.

La Figura 14A muestra un gráfico de dispersión de reacciones ASPE analizadas con polimerasa Klenow en BeadArrays™ en presencia de SSB y ausencia de una muestra de ácido nucleico diana (ntc = sin control de diana).
20 Tal como se demuestra en la Figura 14C, la señal ectópica se redujo en gran medida en presencia de SSB en comparación con la ausencia de SSB. Se obtuvieron resultados similares para reacciones ASPE analizadas con polimerasa Klentaq. Los gráficos mostrados en las Figuras 14C y D se obtuvieron clasificando las señales de los gráficos de dispersión a lo largo del eje X por intensidad creciente. Tal como se muestra en la Figura 14B, la extensión específica de alelo se produjo en niveles detectables para las reacciones ASPE llevadas a cabo en
25 presencia de una muestra diana que contenía una población amplificada de fragmentos de genoma.

Estos resultados demuestran que la inclusión de SSB en un ensayo de extensión de cebador suprime la extensión ectópica manteniendo o mejorando a la vez la extensión específica de alelo. Estudios adicionales han indicado que la inclusión de SSB en una reacción ASPE basada en matriz mejoró la discriminación alélica.

30

EJEMPLO VII

Evaluación de poblaciones de fragmentos de genoma producidos por amplificación de cebadores aleatorios

35 Este ejemplo demuestra que poblaciones de fragmentos de genoma humano producidos mediante amplificación de cebadores aleatorios (RPA) son representativas de sus moldes de genoma, que tienen poco sesgo alélico y que se pueden generar de manera reproducible.

Se usaron reacciones RPA para producir poblaciones amplificadas de fragmentos de genoma procedentes de ADN genómico humano utilizando los métodos descritos en el Ejemplo V. Las reacciones de amplificación se llevaron a cabo en un formato monotubo sin necesidad de aislar los componentes o productos de la reacción antes de incubar las mezclas de reacción con las matrices de sondas. Con la excepción de las modificaciones descritas a continuación, las mezclas de reacción se incubaron con BeadArrays™ tal como se describe en el Ejemplo V y la detección se llevó a cabo mediante ASPE como se describe en el Ejemplo III.

45

Los resultados mostrados en la Figura 15 ilustran la representación conseguida durante el proceso de amplificación. Las reacciones RPA realizadas por duplicado sobre 100 ng de ADN genómico humano (Coriell Cell Repositories, Camden, NJ) en 100 ul dieron como resultado poblaciones de fragmentos de genoma que tienen 1-2 ug de ADN/ul. Las muestras duplicadas de genoma no amplificado consistieron en ADN de placenta humana (Sigma-Aldrich, N° de catálogo D3287) fragmentado con ADNasa I hasta un tamaño promedio de aproximadamente 200 a 300 bases.

50

Las muestras amplificadas y no amplificadas se hibridaron con matrices con sondas diseñadas dirigidas contra regiones no polimórficas del genoma. De este modo, todas las sondas fueron parejas perfectas del genoma, y deberían ampliarse en el ensayo de genotipado. Los valores de intensidad obtenidos para las sondas individuales tras la hibridación de dos muestras diferentes se representaron gráficamente en los gráficos de dispersión de la Figura 15. Tal como se muestra en la Figura 15A, se produjo un elevado grado de correlación entre las muestras duplicadas no amplificadas. De manera similar, y tal como se muestra en la Figura 15B, se observó una fuerte correlación entre las muestras amplificadas duplicadas, lo que indica que los métodos de amplificación produjeron resultados muy reproducibles. El gráfico de dispersión de la Figura 15C que compara las muestras amplificadas y no
55 amplificadas mostró una agrupación más difusa en comparación al observado para los duplicados, e indica que algunos loci estuvieron representados en exceso, mientras que otros estaban menos representados en la muestra amplificada.

60

Sin embargo, los resultados indican una buena representación. El número de sondas (agrupaciones) con cocientes concretos de intensidades de señal entre entradas de ADN no amplificado comparado con el amplificado (cociente amplificado: no amplificado) se representa gráficamente en la Figura 16A. Los datos demostraron que el 90,1% de

65

los loci detectados tienen una varianza de la intensidad en la población amplificada que no superó 0,5 a 2 veces comparado con la intensidad medida para el ADN genómico no amplificado. De esta forma, el 90,1% de los loci detectados en la población amplificada se representaron con una disminución no inferior a 0,5 veces, y con un exceso no superior a 2 veces comparado con sus cantidades relativas en el genoma no amplificado. Además, el 97,4% de los loci detectados en la población amplificada se representaron con una disminución no inferior a 0,3 veces y con un exceso no superior a 3 en comparación con sus cantidades relativas en el genoma no amplificado.

La población de fragmentos de genoma amplificada de forma representativa se comparó con las muestras de ADN de control en el ensayo GoldenGate™ (Illumina, Inc. San Diego, CA). Los datos ilustrativos procedentes de cuatro loci (1824, 2706, 3633 y 6126) se muestran en los Genoplots (también denominados gráficos GenTrain) de la Figura 17. Los genoplots son otra representación gráfica en coordenadas polares de los gráficos de dispersión de genotipado convencionales. Los gráficos de dispersión de genotipado convencionales tienen un eje de la intensidad detectada para un primer canal (que se correlaciona con un primer alelo) comparada con la intensidad detectada para un segundo canal de color (que se correlaciona con un segundo alelo) y representan gráficamente un punto de dispersión para cada locus de acuerdo con su intensidad en cada canal. Los Genoplots son otra representación gráfica de cada punto de dispersión de acuerdo con la distancia a una línea trazada desde el origen del punto de dispersión (R) y el ángulo entre la línea y el eje x (theta). Tal como se muestra en la Figura 17, los puntos de dispersión de los datos generados a partir de mezclas RPA producidas a partir de entradas de genoma de 10 ng, 100 ng o 1 ug dieron como resultado buenas agrupaciones comparadas con las agrupaciones de control (círculos) procedentes de ADN genómico no amplificado, indicando muy poco sesgo alélico.

Se demuestra que el límite de detección (LOD) en ensayos de genotipado aumenta con cantidades crecientes de ADN genómico introducidas en las reacciones RPA. Se llevaron a cabo reacciones RPA independientes (por duplicado) con diferentes cantidades de ADN genómico de entrada. Las cantidades de entrada estuvieron en un intervalo de 1 femtogramo a 100 nanogramos, incluyendo las cantidades representadas en el eje x de la Figura 18A. La Figura 18A es un gráfico de barras que muestra la intensidad promedio detectada en todas las muestras de cada matriz (LOD) tras la hibridación y detección mediante ASPE de las mezclas de reacción RPA generadas a partir de diferentes cantidades de ADN genómico de entrada (entrada). Tal como se muestra en la Figura 18A, cantidades de ADN genómico de entrada de 10 pg (aproximadamente 3 copias del genoma humano) o mayores dieron como resultado valores de LOD que estaban sustancialmente aumentadas en comparación con una reacción RPA de control en la que no se utilizó un ADN genómico de entrada (0 g). El LOD aumentó sustancialmente sobre el fondo cuando se utilizaron al menos 100 pg (30 copias del genoma), 1 ng (300 copias del genoma), 10 ng (3.000 copias del genoma) o 100 ng (30.000 copias del genoma) de ADN genómico humano de entrada en la reacción de RPA tal como se muestra en la Figura 18A.

Se demostró que la representación mejoraba cuando se introducían cantidades crecientes de ADN genómico en las reacciones RPA. El gráfico de barras mostrado en la Figura 18B representa gráficamente PM/(PM+MM) para todas las sondas de una matriz (cociente) cuando se utilizaban mezclas de sondas RPA producidas a partir de cantidades variables de ADN genómico de entrada (entrada). Cantidades de ADN genómico de entrada de 10 pg (aproximadamente 3 copias del genoma humano) más dieron como resultado una mejora sustancial en la representación cuando se compararon con una reacción RPA de control sin ADN genómico de entrada (0 g) o bajos niveles de ADN genómico (cantidades en femtogramos). La representación mejoró sustancialmente de forma adicional cuando se utilizaron al menos 100 pg (30 copias del genoma), 1 ng (300 copias del genoma), 10 ng (3.000 copias del genoma) o 100 ng (30.000 copias del genoma) de ADN genómico de entrada en la reacción RPA.

Estos resultados indican que la RPA se puede usar para producir cientos de microgramos de una población amplificada de fragmentos de genoma a partir de cantidades del molde de ADN genómico tan bajas como unos pocos picogramos. Las poblaciones amplificadas de fragmentos de genoma producidas mediante RPA tienen una buena representación, se pueden fabricar de manera reproducible, y tienen un bajo sesgo alélico. De esta forma, el ADN producido mediante RPA tiene calidad y cantidad suficiente para genotipar el genoma completo.

EJEMPLO VIII

Comportamiento del ensayo de genotipado del genoma completo

Este ejemplo demuestra el genotipado de un genoma completo de una población amplificada de fragmentos de genoma mediante hibridación directa con una matriz de ADN y los procedimientos de extensión de cebador basados en matriz producen resultados de puntuación SNP precisos y de alta calidad para sujetos humanos.

