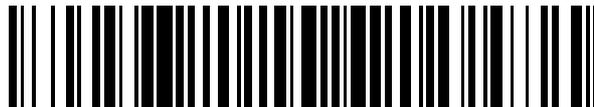


19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 522 912**

51 Int. Cl.:

**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

96 Fecha de presentación y número de la solicitud europea: **22.10.2008 E 08842936 (0)**

97 Fecha y número de publicación de la concesión europea: **10.09.2014 EP 2209913**

54 Título: **Métodos para el enriquecimiento de secuencias basadas en solución y análisis de regiones genómicas**

30 Prioridad:

**23.10.2007 EP 07020660**  
**20.08.2008 US 194574**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**19.11.2014**

73 Titular/es:

**F. HOFFMANN-LA ROCHE AG (100.0%)**  
**Grenzacherstrasse 124**  
**4070 Basel , CH**

72 Inventor/es:

**ALBERT, THOMAS y**  
**RODESCH, MATTHEW**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 522 912 T3**

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

## DESCRIPCIÓN

Métodos para el enriquecimiento de secuencias basadas en solución y análisis de regiones genómicas

## 5 CAMPO DE LA INVENCION

La presente solicitud se refiere al campo del enriquecimiento y análisis de secuencias de ácidos nucleicos por medio de la captura de dichas secuencias sobre un soporte sólido. Más precisamente, la presente invención proporciona un nuevo método para capturar regiones genómicas específicas para el posterior análisis adicional, si la región de interés es demasiado grande para ser amplificada por sólo una o unas pocas reacciones de PCR. En particular, la presente invención proporciona para el enriquecimiento de secuencias específicas en un formato basado en solución.

## 15 ANTECEDENTES DE LA INVENCION

El advenimiento de la tecnología de microchips de ácidos nucleicos, por ejemplo la tecnología de microchips de DNA, hace que sea posible construir una matriz de millones de secuencias de ácido nucleico, por ejemplo secuencias de DNA, en un área muy pequeña, por ejemplo en un portaobjetos de microscopio (por ejemplo, Patentes de Estados Unidos núm. 6.375.903 y 5.143.854). Inicialmente, dichas matrices se crearon mediante la detección de secuencias de DNA previamente sintetizadas sobre portaobjetos. Sin embargo, la construcción de sintetizadores de chips sin máscara (MAS) en los que la luz se utiliza para la síntesis directa de las secuencias de DNA, la dirección de la luz se realiza usando un dispositivo de microespejo digital (DMD) tal como se describe en la Patente de Estados Unidos N° 6.375.903 que ahora permite la síntesis in situ de secuencias de oligonucleótidos directamente en el portaobjetos.

Al utilizar un instrumento MAS, la selección de secuencias de oligonucleótidos o secuencias de DNA que se construirán en el microchip se encuentra bajo el control de un programa de tal manera que ahora es posible crear chips personalizados de forma individual en base a las necesidades particulares de un investigador. En general, la tecnología de síntesis de DNA o microchips de oligonucleótidos basados en MAS permite la síntesis en paralelo de millones de características únicas de oligonucleótidos en un área muy pequeña de un portaobjetos de microscopio estándar. Los microchips se sintetizan generalmente mediante el uso de la luz para dirigir qué oligonucleótidos son sintetizados en lugares específicos en un chip, estos lugares se llaman características.

Con la disponibilidad de los genomas completos de cientos de organismos, para los que una secuencia de referencia ha sido generalmente depositada en una base de datos pública, se han utilizado microchips para llevar a cabo análisis de secuencias de ácidos nucleicos o DNA aislados de una gran variedad de organismos.

La tecnología de ácido nucleico o de microchips de DNA se ha aplicado a muchas áreas de investigación y diagnóstico, como la expresión y el descubrimiento de genes, la detección de mutaciones, comparación de secuencias alélicas y evolutivas, la cartografía del genoma, el descubrimiento de fármacos, y mucho más. Muchas aplicaciones requieren la búsqueda de variantes genéticas y mutaciones a través de todo el genoma humano que subyacen a enfermedades humanas. En el caso de enfermedades complejas, estas búsquedas generalmente resultan en un polimorfismo de un solo nucleótido (SNP) o un conjunto de SNP asociados con enfermedades y / o riesgo de enfermedad. La identificación de estos SNP ha demostrado ser una tarea ardua y muchas veces infructuosa porque se requiere la resecuenciación de grandes regiones de DNA genómico de las personas afectadas o muestras de tejido, por lo general superior a 100 kilobases (Kb), para encontrar un solo cambio de base o para identificar todas las variantes de la secuencia. Otras aplicaciones incluyen la identificación de las ganancias y pérdidas de secuencias cromosómicas que también pueden estar asociadas con el cáncer, como el linfoma (Martínez-Climent JA et al., 2003, Blood 101: 3109 - 3117), el cáncer gástrico (Weiss MM et al., 2004, Cell. Oncol. 26: 307-317), el cáncer de mama (Callagy G et al., 2005, J. Path. 205: 388-396) y el cáncer de próstata (París, PL et al., 2004, Hum. Mol. Gen. 13: 1303-1313). Como tal, la tecnología de microchips es una herramienta tremendamente útil para los investigadores científicos y clínicos en su comprensión de las enfermedades y la eficacia del régimen terapéutico en el tratamiento de estas enfermedades.