Se obtuvo ADN genómico (100 ng) procedente de 95 muestras del Centre d'Etude du Polymorphisme Humain (CEPH) en el conjunto utilizado para control de calidad en el proyecto International HapMap (para más información sobre las muestras, véase International HapMap Consortium, Nature 426:789-796 (2003)). Las reacciones RPA se llevaron a cabo tal como se ha descrito en el Ejemplo V, dando como resultado mezclas de reacción que contenían 188 ug de ADN en 100 ul. Las mezclas de reacción sin diluir se incubaron con los BeadArrays™ que tenían 50-mer de sondas específicas del conjunto de loci 1500 HapMap QC (para más información sobre los loci, véase International HapMap Consortium, Nature 426:789-796 (2003)) usando los métodos descritos en el Ejemplo V,

seguido por ASPE tal como se ha descrito en el Ejemplo III. A continuación, se tomaron imágenes de las matrices en un dispositivo lector con acoplamiento de carga tal como se ha descrito en Gunderson y col., Genome Res. 14: 870-877 (2004). Los genotipos SNP se llamaron mediante el programa informático GenCall (Illumina Inc., San Diego, CA).

5 Las Figuras 19A y 19B muestran Genoplots (también denominados gráficos GenTrain) representativos de los loci 860 y 954, respectivamente. Se obtuvo una Buena separación de agrupamientos para los loci 860 y 954, dando como resultado puntuaciones de agrupaciones de genes (GCS) de 7,5 y 4,4, respectivamente ($GCS = \text{Min}[(\text{Abs}(\theta_{AB} - \theta_{AA}) / (\sigma_{AB} + \sigma_{AA})), (\text{Abs}(\theta_{AB} - \theta_{BB}) / (\sigma_{AB} + \sigma_{BB}))]$), donde θ_{AB} es el valor de θ promedio para la agrupación AB (θ se ha descrito anteriormente con respecto a la Figura 11) y σ_{AB} es la desviación estándar de θ_{AB}). La Figura 19C muestra una distribución de loci de acuerdo con una puntuación de separación entre agrupaciones de genotipo. Más del 75% de los loci obtuvieron un GCS de 3,0 o superior (barras oscuras) y de este modo, se consideraron aceptables para genotipado.

15 En la Tabla 1 se muestra un resumen de la estadística de genotipado obtenida en la interrogación del conjunto de loci HapMap QC en las muestras CEPH. Se evaluó la tasa de conversión del ensayo por recuento del número de loci que consiguieron detectar un alelo minoritario. Los loci no polimórficos y los loci de alto número de copias se contaron como fracasos del ensayo con respecto al desarrollo de un ensayo SNP real. Técnicamente, muchos de los loci no polimórficos eran ensayos correctos pero no se contaron porque no mostraban un alelo minoritario. La tasa de conversión del ensayo comparada con los resultados obtenidos en el ensayo Golden Gate (Illumina, Inc. San Diego, CA) con las mismas muestras de ADN genómico fue del 95%. La tasa de llamada fue bastante elevada del 99,5% y la reproducibilidad fue superior al 99,99%.

25 Se determinó la concordancia entre los resultados del genotipado obtenidos según se ha descrito anteriormente y los resultados del genotipado obtenidos para las mismas muestras y loci mediante el ensayo Golden Gate (Illumina, Inc. San Diego, GA). La concordancia fue superior al 99%.

Tabla 1

Parámetro	Valores	Porcentaje
Conversión del ensayo	819/864	95%
Tasa de llamada	68807/68970	99,5%
Reproducibilidad	8189/8190	99,99%
Concordancia	137.456/137.614	99,9%

30 Estos resultados indican que los ensayos de genotipado del genoma completo proporcionan datos de genotipado de alta calidad, equivalentes a los del ensayo Golden Gate que se utiliza en la actualidad para genotipar una gran parte del genoma en el proyecto International HapMap.

35 **EJEMPLO IX**

Matrices de arrastre para eliminar dianas hibridadas antes de la detección

40 Este ejemplo demuestra la eliminación de dianas hibridadas de una matriz mediante arrastre con NaOH 0,1 N tras modificar las sondas mediante extensión de la polimerasa dependiente de la diana.

45 Se obtuvo ADN genómico del Coriell Cell Repositories (Camden, NJ). Se llevaron a cabo reacciones RPA tal como se ha descrito en el Ejemplo VII. Las mezclas de reacción resultantes se hibridaron con BeadArrays™ y se llevaron a cabo reacciones ASPE tal como se ha descrito en el Ejemplo III. Tras la reacción ASPE y antes de la detección de la señal de fluorescencia, las matrices se trataron con NaOH 0,1 N en agua (+NaOH) o tampón de hibridación 1X, que carecía de formamida (-NaOH). Las matrices se detectaron tal como se ha descrito en el Ejemplo VIII.

50 Tal como se muestra en la Figura 20, el arrastre posterior a la extensión de la matriz con NaOH redujo la señal de fondo de las sondas emparejadas incorrectamente, y dio como resultado una mayor diferencia ratiométrica entre la señal de las sondas con emparejamiento incorrecto y las de emparejamiento perfecto.

Estos resultados indican que las matrices de arrastre tras la modificación de la sonda, aunque no son necesarias, se pueden utilizar para mejorar en gran medida la especificidad del ensayo.

55 **EJEMPLO X**

Amplificación del genoma completo de ADN tratado con bisulfito

Este ejemplo describe métodos para amplificar el genoma completo de ADN tratado con bisulfito. De forma típica, el tratamiento del ADN con bisulfito genera una despurinación importante y la fragmentación paralela del ADN. Este producto fragmentado se amplifica de forma típica con bajo rendimiento usando polimerasas de desplazamiento de cadena en la solución de amplificación del genoma con cebadores aleatorios. Se describen aquí dos soluciones para mejorar el rendimiento de amplificación. La primera solución es la concatenación de la muestra fragmentada, y el uso de los productos concatenados más largos como moldes en la amplificación con cebadores aleatorios con desplazamiento de cadena. La segunda solución crea una representación a partir de las dianas fragmentadas por unión de sitio de cebado universal a los extremos de los fragmentos.

- 5
- 10 De forma típica, el tratamiento del ADN genómico con bisulfito se utiliza para detectar la metilación basada en una reacción mediante la cual la citosina se convierte en uracilo, pero la 5-metilcitosina permanece no reactiva (véase, por ejemplo, Feil y col. *Nucleic Acids Res*, 22; 695-696 (1994); Frommer y col., *Proc Natl Acad Sci USA*, 89; 1827-1831 (1992)). Una reacción adicional del ADN con bisulfito es una despurinación y fragmentación paralela. Los fragmentos de ADN producidos mediante el tratamiento con bisulfito contienen un grupo fosfato en el extremo 3'.
- 15 Este grupo fosfato bloquea eficazmente la reacción del extremo 3' con nucleótidos o polinucleótidos únicos mediante el uso de varias enzimas biológicas.

Concatenación de ADN genómico tratado con bisulfito

- 20 El grupo 3' fosfato del ADN genómico tratado con bisulfito se eliminó por tratamiento con fosfatasa alcalina o la actividad 3' fosfatasa de la ADN quinasa de T4 en condiciones normalizadas recomendadas por el proveedor. La ADN quinasa de T4 mantiene intacto el 5' fosfato eliminando al mismo tiempo el 3' fosfato (en presencia de ATP), dando como resultado un producto que tiene un 5' fosfato y un 3' hidroxilo (véase la Figura 21A). Por el contrario, la fosfatasa alcalina elimina los dos, 5' fosfato y 3' fosfato, dando como resultado un producto que tiene ambos 5' hidroxilo y 3' hidroxilo (véase la Figura 21A).

- Tras la eliminación del 3' fosfato mediante la ADN quinasa de T4 los productos se incubaron a continuación con ARN ligasa de T4 para crear concatémeros usando las condiciones descritas en McCoy y col., *más arriba* (1980). Los concatémeros lineales y circulares resultantes que tienen diferentes tamaños se amplificaron mediante la amplificación con cebadores aleatorios tal como se ha descrito en el presente documento, por ejemplo, en el Ejemplo V. Este producto amplificado se utilizó a continuación para genotipado tal como se ha descrito en el presente documento, por ejemplo, en el VII, y proporciona un medio para llevar a cabo un perfil de metilación del genoma completo.
- 30

Colas de ADN genómico tratado con bisulfito

Los 3' fosfatos de los fragmentos tratados con bisulfito se convirtieron en 3' hidroxilos tal como se ha descrito anteriormente. Se añadieron colas universales al producto usando uno de tres métodos diferentes.

- 40 El primer método es el tratamiento de los fragmentos de ADN con desoxinucleótido transferasa terminal (TdT) y dGTP para añadir una cola de poliguanilato al extremo 3' (véase la Figura 21C). Se añadió una cola universal al extremo 5 del fragmento incubado con ADN ligasa y un oligonucleótido que tiene un duplete aleatorio tetramero adaptador en 3' y una secuencia 5' universal en el sitio del cebador (Figura 21 C) usando condiciones normalizadas recomendadas por el proveedor. Los fragmentos resultantes se amplificaron mediante la reacción en cadena de la polimerasa mediante el uso de un cebador universal (cebador A en la Figura 21C) que complementa la cola 5' universal en el sitio del cebador de los fragmentos y un cebador de policitidilato (cebador B en la Figura 21C) que complementa la cola de 3' poliguanilato de los fragmentos.
- 45

- En el segundo método, se añade una cola 5' mediante ligadura mediada por ARN ligasa de T4 de un oligonucleótido que tiene un sitio de cebador universal usando condiciones normalizadas recomendadas por el proveedor. Tal como se muestra en la Figura 21D, se lleva a cabo en dos etapas. En la primera etapa, un oligonucleótido que tiene un sitio de cebador universal que tiene un 5' fosfato pero que carece de 3' hidroxilo se hace reaccionar con el fragmento de manera tal que se añade al fragmento una cola 3'. En la segunda etapa, un oligonucleótido que tiene un sitio de cebador universal que tiene un 3' hidroxilo pero que carece de un 5' fosfato se hace reaccionar con el fragmento de manera tal que se añade al fragmento una cola 5'. El uso de nucleótidos bloqueados en ambas etapas reduce las reacciones secundarias no deseadas debido a la autoligadura de los oligonucleótidos que tienen un sitio de cebador universal. Los fragmentos resultantes se amplificaron mediante la reacción en cadena de la polimerasa mediante el uso de un cebador universal (cebador A en la Figura 21D) que complementa la cola 5' universal en el sitio del cebador de los fragmentos y un cebador universal (cebador B en la Figura 21D) que complementa el sitio del cebador universal 3' de los fragmentos. Este producto amplificado se utiliza a continuación para genotipado tal como se ha descrito en el presente documento, por ejemplo, en el Ejemplo VII, y proporciona un medio para llevar a cabo un perfil de metilación del genoma completo.
- 50
- 55
- 60

- El tercer método emplea ligadura directa de oligonucleótidos que tienen un sitio de cebador universal en ambos extremos 3' y 5' usando ARN polimerasa de T4 usando condiciones normalizadas recomendadas por el proveedor. A continuación se utilizan cebadores universales complementarios para amplificar los fragmentos mediante la
- 65

reacción en cadena de la polimerasa. Este producto amplificado se utiliza a continuación para genotipado tal como se ha descrito en el presente documento, por ejemplo, en el Ejemplo VII, y proporciona un medio para llevar a cabo un perfil de metilación del genoma completo.

- 5 En el presente documento, se pretende que el término “que comprende” sea no inclusivo, incluyendo no solamente los elementos citados, sino que también abarque cualesquiera elementos adicionales.

La divulgación también se puede definir con respecto a las siguientes cláusulas:

- 10 1. Un método para detectar loci tipables en un genoma, que comprende (a) proporcionar una población representativa amplificada de fragmentos de genoma que comprenden dichos loci tipables, donde dicha población comprende una representación de complejidad elevada; (b) poner en contacto dichos fragmentos de genoma con una pluralidad de diferentes sondas de ácidos nucleicos que tienen secuencias correspondientes a dichos loci tipables en condiciones en que se forman híbridos sonda-fragmento; y (c) detectar directamente los loci tipables de dichos híbridos sonda-fragmento.
- 15
- 20 2. Un método, que comprende (a) poner en contacto una población representativa amplificada de fragmentos de genoma con una pluralidad de sondas de ácido nucleico inmovilizado diferentes en condiciones en que se forman híbridos sonda-fragmento; (b) modificar dichas sondas inmovilizadas mientras están hibridadas a dichos fragmentos de genoma, formando de este modo sondas inmovilizadas modificadas; (c) retirar dichos fragmentos de genoma de dichos híbridos sonda-fragmento; y (d) detectar dichas sondas inmovilizadas modificadas después de retirar dichos fragmentos de genoma, detectando de este modo los loci tipables de dichos híbridos sonda-fragmento..
- 25
- 30 3. Un método, que comprende (a) poner en contacto una población representativa amplificada de fragmentos de genoma con una pluralidad de sondas de ácido nucleico inmovilizado diferentes en condiciones en que se forman híbridos sonda-fragmento; (b) modificar dichos híbridos sonda-fragmento inmovilizados por adición de un resto de detección a la sonda, formando de esta manera sondas con ligando marcado por afinidad; (c) poner en contacto dichas sondas con ligando marcado por afinidad con un resto de unión y un reactivo de amplificación, donde dicho resto de unión tiene uno o más sitios capaces de unir dicho ligando, y donde dicho reactivo de amplificación tiene afinidad por dicho resto de unión, mediante lo cual se forman complejos multiméricos entre dichas sondas con ligando marcado por afinidad, dichos restos de unión y dichos reactivos de amplificación; y (d) detectar dichos complejos multiméricos, detectando de esta forma los loci tipables de dichos fragmentos de genoma.
- 35
- 40 4. El método de la cláusula 1,2 o 3, donde dicha pluralidad de sondas de ácidos nucleicos diferentes comprenden una matriz de dichas sondas adheridas a una superficie.
5. El método de la cláusula 1,2 o 3, donde dicha pluralidad de sondas de ácidos nucleicos está adherida a partículas.
- 45 6. El método de la cláusula 5, donde cada una de dichas partículas está unida a un único tipo de sonda de ácido nucleico.
7. El método de la cláusula 5, donde dichas partículas están unidas a un sustrato.
8. El método de la cláusula 1,2 o 3, donde dichas sondas tienen un máximo de 125 nucleótidos de longitud.
- 50 9. El método de la cláusula 1, 2 o 3, que comprende además amplificar de forma representativa un genoma natural, proporcionando de esta forma dicha población representativa amplificada de fragmentos de genoma.
10. El método de la cláusula 9, que comprende amplificar de forma representativa dicho genoma nativo en una única etapa de reacción.
- 55 11. El método de la cláusula 9, que comprende amplificar de forma representativa dicho genoma nativo en condiciones isotermas.
12. El método de la cláusula 9, donde dicha amplificación representativa comprende una PCR del adaptador-enlazador.
- 60 13. El método de la cláusula 9, donde dicha amplificación representativa comprende una amplificación mediante cebadores aleatorios.
- 65 14. El método de la cláusula 13, que comprende una amplificación mediante el uso de una polimerasa con baja capacidad de procesamiento.

15. El método de la cláusula 14, donde dicha baja capacidad de procesamiento es inferior a 100 bases por evento de polimerización.
- 5 16. El método de la cláusula 9, donde dicha amplificación representativa comprende tratar dicho ADN genómico con una endonucleasa.
17. El método de la cláusula 9, donde dicha amplificación representativa comprende tratar dichos fragmentos de genoma con una endonucleasa.
- 10 18. El método de la cláusula 9, donde se usa un máximo de $1 < 10^6$ copias de dicho genoma nativo como molde de dicha amplificación representativa.
19. El método de la cláusula 9, donde dicho genoma nativo es genoma humano.
- 15 20. El método de la cláusula 9, donde dicha población representativa amplificada de fragmentos de genoma comprende secuencias idénticas en al menos un 5% de dicho genoma nativo.
21. El método de la cláusula 1, 2 o 3, donde dicha población representativa amplificada de fragmentos de genoma comprende una concentración de al menos 1 ug/ul de ADN.
- 20 22. El método de la cláusula 1,2 o 3, donde dicha población representativa amplificada de fragmentos de genoma comprende una complejidad de al menos 1 Gigabases.
23. El método de la cláusula 1,2 o 3, donde dicha población representativa amplificada de fragmentos de genoma comprende al menos 100 ug de ADN.
- 25 24. El método de la cláusula 1,2 o 3, donde al menos 100.000 de dichas sondas de ácido nucleico diferentes se hibridan con fragmentos de genoma para formar híbridos sonda-fragmento.
- 30 25. El método de la cláusula 1, donde dicha detección comprende modificar dichas sondas mientras están hibridadas a dichos fragmentos de genoma.
26. El método de la cláusula 25, donde dichas sondas de ácidos nucleicos están inmovilizadas.
- 35 27. El método de la cláusula 2 o 26, que comprende además exponer dichas sondas inmovilizadas modificadas a condiciones de desnaturalización antes de dicha detección, eliminando por tanto dichos fragmentos de genoma.
28. El método de la cláusula 2 o 25, donde dicha modificación comprende la adición de un resto de detección a la sonda, formando de esta manera sondas con ligando marcado por afinidad.
- 40 29. El método de la cláusula 28, que comprende además poner en contacto dichas sondas con ligando marcado por afinidad con un receptor y un reactivo de amplificación, donde dicho receptor tiene uno o más sitios que pueden unir dicho ligando, y donde dicho reactivo de amplificación tiene afinidad por dicho receptor, donde se forman complejos multiméricos entre dichas sondas con ligando marcado por afinidad, dicho receptor y dicho reactivo de amplificación.
- 45 30. El método de la cláusula 29, donde dicha detección comprende detectar dichos complejos multiméricos.
- 50 31. El método de la cláusula 2,3 o 25, donde dicha modificación comprende la adición de nucleótidos o análogos de nucleótido mediante una polimerasa.
32. El método de la cláusula 2,3 o 25, donde dicha modificación comprende la unión de sondas a dichas sondas de ácido nucleico inmovilizado.
- 55 33. El método de la cláusula 2,3 o 25, donde dicha modificación comprende la escisión de dichas sondas de ácido nucleico inmovilizado.
34. El método de la cláusula 2,3 o 25, que comprende además poner en contacto dichas sondas con una proteína de unión a ácido nucleico monocatenario.
- 60 35. El método de la cláusula 1, 2 o 3, donde dichos híbridos sonda-fragmento se han formado en una celda de flujo con hueco capilar.
- 65 36. El método de la cláusula 1,2 o 3, donde se detectan al menos 100 loci tipables simultáneamente.

37. El método de la cláusula 1, que comprende además la producción de un informe que identifica dichos loci tipables que se detectan.