El genoma normalmente es demasiado complejo para ser estudiado como un todo, y deben utilizarse técnicas para reducir la complejidad del genoma. Para abordar este problema, una solución es reducir ciertos tipos de secuencias abundantes de una muestra de ácido nucleico o DNA genómico, como se describe en la patente de EE.UU. 6.013.440. Una alternativa es utilizar métodos y composiciones para el enriquecimiento de las secuencias genómicas como se describe, por ejemplo; en Albert et al. (2007, Nat. Meth., 4: 903-5), Okou et al. (2007, Nat. Meth. 4: 907-9), Olson M. (2007, Nat. Meth. 4: 891-892), Hodges et al. (2007, Nat. Genet. 39: 1522-1527) y como se muestra en la Solicitud de Patente de Estados Unidos con números de serie 11/638004, 11/970949 y 61/032594. Albert et al. describe una alternativa que es a la vez rentable y rápida en la reducción efectiva de la complejidad de una muestra genómica de una manera definida por el usuario para permitir su posterior procesamiento y análisis. Lovett et al. (1991, Proc. Natl. Acad. Sci. 88: 9628-9632) también describe un método para la selección genómica utilizando cromosomas artificiales bacterianos. Sin embargo, los métodos existentes están limitados, por ejemplo, por su facilidad de uso y la inflexibilidad de los materiales y métodos.

La tecnología de microchips previa, ya sea la tecnología de enriquecimiento u otra, suele ser una tecnología asociada al sustrato con una variabilidad inherente, como portaobjetos de microchips, chips, y similares. La variabilidad puede tomar muchas formas, por ejemplo la variabilidad en el ruido de fondo, la cinética de hibridación / sonda, fuente de vidrio y similares. La variabilidad juega un papel importante en la interpretación experimental y puede definir el éxito o fracaso de un experimento.

Por lo tanto, lo que se necesita son métodos, sistemas y composiciones para proporcionar enriquecimiento de secuencias específicas en un formato de microchip que no sea el habitual de tipo sustrato. La aparición de nuevos formatos de microchips proporcionará herramientas adicionales para los investigadores y clínicos en el avance del conocimiento de las enfermedades y estados patológicos.

## RESUMEN DE LA INVENCION

La presente descripción proporciona métodos y sistemas para la captura y el enriquecimiento de ácidos nucleicos diana y el análisis de los ácidos nucleicos diana enriquecidos. En particular, la presente descripción proporciona el enriquecimiento de secuencias específicas en un formato basado en solución. Los métodos y sistemas de la presente descripción son útiles para ayudar a los investigadores y clínicos en la identificación, el estudio y seguimiento de regímenes de tratamiento asociados con la enfermedad y estados de la enfermedad.

La presente descripción se resume como un nuevo método para reducir la complejidad de una gran muestra de ácidos nucleicos, tal como una muestra genómica, biblioteca de DNAC o RNAm o biblioteca de RNAm para facilitar su posterior procesamiento y análisis genético. Las realizaciones de la presente descripción comprenden sondas de ácido nucleico inmovilizadas (pre-seleccionadas) o no inmovilizadas para capturar las secuencias diana de ácidos nucleicos a partir de, por ejemplo, una muestra genómica mediante hibridación de la muestra a las sondas, o amplicones derivados de sonda, sobre un soporte sólido o en solución. Los ácidos nucleicos diana capturados son preferiblemente lavados y eluidos de las sondas. Las secuencias genómicas eluidas son más susceptibles de análisis genético detallado que una muestra que no ha sido sometida a los métodos descritos en el presente documento. La presente descripción proporciona métodos y sistemas para la captura y el enriquecimiento de ácidos nucleicos diana y el análisis de los ácidos nucleicos diana enriquecidos. En algunas realizaciones, la presente descripción proporciona el enriquecimiento de secuencias específicas en un formato basado en solución. En algunas realizaciones, la presente descripción proporciona métodos y sistemas para la captura y el enriquecimiento de ácidos nucleicos diana basados en solución (por ejemplo, DNA genómico, RNA, DNAC, RNAm, etc.).

Los métodos descritos proporcionan una aproximación rentable, flexible y eficiente para reducir la complejidad de una muestra genómica. Las muestras genómicas se usan aquí con fines descriptivos, pero se entiende que otras muestras no genómicas, también grandes muestras no genómicas, podrían someterse a los mismos procedimientos. Los métodos y sistemas descritos en el presente documento proporcionan el enriquecimiento de las secuencias diana en un enfoque basado en la solución proporcionando de este modo una alternativa a los métodos de microchips basados en sustrato para el uso en la investigación y terapéutica asociada con la enfermedad y estados de enfermedad tales como cáncer (Durkin et al., 2008, Proc. Natl. Acad. Sci. 105: 246-251; Natrajan et al., 2007, Genes, Chr. and Cancer 46: 607-615; Kim et al., 2006, Cell 125: 1269-1281; Stallings et al., 2006 Can. Res. 66: 3673 a 3680), trastornos genéticos (Balciuniene et al., Am. J. Hum. Genet. En prensa), enfermedades mentales (Walsh et al., 2008, Science 320: 539-543; Roohi et al., 2008, J. Med. Genet. Epub 18 de marzo 2008; Sharp et al., 2008, Nat. Genet. 40: 322-328; Kumar et al., 2008, Hum. Mol. Genet. 17: 628-638;) e investigación básica y evolutiva (Lee et al., 2008, Hum. Mol. Gen. 17: 1127-1136; Jones et al., 2007, BMC Genomics 8:402; Egan et al., 2007, Nat. Genet. 39: 1384-1389; Levy et al., 2007, PLoS Biol. 5: e254; Ballif et al., 2007, Nat. Genet. 39: 1071-1073; Scherer et al., 2007, Nat. Genet. S7-S15; Feuk et al., 2006, Nat. Rev. Genet. 7: 85-97), por nombrar unos pocos.

En una realización, la presente descripción proporciona métodos para reducir la complejidad genética de una población de moléculas de ácido nucleico, comprendiendo el método los pasos de exponer las moléculas fragmentadas, desnaturalizadas de ácido nucleico de dicha población a múltiples, diferentes sondas de oligonucleótidos en condiciones de hibridación seguido por la unión de los complejos de moléculas hibridadas a un soporte sólido para capturar las moléculas de ácido nucleico que hibridan específicamente con dichas sondas, en los que dichas moléculas de ácido nucleico desnaturalizados fragmentados tienen un tamaño medio de aproximadamente 100 a aproximadamente 1000 residuos de nucleótidos, preferiblemente entre aproximadamente 250 a aproximadamente 800 residuos de nucleótidos y lo más preferiblemente entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos, separando los ácidos nucleicos no unidos hibridados de forma inespecífica de las moléculas capturadas, eluyendo las moléculas capturadas, y opcionalmente repitiendo los procesos antes mencionados durante al menos un ciclo más con las moléculas capturadas eluidas.