38. Un informe producido según el método de la cláusula 37.

5

REIVINDICACIONES

1. Un método para indicar la presencia de un polimorfismo de nucleótido único (SNP) de interés en los loci tipables de un genoma, que comprende las etapas de:
- 5 (a) amplificar representativamente un genoma nativo para obtener una población representativa de fragmentos de genoma que comprende loci tipables, teniendo cada uno una posición de interrogación que corresponde a un SNP;
- 10 (b) poner en contacto dichos fragmentos con una matriz de diferentes sondas de ácidos nucleicos inmovilizados en condiciones en las que se forman los híbridos sonda-fragmentos, donde dichos fragmentos tienen una concentración de al menos 1 µg/µl de ADN y donde dichas sondas de ácidos nucleicos inmovilizados comprenden una región que es complementaria de los loci tipables de un genoma;
- 15 (c) poner en contacto la pluralidad de sondas de ácidos nucleicos con un inhibidor de la extensión ectópica (EEI) en condiciones en que se forman híbridos de EEI de la sonda;
- (d) poner en contacto dichos híbridos sonda-fragmento con una enzima de modificación, donde si la posición de interrogación de un locus tipable contiene un SNP de interés, un resto de detección se une a la sonda hibridada; y
- 20 (e) detectar la sonda, indicando de esta forma la presencia de dichos polimorfismos de nucleótidos únicos de interés en los loci tipables de dicho genoma;
- y
donde la etapa (e) detecta los híbridos sonda-fragmento selectivamente en comparación con los híbridos sonda-EEI.
- 25 2. El método de la reivindicación 1, donde una complejidad de al menos 0,1, 0,2, 0,5, 1, 2 o 3 Gigabases de una única secuencia de dicho genoma nativo está presente en dicha población de fragmentos de genoma.
3. El método de la reivindicación 1, donde dicha amplificación representativa comprende una amplificación del cebador aleatorio.
- 30 4. El método de la reivindicación 1, donde dicha población representativa amplificada de fragmentos de genoma se amplifica en condiciones isotermas.
5. El método de la reivindicación 1, donde dicha población representativa de fragmentos de genoma que se pone en contacto con dicha pluralidad de diferentes sondas de ácido nucleico comprende al menos 10 µg, 50 µg, 100 µg, 150 µg, 200 µg o 300 µg de ADN.
- 35 6. El método de la reivindicación 1, donde al menos 100, 500, 1000, 5000, 10,000, 50,000 100,000, 500,000 o 10⁶ de dichas sondas de ácidos nucleicos diferentes se hibridan con fragmentos de genoma para formar híbridos sonda-fragmento.
- 40 7. El método de la reivindicación 1, donde las regiones complementarias de dichas sondas tienen una longitud máxima de 10, 20, 30, 40, 50, 60, 80, o 100 nucleótidos.
- 45 8. El método de cualquiera de las reivindicaciones anteriores, donde dichas sondas de ácidos nucleicos inmovilizados se unen a cualquiera de partículas, el interior de una celda de flujo o un sustrato plano.
9. El método de cualquiera de las reivindicaciones anteriores, donde dichas sondas de ácidos nucleicos inmovilizados se configuran como una matriz de matrices, que tiene una pluralidad de matrices individuales configuradas para permitir el procesamiento de múltiples muestras y/u opcionalmente donde dichas matrices individuales se configuran para disponerse en los pocillos de una placa de microvaloración, y el método comprende además disponer dichas matrices individuales en los pocillos de la placa de microvaloración.
- 50 10. El método de cualquiera de las reivindicaciones anteriores, donde el resto de detección es un ligando de afinidad y la etapa (e) comprende poner en contacto dichas sondas marcadas con el ligando de afinidad con un receptor y un reactivo de amplificación, donde dicho receptor tiene uno o más sitios capaces de unirse a dicho ligando, y donde dicho reactivo de amplificación tiene afinidad por dicho receptor, por lo cual se forman complejos multiméricos entre dichas sondas marcadas con el ligando de afinidad y dicho reactivo de amplificación.
- 55 60 11. El método de cualquier reivindicación anterior, en el que el EEI es una proteína de unión al ácido nucleico monocatenario (SSB).
- 65 12. El método de cualquier reivindicación anterior, en el que el EEI es un oligonucleótido de bloqueo que es complementario del extremo 3' de una sonda de ácido nucleico.

13. El método de cualquiera de las reivindicaciones anteriores, donde al menos una de la pluralidad de diferentes sondas de ácidos nucleicos comprende una primera región y una segunda región, donde la segunda región es complementaria de la primera región, por lo que una sonda que no forma un híbrido sonda-fragmento forma una estructura en horquilla que no puede añadir un nucleótido o análogo de nucleótido mediante una polimerasa.

5 14. El método de la reivindicación 13, donde la primera y segunda regiones de la sonda del ácido nucleico no se hibridan sustancialmente a las temperaturas utilizadas durante la hibridación sonda-fragmento, pero se hibrida para formar estructuras en horquilla a las temperaturas utilizadas durante la adición de la polimerasa.

10

Figura 1

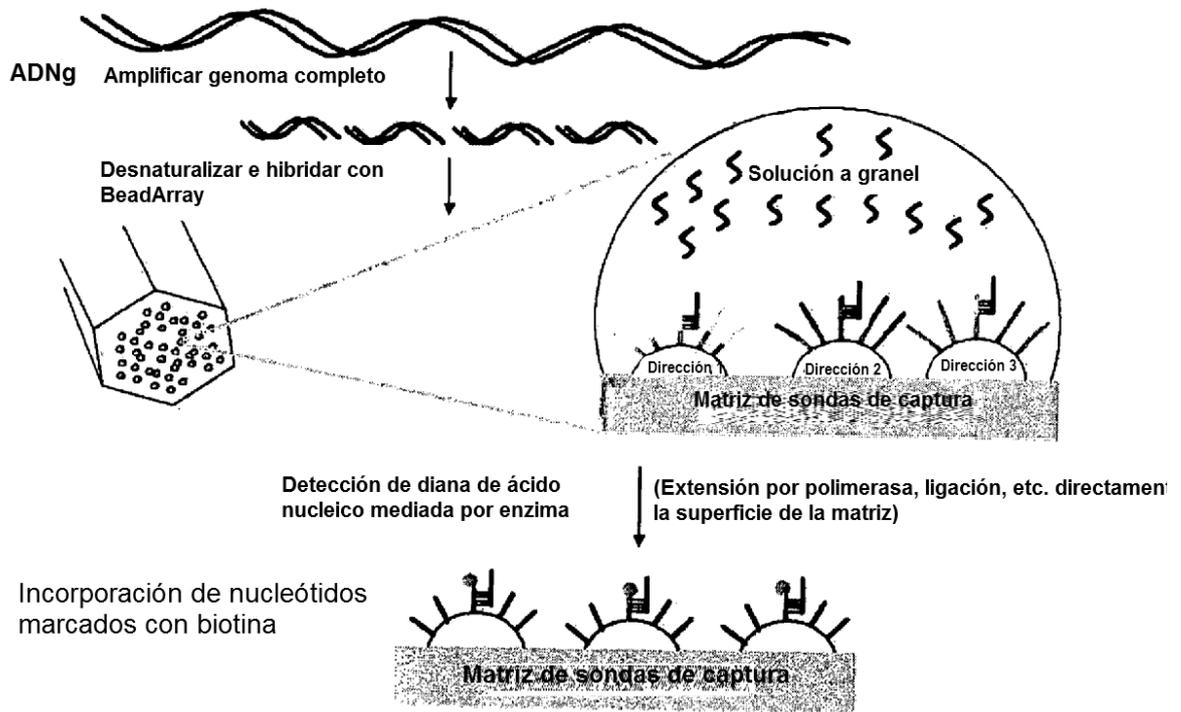


Figura 2

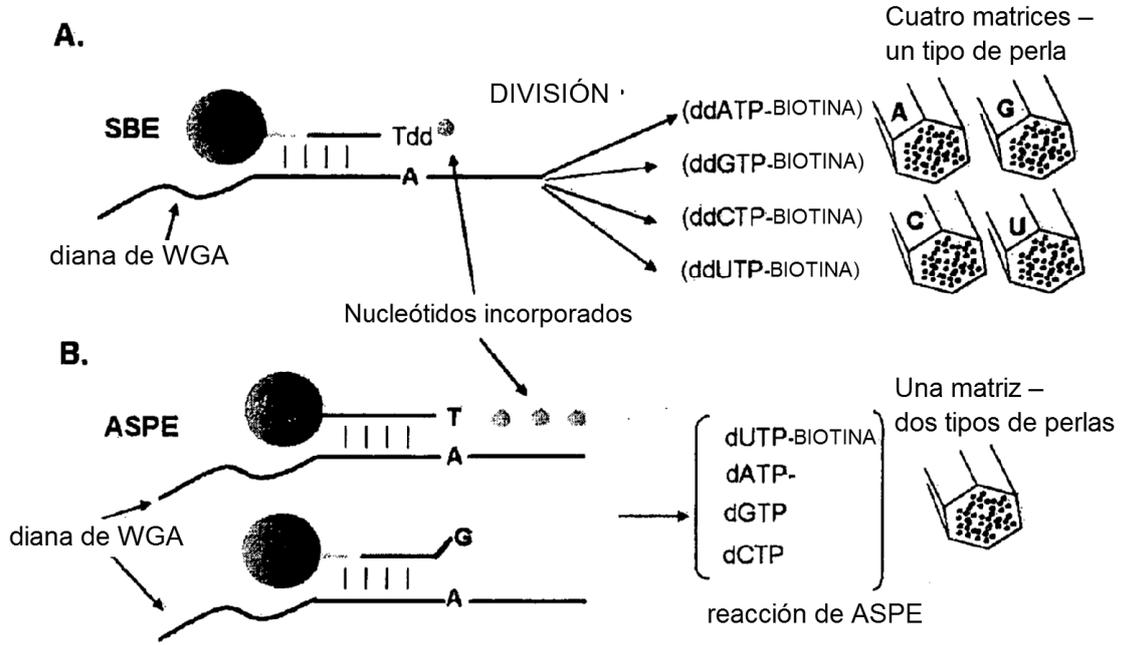
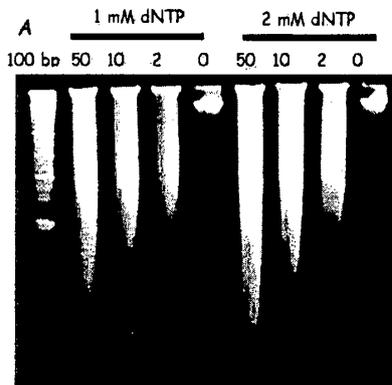


Figura 3



Condición	dNTP (mM)	cebador (uM)	Rendimiento (ug)	Amplificación
1	1	50	49,5	4950
2	1	10	28,5	2850
3	1	2	12,0	1200
4	1	0	6,0	600
5	2	50	79,5	7950
6	2	10	42,0	4200
7	2	2	21,0	2100
8	2	0	6,0	600

Figura 4

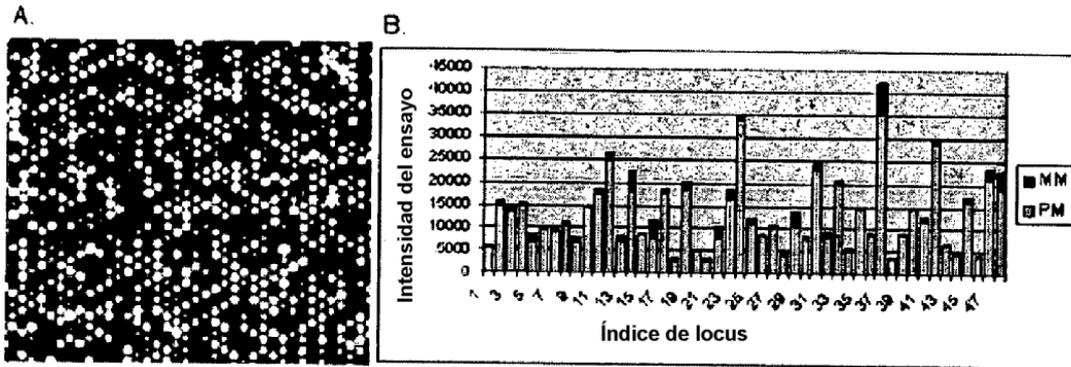


Figura 5

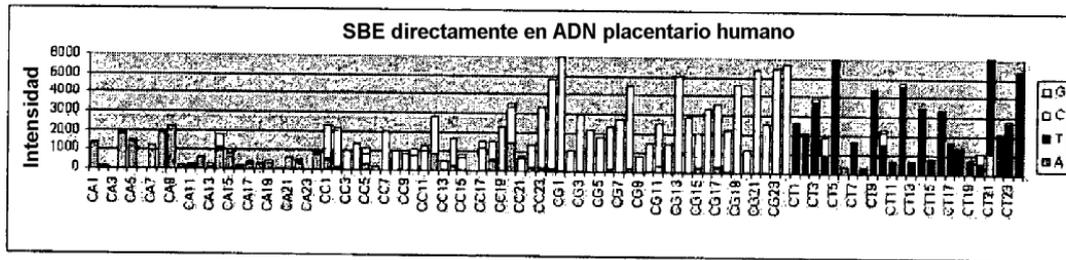


Figura 6

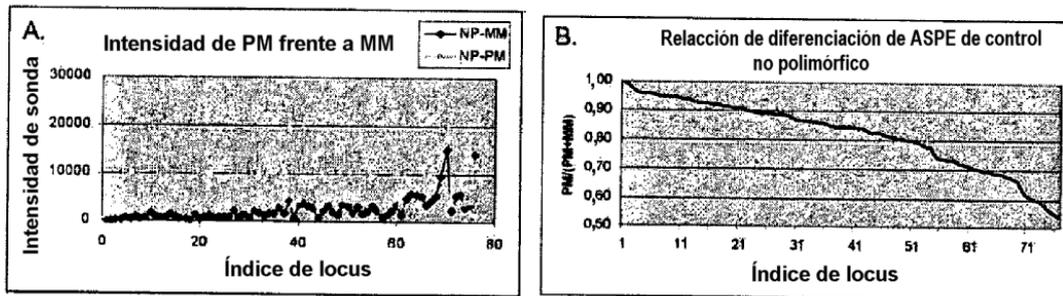


Figura 7

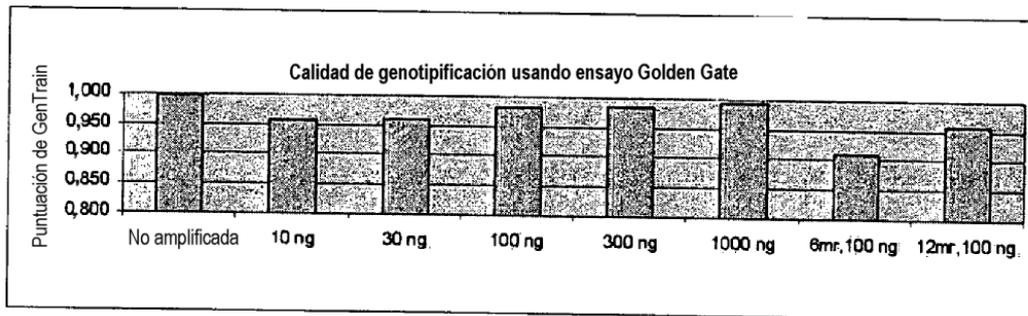


Figura 8

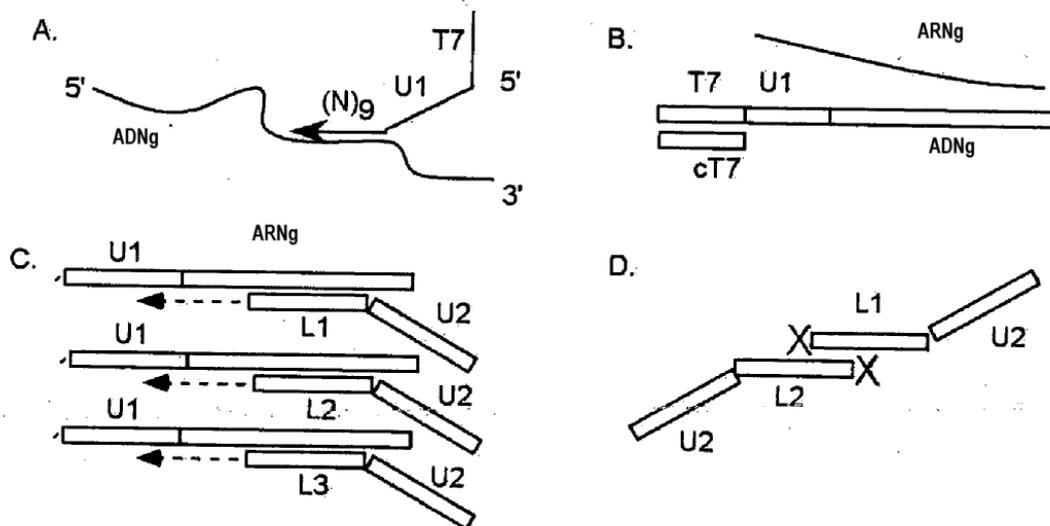


Figura 11

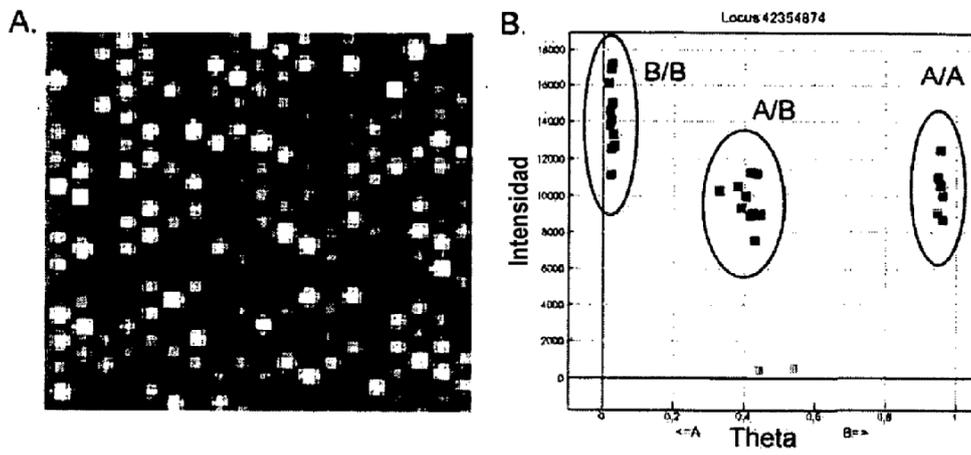


Figura 12

A.