En otra realización de la presente descripción, los métodos de captura basados en solución comprenden amplicones derivados de sondas en los que dichas sondas para la amplificación se fijan a un soporte sólido. El soporte sólido comprende sondas de ácido nucleico inmovilizadas en el soporte para capturar secuencias de ácido nucleico específicas (por ejemplo, ácidos nucleicos diana) de, por ejemplo, una muestra genómica. La amplificación de la sonda proporciona amplicones de sonda en solución que se hibridan a secuencias diana. Después de la hibridación

de los amplicones de la sonda a secuencias diana, las secuencias de ácidos nucleicos diana presentes en la muestra están enriquecidas mediante la captura (por ejemplo, a través de un enlazante químico tal como biotina, digoxigenina, etc.) y el lavado de las sondas y la elución de los ácidos nucleicos diana hibridados de las sondas capturadas (Figura 6). La(s) secuencia(s) de ácido nucleico diana puede(n) amplificarse adicionalmente usando, por ejemplo, la PCR mediada por ligación inespecífica (LM-PCR), resultando en un grupo de productos de PCR amplificados de complejidad reducida en comparación con la muestra diana original.

En algunas realizaciones, la hibridación entre las sondas y los ácidos nucleicos diana se realiza en condiciones preferiblemente rigurosas suficientes para soportar la hibridación entre los amplicones de la sonda basados en solución, en el que dichas sondas comprenden el enlazante químico y las regiones complementarias de la muestra de ácido nucleico diana para proporcionar complejos de hibridación sonda / diana. Los complejos se capturan posteriormente a través del enlazante químico y se lavan en condiciones suficientes para eliminar los ácidos nucleicos unidos de manera inespecífica y las secuencias de ácidos nucleicos diana hibridadas se eluyen de los complejos sonda / diana capturados.

En algunas realizaciones, las múltiples sondas de oligonucleótidos diferentes, comprenden un grupo químico o enlazante químico, por ejemplo, una porción de unión tal como biotina, digoxigenina, etc., que es capaz de unirse a un soporte sólido. El soporte sólido para la unión comprende la molécula de captura correspondiente, por ejemplo, estreptavidina para biotina y anticuerpo anti-digoxigenina para la digoxigenina. Un experto en la materia reconocerá que la presente descripción no está limitada por el enlazante químico utilizado y otras moléculas enlazantes alternativas son igualmente susceptibles a los métodos y sistemas de la presente invención.

La población o pluralidad de moléculas de ácido nucleico diana contienen preferiblemente todo el genoma o al menos un cromosoma de un organismo o al menos una molécula de ácido nucleico con al menos aproximadamente 100 kb. En particular, el(los) tamaño(s) de la(s) molécula(s) de ácido nucleico es(son) al menos aproximadamente 200 kb, al menos aproximadamente 500 kb, al menos aproximadamente 1 Mb, al menos aproximadamente 2 Mb o al menos aproximadamente 5 Mb, especialmente un tamaño entre aproximadamente 100 kb y aproximadamente 5 Mb, entre aproximadamente 200 kb y aproximadamente 5 Mb, entre aproximadamente 500 kb y aproximadamente 5 Mb, entre aproximadamente de 1 Mb y aproximadamente 2 Mb o entre aproximadamente 2 Mb y aproximadamente 5 Mb.

En algunas realizaciones, las moléculas de ácidos nucleicos diana se seleccionan de un animal, una planta o un microorganismo, en realizaciones preferidas, el organismo es un ser humano. Si sólo están disponibles muestras limitadas de ácidos nucleicos (por ejemplo, del genoma humano), los ácidos nucleicos se pueden amplificar, por ejemplo, mediante amplificación del genoma entero, antes de practicar los métodos de la presente invención. Puede ser necesario antes de la amplificación realizar el método de la invención, por ejemplo, para fines forenses (por ejemplo, en medicina forense para los propósitos de identidad genética).

En algunas realizaciones, la población o pluralidad de moléculas de ácido nucleico diana es una población de moléculas de DNA genómico. Las sondas se pueden seleccionar de una pluralidad de sondas o secuencias que, por ejemplo, definen una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos, una pluralidad de sondas que definen la secuencia completa de al menos un locus genético único, dicho locus tiene un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb, o al menos uno de los tamaños como se especifica anteriormente, una pluralidad de sondas que definen polimorfismos de nucleótido único (SNP), o una pluralidad de sondas que definen un chip, por ejemplo, un microchip de secuencia contigua diseñado para capturar la secuencia completa de al menos un cromosoma completo.

En algunas realizaciones, el soporte sólido es o bien un microchip de ácido nucleico o una población de cuentas.

En algunas realizaciones, la presente invención comprende el paso de ligación de moléculas adaptadoras para uno o ambos, preferiblemente ambos extremos de las moléculas de ácido nucleico antes o después de la exposición de muestras de ácidos nucleicos fragmentados a las sondas para la hibridación.

En algunas realizaciones, los métodos comprenden además la amplificación de las moléculas de ácido nucleico diana con al menos un cebador, dicho cebador comprende una secuencia que se hibrida específicamente con la secuencia de dicha molécula adaptadora.

En algunas realizaciones, las secuencias de ácido nucleico diana amplificado puede secuenciarse, hibridarse a una resecuenciación o a al chip de SNP y la secuencia o genotipos pueden ser analizados.

En algunas realizaciones, la presente descripción proporciona un método de enriquecimiento de secuencias de ácidos nucleicos diana en una muestra genómica, tales como exones o variantes, preferiblemente sitios de SNP. Esto se puede lograr mediante la programación de sondas genómicas específicas para una región del genoma a sintetizar en un microchip o mediante la síntesis de sondas genómicas específicas para una región del genoma para capturar secuencias de ácido nucleico diana complementaria contenidas en una muestra genómica compleja.