Precisión de la genotipificación:

Tasa de llamada	Precisión	Llamada	Llamadas correctas
100%	99,95%	4092	4090
99,70%	100%	4078	4078

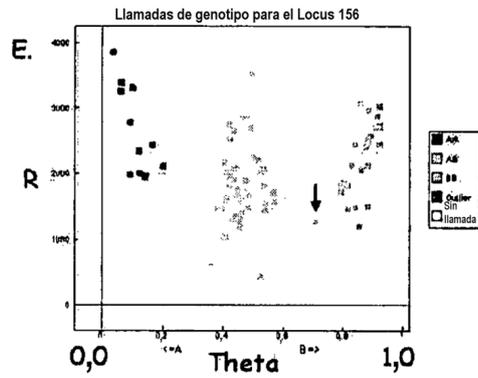
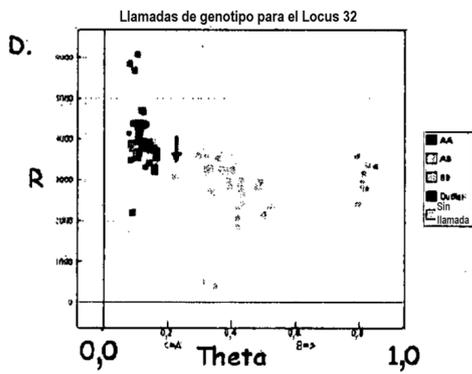
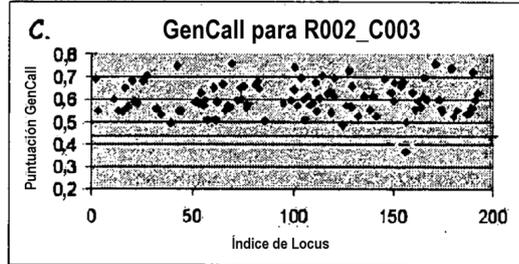
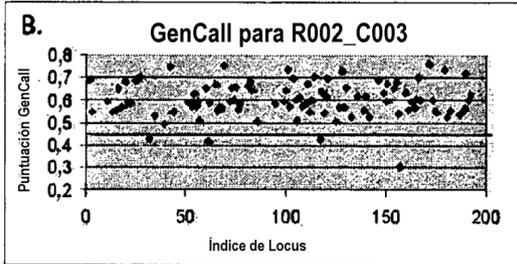


Figura 13

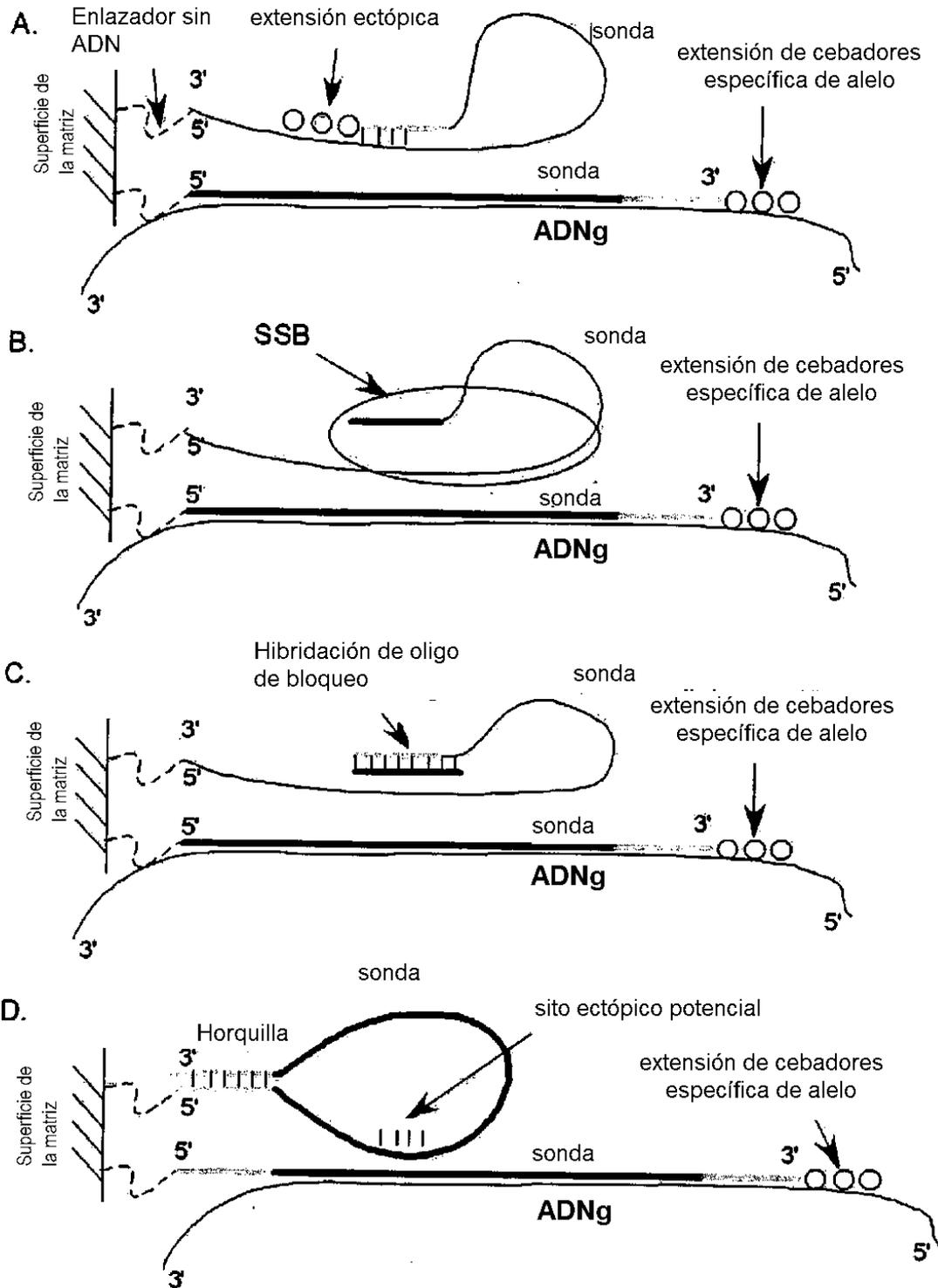


Figura 14

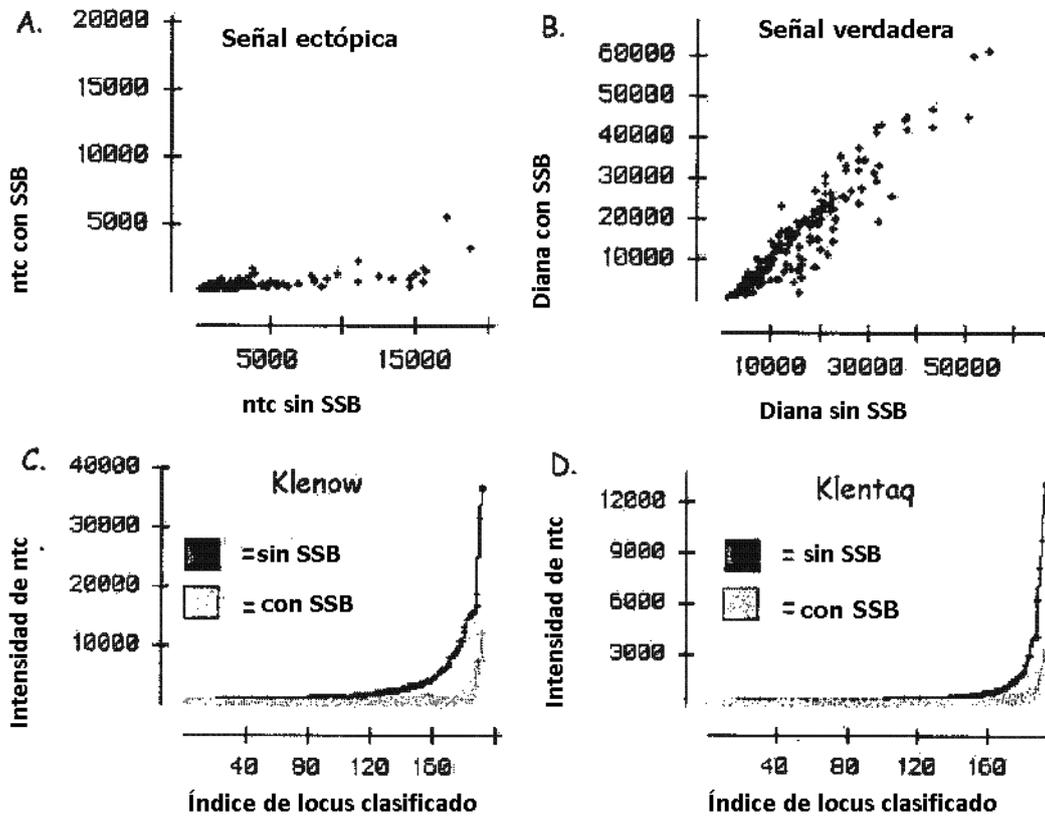


Figura 15 A

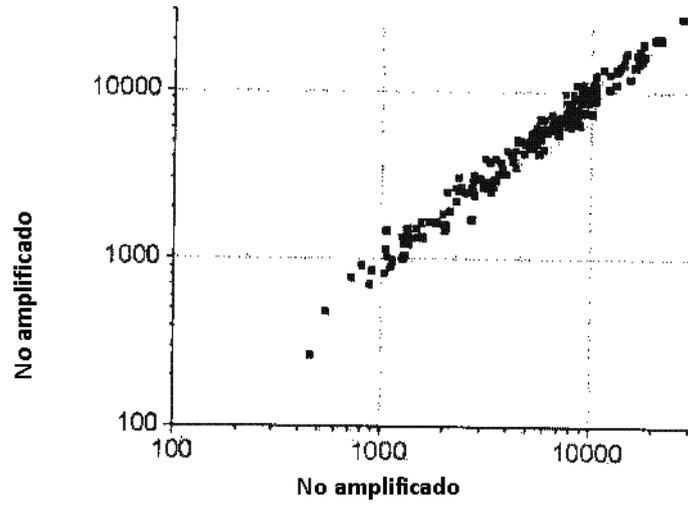


Figura 15 B

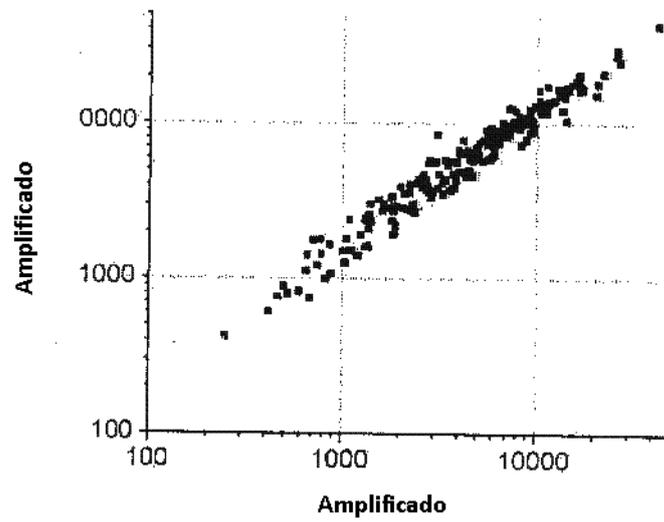


Figura 15 C

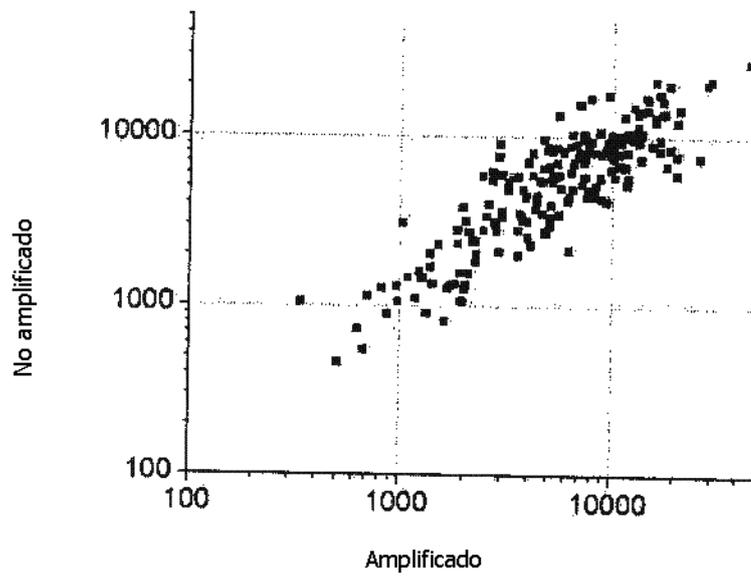


Figura 16 A

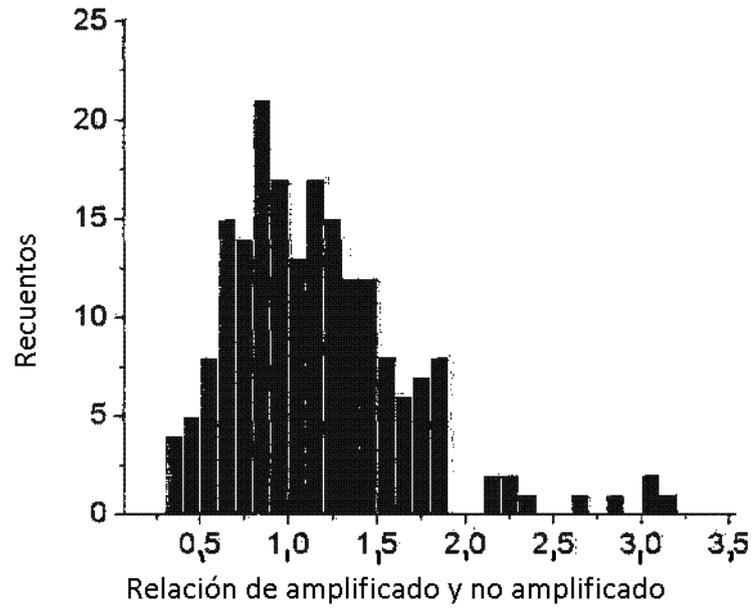


Figura 16 B

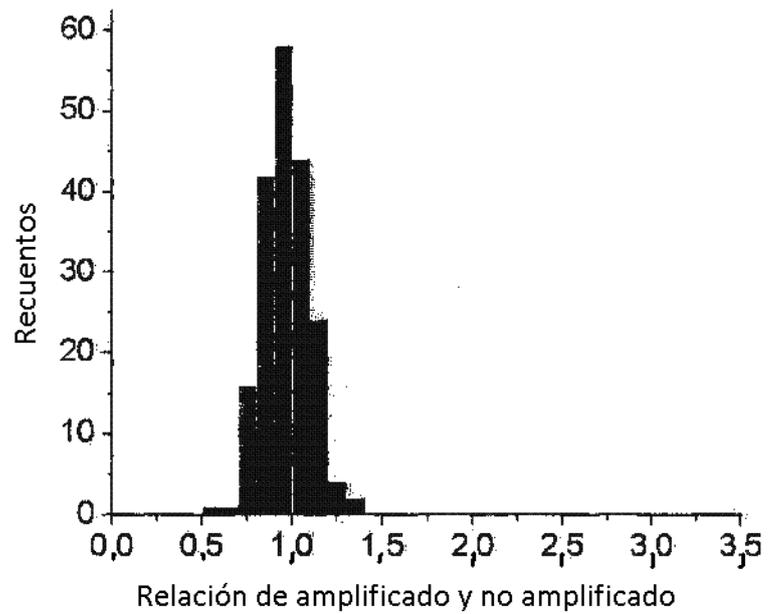