En algunas realizaciones, la presente descripción está dirigida a un método para determinar la información de secuencia de un ácido nucleico de alrededor al menos una región de ácido nucleico, en particular el ácido nucleico genómico, por ejemplo, todo el genoma o al menos un cromosoma, por ejemplo con un tamaño tal como se especifica anteriormente, específicamente en una muestra, comprendiendo el método los pasos de realizar el método como se describe anteriormente y determinar la secuencia de ácido nucleico de las moléculas capturadas (eluidas), en particular mediante la realización de secuenciación mediante reacciones de síntesis.

En algunas realizaciones, la presente descripción se refiere a un método para detectar la variación en la región codificante relativa a un genoma de referencia, en particular, en relación a un genoma de referencia que comprende, moléculas de ácido nucleico genómico desnaturalizado fragmentado, el método tal como se describió previamente, comprende además la determinación de la secuencia de ácido nucleico de las moléculas diana capturadas (y eluidas), en particular mediante la realización de la secuenciación mediante reacciones de síntesis y la comparación de la secuencia determinada con una secuencia de una base de datos, en particular, con una secuencia en una base de datos de polimorfismos en el genoma de referencia para identificar las variantes del genoma de referencia.

En algunas realizaciones, la presente descripción se refiere a un equipo que comprende un soporte sólido y los reactivos para realizar un método según la presente invención. Dicho equipo puede comprender una molécula adaptadora de doble cadena, y un soporte sólido con múltiples, diferentes sondas de oligonucleótidos, en donde las sondas se seleccionan de entre una pluralidad de sondas que definen una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos, una pluralidad de sondas que definen la secuencia completa de al menos un locus genético único, dicho locus tiene un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb, o al menos uno de los tamaños como se ha especificado anteriormente, una pluralidad de sondas que definen los sitios conocidos por contener SNP, o una pluralidad de sondas que definen un microchip de secuencia contigua diseñado para capturar la secuencia completa de al menos un cromosoma completo. Preferiblemente, el equipo comprende dos moléculas adaptadoras bicatenarias diferentes (A y B). El soporte sólido es de nuevo una pluralidad de cuentas o un microchip. El equipo puede comprender además al menos uno o más de otros componentes seleccionados a partir de DNA polimerasa, polinucleótido quinasa de T4, DNA ligasa de T4, una solución de hibridación de chip, una solución de lavado de chip, y / o una solución de elución de chip.

En otras realizaciones, la presente descripción se refiere a un equipo que comprende composiciones y reactivos para realizar un método según la presente invención. Dicho equipo puede comprender, pero no se limita a, una molécula adaptadora de cadena doble, sondas múltiples de oligonucleótidos diferentes, un soporte sólido para la captura de dichas sondas, donde las sondas se seleccionan de entre una pluralidad de secuencias que definen una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos, una pluralidad de sondas que definen la secuencia completa de al menos un locus genético único, un locus que tiene un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb, o al menos uno de los citados tamaños como se ha especificado anteriormente, una pluralidad de sondas que definen los sitios que se sabe contienen SNP, o una pluralidad de sondas que definen un microchip de secuencia contigua diseñado para capturar la secuencia completa de al menos un cromosoma completo. En algunas formas de realización, un equipo comprende una pluralidad de cuentas o un sustrato de microchips (por ejemplo, un portaobjetos, un chip, etc). En algunas formas de realización, un equipo comprende dos moléculas adaptadoras de doble cadena diferente. Un equipo puede comprender además al menos uno o más de otros componentes seleccionados a partir de DNA polimerasa, polinucleótido quinasa de T4, DNA ligasa de T4, la solución de hibridación, la solución de lavado, y / o solución de elución.

En algunas realizaciones, las sondas de captura de ácido nucleico (previamente seleccionados) se inmovilizan sobre un soporte sólido (por ejemplo, portaobjetos, chip, cuentas, etc.) utilizando cualquier número de métodos reconocidos (por ejemplo, deposición de muestras, fotolitografía, síntesis in situ, etc.). En realizaciones preferidas, las sondas se sintetizan in situ mediante síntesis de chip sin máscara sobre un sustrato y posteriormente amplificadas mediante, por ejemplo, PCR resultando en amplicones derivados de la sonda en solución. En algunas realizaciones, las secuencias de la sonda tal como se sintetiza comprenden sitios de unión del cebador para la amplificación en uno o ambos extremos 3' y 5' terminales (por ejemplo, en o cerca de los extremos) de las sondas. En algunas realizaciones, la secuencia de los sitios de unión al cebador en las sondas son los mismos tanto en el extremo de cebamiento 3' como 5' o las sondas, mientras que en otras realizaciones, la secuencia de los sitios de unión del cebador es diferente al primer extremo 3' que la secuencia en el extremo de cebamiento 5'. En algunas realizaciones, los cebadores de amplificación para la amplificación de la sonda comprenden además un sitio de endonucleasa de restricción, por ejemplo un sitio de MlyI para facilitar la extracción de secuencias de los cebadores de la diana final capturada, en el que uno de los cebadores (por ejemplo, el cebador directo o inverso) comprende además un enlazante químico tal como una porción de unión o de secuencia (por ejemplo, biotina, digoxigenina, cola HIS, etc) y se depositan sobre el soporte con las sondas inmovilizadas junto con los reactivos necesarios para la amplificación PCR exponencial (por ejemplo, procedimientos de PCR para la amplificación exponencial de las dianas tal como los conoce un experto en la materia). La PCR se realiza creando de este modo amplicones de secuencias de sonda de captura de tal manera que una de las hebras comprende el enlazante químico, como una porción o secuencia de unión. La solución que contiene el amplicón se transfiere a un recipiente (por ejemplo, tubo, pocillo de una placa de 96 pocillos, etc.) y en algunas realizaciones, se purifica a partir de componentes de la reacción. Una nueva ronda de amplificación se lleva a cabo preferentemente en los amplicones derivados de la sonda mediante PCR asimétrica, en el que el cebador enlazante químico marcado está en abundancia en comparación con el

cebador no marcado para sintetizar preferentemente amplicones marcados con fracción / secuencia de unión monocatenaria. Los amplicones se purifican lejos de los componentes de reacción y se transfieren a un recipiente, se añade muestra de ácido nucleico desnaturalizado, y se permite que se produzca la hibridación.

5 Tras la hibridación, se capturan los complejos de ácido nucleico / amplicón marcados. Por ejemplo, cuando la biotina es la porción de unión se utiliza un sustrato revestido de estreptavidina (SA) tal como cuentas recubiertas de SA (por ejemplo, cuentas / partículas paramagnéticas) para capturar el complejo amplicón / diana marcado con biotina. El complejo unido a la SA se lava y los ácidos nucleicos diana hibridados se eluyen desde el complejo y se utilizan en aplicaciones posteriores, tales como aplicaciones de secuenciación.

10 En algunas realizaciones, la presente descripción proporciona métodos para aislar y reducir la complejidad de una pluralidad de secuencias de ácido nucleico que comprende proporcionar un soporte sólido en el que dicho soporte sólido comprende sondas de hibridación hibridables con secuencias diana de ácido nucleico y proporcionar una muestra de ácido nucleico fragmentado que comprende secuencias de ácido nucleico diana, la amplificar las sondas de hibridación en el que los productos de amplificación comprenden una porción de unión y en el que los productos de amplificación están en solución, hibridar la muestra de ácido nucleico a los productos de amplificación en solución bajo condiciones tales que la hibridación entre los productos de amplificación y secuencias de ácido nucleico diana se permite que se produzca, separar los complejos de producto de amplificación / secuencias de ácido nucleico diana hibridado de los ácidos nucleicos no hibridados específicamente por dicha porción de unión, y eluyendo las secuencias de ácido nucleico diana hibridado del complejo de esta manera aislando y reduciendo la complejidad de una pluralidad de secuencias de ácido nucleico. En algunas realizaciones, las secuencias de ácidos nucleicos diana eluidas se secuencian. En algunas realizaciones, el soporte sólido es un portaobjetos de microchip. En algunas realizaciones, la muestra de ácido nucleico diana es DNA genómico fragmentado con o sin moléculas adaptadoras en uno o ambos extremos de los fragmentos. En algunas realizaciones, las sondas de hibridación comprenden un sitio de endonucleasa de restricción, por ejemplo un sitio MlyI. En algunas realizaciones, la sonda de amplificación comprende la reacción en cadena de polimerasa exponencial, y puede comprender además amplificación no exponencial asimétrica. En algunas realizaciones, la porción de unión es biotina y el sustrato de captura, como una cuenta, por ejemplo, una partícula paramagnética, está recubierta con estreptavidina para la separación del complejo de producto de amplificación / ácido nucleico diana de los ácidos nucleicos diana hibridados de forma no específica. En algunas formas de realización, los complejos de productos de amplificación / ácido nucleico diana capturado se lavan antes de la elución de los ácidos nucleicos diana unidos.

En algunas realizaciones, la presente descripción proporciona un equipo que comprende secuencias de sonda de hibridación que comprende una porción de unión y un sitio de enzima de restricción en el que dichas secuencias de sondas se diseñan para hibridarse a una o más secuencias diana de ácido nucleico y en el que dichas secuencias de sonda están en solución, un sustrato que comprende una pareja de unión para la unión de dicha porción de unión, y la instrucción para realizar los métodos de la presente invención. En algunas formas de realización, un equipo comprende además una o más soluciones tales como la(s) solución(es) de hibridación, lavado, y de elución. En algunas formas de realización, un equipo comprende un imán. En algunas formas de realización, un equipo comprende una o más enzimas y reactivos correspondientes, tampones, y similares, por ejemplo, una enzima de restricción tal como MlyI y tampones / reactivos para llevar a cabo reacciones de enzimas de restricción utilizando MlyI.

45 Otros objetos, ventajas y características de la presente invención se harán evidentes a partir del conjunto de la siguiente especificación tomada con los figuras adjuntas.

#### FIGURAS

50 Fig. 1 es una representación gráfica general del diagrama de flujo de un proceso de selección genómico directo utilizando un microchip.

Fig. 2 es otra representación gráfica de diagrama de flujo de un proceso de selección genómico directo utilizando un microchip.

55 La figura 3 (a-b) muestra los resultados de un proceso de selección genómico directo utilizando un microchip de acuerdo con el Ejemplo 2. (a) Detalle del mapa de lectura de secuencia ~190Kb del cromosoma 16 de tres réplicas del microchip de selección genómica, lo que indica la reproducibilidad de la secuenciación dirigida. El DNA genómico de una línea celular de linfoma de Burkett se purificó y se fragmentó. Los exones del programa de secuenciación de tumores (6726 regiones del genoma de 500 pb de tamaño), fueron capturados utilizando un microchip de oligonucleótidos NimbleGen y se secuenciaron utilizando un secuenciador 454. (1) Posición del cromosoma, (2, 3, 4) mapa de lectura de la mayor puntuación BLAST para las lecturas del secuenciador 454 de la selección de tres microchips independientes y experimentos de secuenciación (5) regiones seleccionadas mediante sondas de microchips. (b) detalle de mapa de lectura de la secuencia ~2,000 bases de un cromosoma 17 entre una selección de microchips de una región contigua de 2 Mb que contiene el gen BRCA1. (1) Posición de cromosoma, (2) sondas de selección de microchips. Las sondas están espaciadas cada 10pb y escalonadas a lo largo del eje y. (3)

Cobertura de secuencia por base. La cobertura es de 0 a 100 veces. (4) Mapa de lectura de las puntuaciones BLAST más altas para las lecturas del secuenciador 454.