Figura 17

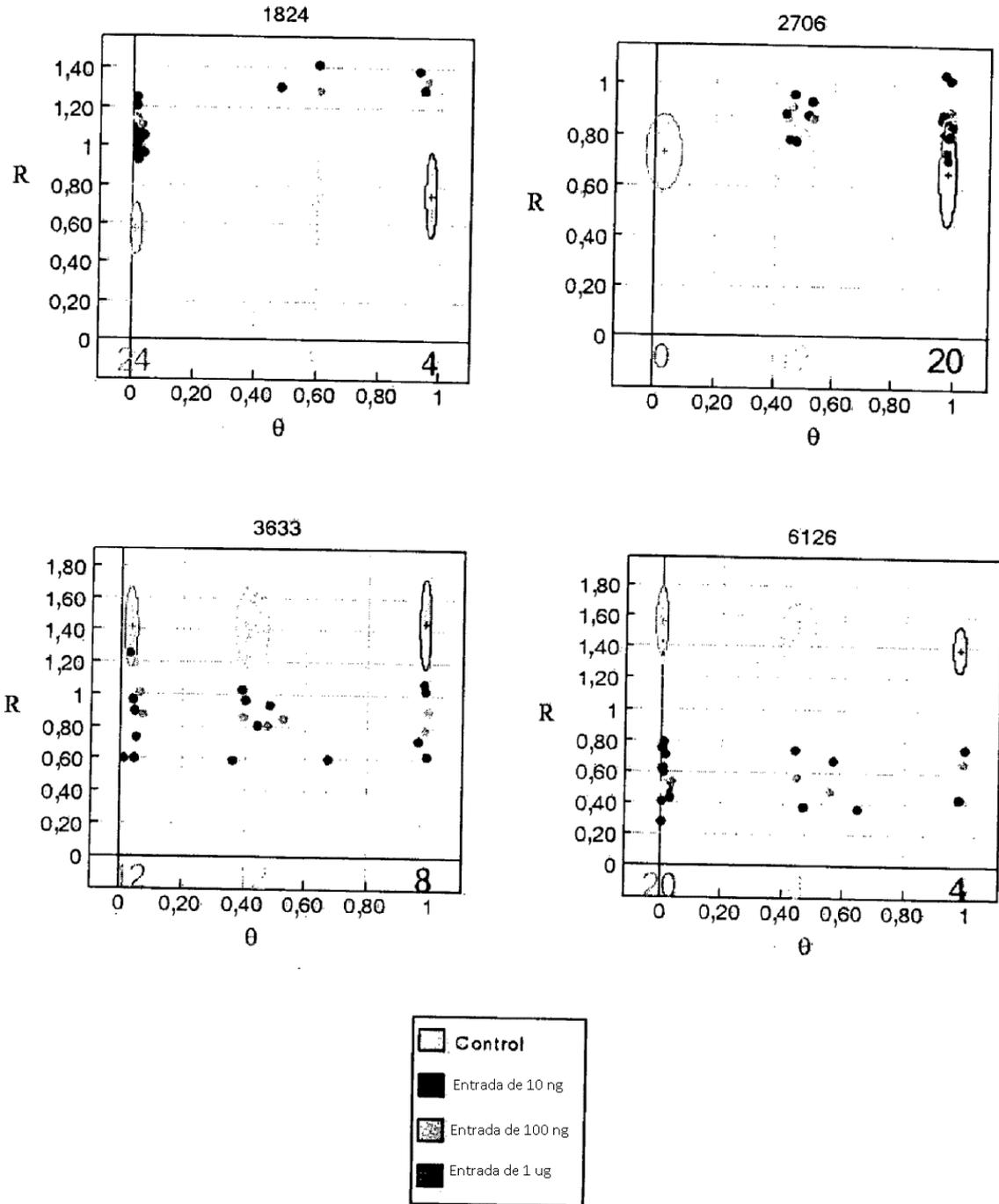


Figura 18A

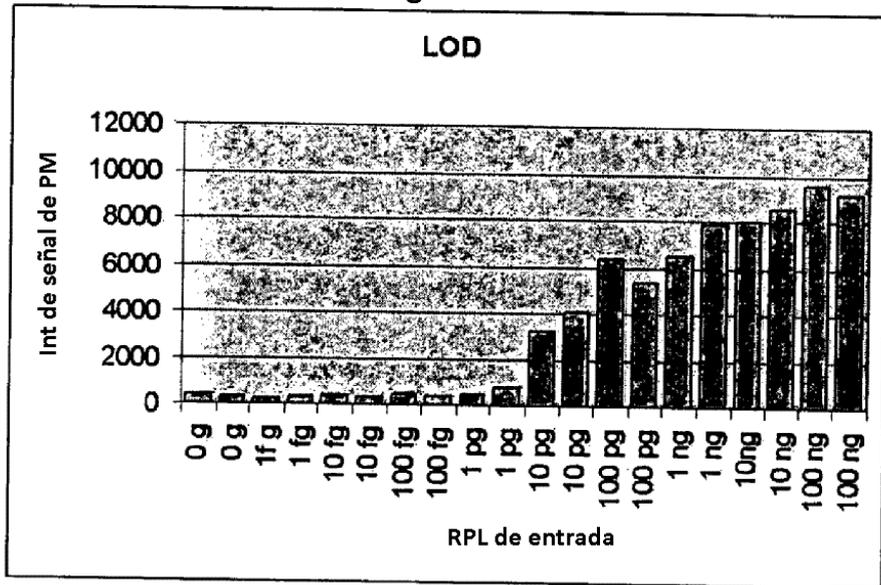


Figura 18B

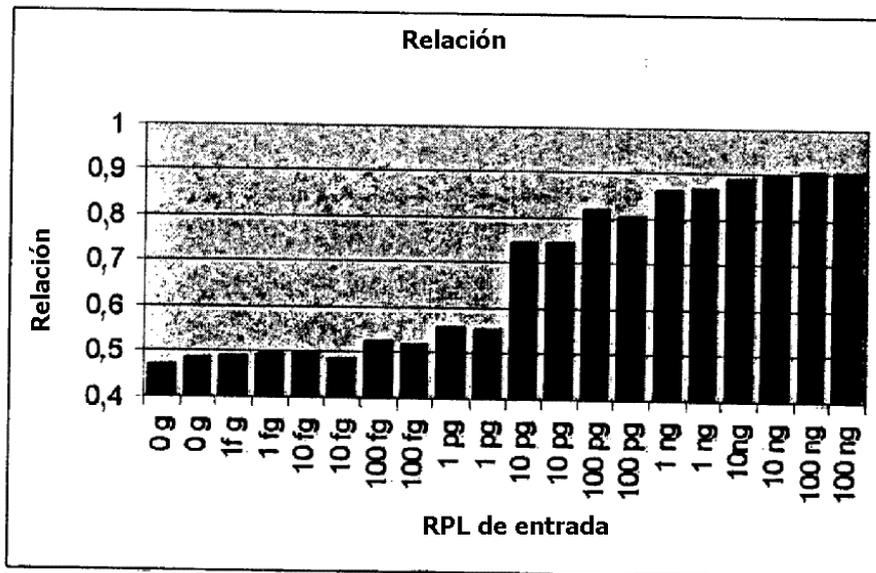


Figura 19A

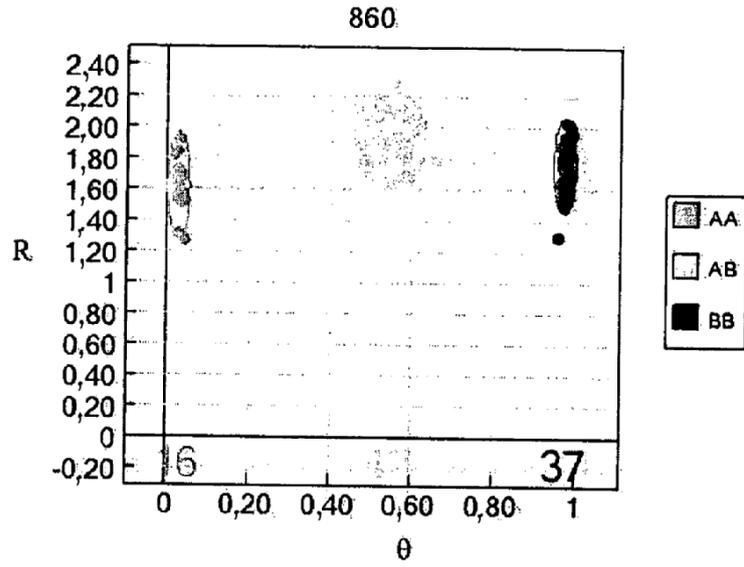


Figura 19B

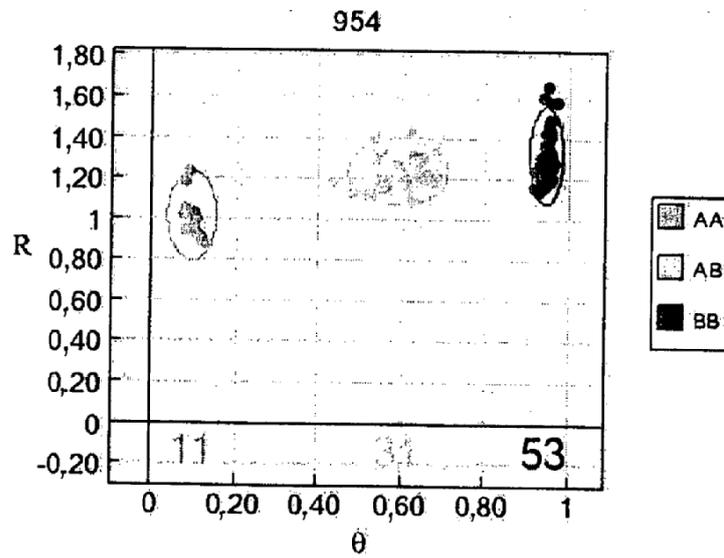


Figura 19C

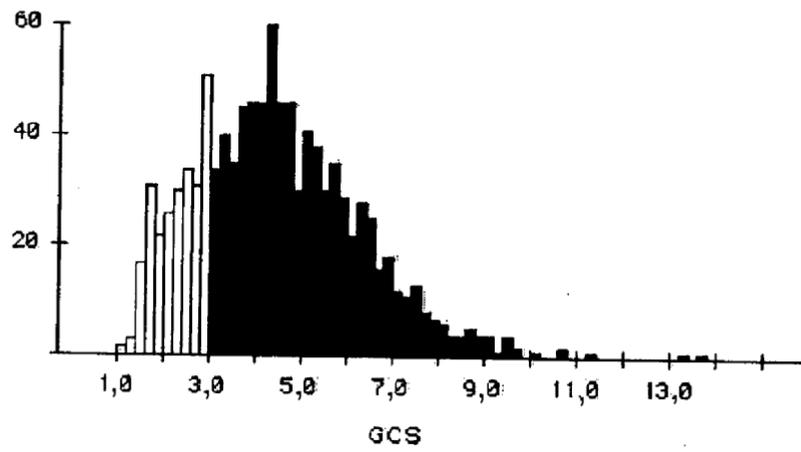


Figura 20

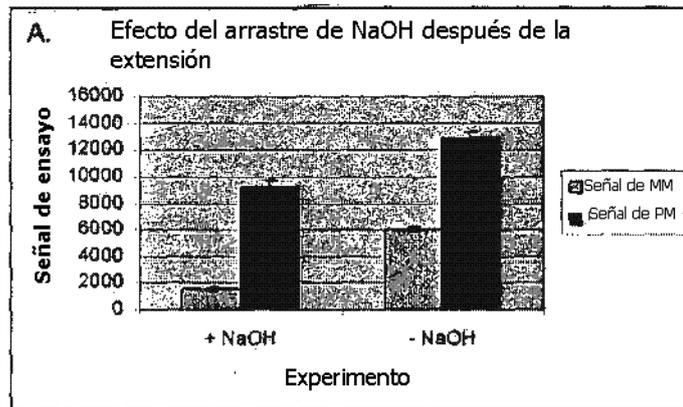


Figura 21