Fig. 4 (a-c) muestra los resultados de la síntesis de sondas en un microchip, la liberación de las sondas del microchip y la inmovilización de las sondas sobre un soporte para su uso en un método para la captura de polinucleótidos diana de interés. (a) Comparación de la profundidad de cobertura para la selección y secuenciación "Exónico" y "Locus" como se describe en el Ejemplo 2. El gráfico muestra la fracción de bases de cada región diana agregada y la correspondiente profundidad de cobertura de secuencia acumulada después de un pase con 454-FLX. La Muestra "Exónica" representa 6726 regiones de tamaño exón. La región 2 Mb BRCA1 fue seleccionada desde las posiciones 37.490.417 a 39.490.417 sobre el cromosoma humano 17. Sólo la fracción única fue seleccionada con las sondas de selección. (b) Histograma de profundidad de cobertura por base de secuencia para el experimento Exónico, como se describe en el Ejemplo 2 (c) Histograma de profundidad de cobertura por base para el ejemplo 2Mb Locus de acuerdo con el Ejemplo 3.

Fig. La figura 5 ilustra un detalle del mapa de lectura para un locus sobre el cromosoma 16 a partir de tres muestras genómicas. Los datos fueron generados por la secuenciación selectiva de 6726 exones que fueron capturados en solución. Los oligonucleótidos de captura se escindieron y se amplificaron a partir de un microchip, utilizando el protocolo descrito en el Ejemplo 4. Los datos presentados representan un ejemplo de mapa genético del cromosoma 3. (1) Posición de cromosoma, (2) mapa de lecturas de secuenciación de un pase de secuenciación con 454-FLX, y (3) regiones diana. El análisis de los datos de captura en fase de solución indica que el 83,8% de las lecturas mapean de nuevo las regiones diana, lo que indica un rendimiento similar a los protocolos de captura basados en chips.

Fig. 6 ejemplifica una realización de la presente invención; un diagrama de flujo generalizado de un proceso de enriquecimiento en el que dichos métodos de enriquecimiento se utilizan para aislar y enriquecer para una pluralidad de secuencias de ácidos nucleicos en una solución acuosa. Las sondas de hibridación, fijadas a un sustrato de microchips, se amplifican in situ para producir amplicones derivados de sonda en solución de los amplicones que comprenden una porción de unión. Los ácidos nucleicos fragmentados (por ejemplo, marcados con una porción de detección) se hibridaron en solución con los amplicones de sonda marcados, los complejos se capturaron posteriormente (por ejemplo, mediante partículas paramagnéticas de captura). Los complejos hibridados capturados e inmovilizados se lavan y las dianas unidas específicamente se eluyen de los amplicones sonda inmovilizados unidos. Las secuencias diana eluidas (por ejemplo, aisladas y enriquecidas) se aplican a experimentos posteriores, como la secuenciación.

Fig. 7 demuestra la conformidad de la resecuenciación usando los métodos de captura en solución de la presente invención. El ensayo de resecuenciación se compone de un subconjunto de las regiones diana capturadas. El eje x representa un conjunto arbitrario de las regiones de la región diana más grande, que sirve como una representación de la región de captura diana como un todo. El eje y representa el porcentaje de conformidad de secuencia con las secuencias diana conocidas.

## DEFINICIONES

Tal como se utiliza en este documento, el término "muestra" se utiliza en su sentido más amplio. En un sentido, se pretende incluir una muestra o cultivo obtenido de cualquier fuente, preferiblemente de una fuente biológica. Las muestras biológicas pueden obtenerse a partir de animales (incluidos los seres humanos) y abarcan: líquidos, sólidos, tejidos y gases. Las muestras biológicas incluyen productos sanguíneos, tales como plasma, suero y similares. Como tal, una "muestra de ácidos nucleicos" o una "muestra de ácido nucleico", una "muestra diana" comprende ácidos nucleicos (por ejemplo, DNA, RNA, DNAc, RNAm, RNAt, RNAmi, etc.) de cualquier fuente. En la presente solicitud, una muestra de ácido nucleico deriva preferiblemente de una fuente biológica, tal como una célula, tejido y similares de origen humano o no humano. El término "no humano" se refiere a todos los animales no humanos y entidades que incluyen, pero no se limitan a, vertebrados tales como roedores, primates no humanos, ovinos, bovinos, rumiantes, lagomorfos, porcinos, caprinos, equinos, caninos, felinos, aves, etc. No humanos también incluye los invertebrados y organismos procariontes tales como bacterias, plantas, levaduras, virus, y similares. Como tal, una muestra de ácido nucleico usada en los métodos y sistemas de la presente invención es una muestra de ácido nucleico derivado de cualquier organismo, ya sea eucariota o procarionte.

Tal como se utiliza aquí, el término "hibridación" se usa en referencia al emparejamiento de ácidos nucleicos complementarios. La hibridación y la fuerza de hibridación (por ejemplo, la fuerza de la asociación entre los ácidos nucleicos) se ve afectada por factores tales como el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones, la temperatura de fusión ( $T_m$ ) del híbrido formado, y la relación G:C de los ácidos nucleicos. Aunque la invención no se limita a un conjunto particular de condiciones de hibridación, preferiblemente se emplean condiciones de hibridación rigurosas. Las condiciones de hibridación rigurosas son dependientes de la secuencia y difieren al variar los parámetros ambientales (por ejemplo, concentraciones de sal, presencia de materiales orgánicos, etc.). Generalmente, se seleccionan las condiciones "rigurosas" entre alrededor de 50 °C a aproximadamente 20 °C por debajo de la  $T_m$  para la secuencia de ácido nucleico específica a una fuerza iónica y pH definidos. Preferiblemente, las condiciones rigurosas son entre alrededor de 5 °C a 10 °C por debajo del punto de

fusión térmico para un ácido nucleico específico unido a un ácido nucleico complementario. La  $T_m$  es la temperatura (bajo fuerza iónica y pH definidos) a la cual el 50% de un ácido nucleico (por ejemplo, ácido nucleico diana) se hibrida a una sonda perfectamente coincidente.

5 "Condiciones rigurosas" o "condiciones de alta rigurosidad", por ejemplo, puede ser la hibridación en formamida al 50%, 5x SSC (NaCl 0,75 M, citrato de sodio 0,075 M), 50 mM de fosfato de sodio (pH 6,8), 0,1% de pirofosfato de sodio, 5x solución de Denhardt, DNA de esperma de salmón sonificado (50 mg / ml), 0,1% de SDS, y 10% de sulfato de dextrano a 42 °C, con lavados a 42 °C en 0,2% SSC (cloruro sódico / citrato sódico) y 50 % de formamida a 55 °C, seguido de un lavado con 0,1 x SSC que contiene EDTA a 55 °C. A modo de ejemplo, pero no de limitación, se  
10 contempla que los tampones que contienen 35% de formamida, 5x SSC, y 0,1% (p / v) de dodecil sulfato de sodio (SDS) son adecuados para hibridarse bajo condiciones moderadamente no rigurosas a 45 °C durante 16-72 horas.

Además, se contempla que la concentración de formamida se puede ajustar adecuadamente entre un rango de 20-45% dependiendo de la longitud de la sonda y el nivel de rigurosidad deseado. Ejemplos adicionales de las condiciones de hibridación se proporcionan en varias fuentes, incluyendo Molecular Cloning: A Laboratory Manual, Eds. Sambrook et al., Cold Spring Harbour press.  
15

De manera similar, las condiciones de lavado "rigurosas" se determinan empíricamente de forma ordinaria para la hibridación de una diana con una sonda, o en la presente invención, un amplicón derivado de sonda. El amplicón / diana se hibridan (por ejemplo, bajo condiciones de hibridación rigurosas) y después se lavan con tampones que contienen concentraciones de sales sucesivamente más bajas, o mayores concentraciones de detergentes, o a temperaturas incrementales hasta que la relación de señal a ruido para la hibridación inespecífica es lo suficientemente alta para facilitar la detección de la hibridación específica. Las condiciones de temperatura rigurosas generalmente incluirán temperaturas en exceso de aproximadamente 30 °C, más usualmente en exceso de  
20 aproximadamente 37 °C, y ocasionalmente en exceso de aproximadamente 45 °C. Las condiciones rigurosas de sal serán normalmente inferiores a aproximadamente 1000 mM, habitualmente inferiores a aproximadamente 500 mM, más habitualmente inferiores a aproximadamente 150 mM (Wetmur et al, 1966, J. Mol. Biol., 31: 349-370; Wetmur, 1991, Critical Reviews in Biochemistry and Molecular Biology, 26: 227-259.  
25

Tal como se utiliza aquí, el término "cebador" se refiere a un oligonucleótido, ya sea de origen natural como en una digestión de restricción purificada o producido sintéticamente, que es capaz de actuar como un punto de iniciación de síntesis cuando se coloca bajo condiciones en las que se induce la síntesis de un producto de extensión del cebador que es complementario a una cadena de ácido nucleico, (por ejemplo, en presencia de nucleótidos y un agente inductor tal como DNA polimerasa y a una temperatura y pH adecuados). El cebador es preferentemente de  
30 cadena sencilla para una máxima eficiencia en la amplificación. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluyendo la temperatura, la fuente del cebador y el uso del método.  
35

Tal como se utiliza aquí, el término "sonda" se refiere a un oligonucleótido (por ejemplo, una secuencia de nucleótidos), ya sea de origen natural como en una digestión de restricción purificada o producido sintéticamente, de forma recombinante o mediante amplificación por PCR, que es capaz de hibridar con al menos una parte de otro oligonucleótido de interés, por ejemplo, secuencias de ácido nucleico diana. Una sonda puede ser de cadena sencilla o de cadena doble. Las sondas son útiles en la detección, identificación y aislamiento de secuencias génicas  
40 particulares.  
45

Tal como se usa aquí, el término "moléculas de ácido nucleico diana" y "secuencias de ácido nucleico diana" se usan indistintamente y se refieren a moléculas o secuencias de una región genómica diana a estudiar. Las sondas preseleccionadas determinan el rango de moléculas de ácido nucleico dianas. Por lo tanto, la "diana" se busca entre otras secuencias de ácidos nucleicos. Un "segmento" se define como una región de ácido nucleico dentro de la secuencia diana, como es un "fragmento" o una "porción" de una secuencia de ácido nucleico.  
50

Tal como se utiliza aquí, el término "aislado" cuando se usa en relación con un ácido nucleico, como en "aislar un ácido nucleico" se refiere a una secuencia de ácido nucleico que se identifica y separa de al menos un componente o contaminante con la que se asocia habitualmente en su fuente natural. Los ácidos nucleicos aislados está en una forma o configuración que es diferente de aquel en el que se encuentra en la naturaleza. Por contra, los ácidos nucleicos no aislados como los ácidos nucleicos tales como el DNA y RNA que se encuentran en el estado en que existen en la naturaleza. El ácido nucleico aislado, oligonucleótido o polinucleótido puede estar presente en forma de una sola hebra o de doble hebra.  
55

## 60 DESCRIPCIÓN DETALLADA DE LA INVENCION

La presente invención se refiere en general a métodos rentables, flexibles y rápidos para reducir la complejidad de las muestras de ácido nucleico para enriquecer los ácidos nucleicos diana de interés y para facilitar su posterior procesamiento y análisis, como la secuenciación, resecuenciación y SNP. Las secuencias de ácidos nucleicos diana capturadas, que son de una población genómica más definida menos compleja son más susceptibles de análisis  
65

genético detallado, por ejemplo, análisis genético de enfermedades y estados de enfermedad (por ejemplo, cáncer, mutaciones genéticas, enfermedades genéticas, etc.). La presente descripción proporciona métodos y sistemas útiles, por ejemplo, en la búsqueda de variantes genéticas y mutaciones tales como polimorfismos de nucleótidos únicos (SNP), conjuntos de SNP, inserciones genómicas, deleciones, etc., que subyacen en las enfermedades humanas.

Por lo tanto, la descripción proporciona para el método de enriquecimiento de ácido nucleico diana en una muestra de ácido nucleico compleja.

En una realización, se describe un método para reducir la complejidad genética de una población de moléculas de ácido nucleico, comprendiendo el método los pasos de exponer moléculas fragmentadas, desnaturalizadas de ácido nucleico de dicha población a múltiples sondas de oligonucleótidos diferentes que se unen sobre un soporte sólido en condiciones de hibridación para capturar moléculas de ácido nucleico que hibridan específicamente con dichas sondas o exponiendo las moléculas de ácido nucleico fragmentadas, desnaturalizadas de dicha población a múltiples sondas de oligonucleótidos diferentes en condiciones de hibridación, seguido por la unión de los complejos de moléculas hibridadas sobre un soporte sólido para capturar moléculas de ácido nucleico que hibridan específicamente con dichas sondas, en donde dichas moléculas fragmentadas, desnaturalizadas de ácido nucleico tienen un tamaño medio de aproximadamente 100 a aproximadamente 1000 residuos de nucleótidos, preferiblemente entre aproximadamente 250 a aproximadamente 800 residuos de nucleótidos y lo más preferiblemente entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos, separando los ácidos nucleicos no unidos e hibridados inespecíficamente de las moléculas capturadas; eluyendo las moléculas capturadas desde el soporte sólido, y opcionalmente repetir los pasos (a) a (c) durante al menos un ciclo más con las moléculas capturadas eluidas.

En una realización adicional, se describe un equipo que comprende moléculas adaptadoras de doble cadena, y un soporte sólido con múltiples sondas de oligonucleótidos diferentes, en el que dichas sondas se seleccionan de: una pluralidad de sondas que definen una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos, una pluralidad de sondas que definen la secuencia completa de al menos un locus genético único, dicho locus que tiene un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb, o al menos uno de los tamaños especificados anteriormente, una pluralidad de sondas que definen los sitios que se sabe contienen SNP, o una pluralidad de sondas que definen un microchip de secuencia contigua diseñado para capturar la secuencia completa de al menos un cromosoma completo.

En una realización, una muestra que contiene moléculas de ácido nucleico desnaturalizado (es decir, de una sola cadena), preferiblemente moléculas de ácido nucleico genómico, que pueden ser moléculas fragmentadas, se expone en condiciones de hibridación a una pluralidad de sondas de oligonucleótidos, que se inmovilizan en un soporte sólido antes o después de la hibridación con una pluralidad de sondas de oligonucleótidos para capturar a partir de las moléculas de ácidos nucleicos diana de la muestra que se hibridan con las sondas inmovilizadas. Las regiones no hibrida del genoma o cualquier otro ácido nucleico de la muestra permanecen en solución.

Los ácidos nucleicos son típicamente ácidos desoxirribonucleicos o ácidos ribonucleicos, e incluyen productos sintetizados in vitro mediante la conversión de un tipo de molécula de ácido nucleico (por ejemplo, DNA, RNA y DNAC) en otro, así como moléculas sintéticas que contienen análogos de nucleótidos, tales como los PNA. Moléculas de DNA genómico desnaturalizadas son en particular las moléculas derivadas del genoma que son más cortas que las moléculas de ácido nucleico genómico que aparecen de forma natural. El experto en la materia puede producir moléculas de tamaño aleatorio o no aleatorio a partir de moléculas más grandes mediante fragmentación química, física o enzimática o utilizando protocolos bien conocidos. La fragmentación química puede emplear metales ferrosos (por ejemplo, Fe-EDTA). Los métodos físicos pueden incluir sonicación, fuerza hidrodinámica o nebulización (véase la solicitud de patente europea EP 0 552 290). Los protocolos enzimáticos pueden emplear nucleasas, tales como la nucleasa microcócica (Mnasa) o exo-nucleasas (tales como Bal31 o Exol) o endonucleasas de restricción. El protocolo por el cual los fragmentos se generan no debería afectar el uso de los fragmentos en los métodos. Puede ser ventajoso durante el enriquecimiento, emplear fragmentos en un intervalo de tamaño compatible con la tecnología post-enriquecimiento en la que se utilizarán los fragmentos enriquecidos. Un tamaño de fragmento adecuado puede estar en el intervalo de entre aproximadamente 100 y aproximadamente 1000 residuos de nucleótidos o pares de bases, o entre aproximadamente 250 y aproximadamente 800 residuos de nucleótidos o pares de bases, y puede ser de aproximadamente 400 a aproximadamente 600 residuos de nucleótidos o pares de bases, en particular alrededor de 500 residuos de nucleótidos o pares de bases.

Las sondas se corresponden en secuencia con al menos una región del genoma y se pueden proporcionar en un soporte sólido en paralelo utilizando la tecnología de síntesis de chip sin máscara. Alternativamente, las sondas se pueden obtener en serie utilizando un sintetizador de DNA estándar y aplicándolas después al soporte sólido o pueden obtenerse a partir de un organismo y a continuación, inmovilizarlas sobre el soporte sólido. Después de la hibridación, los ácidos nucleicos que no hibridan o que se hibridan de forma inespecífica a las sondas se separan de las sondas unidas al soporte mediante un lavado. Los ácidos nucleicos restantes, unidos específicamente a las sondas, se eluyen del soporte sólido en, por ejemplo agua caliente o en un tampón de elución que contiene ácido

nucleico, por ejemplo tampón TRIS y / o EDTA para producir un eluato enriquecido de moléculas de ácido nucleico diana.

En algunas realizaciones, los enlazantes de doble cadena se proporcionan al menos en uno de los extremos de las moléculas de ácido nucleico (genómico) antes de que los fragmentos se desnaturalicen y se hibriden con las sondas inmovilizadas. En tales realizaciones, las moléculas de ácido nucleico diana pueden amplificarse después de la elución para producir un grupo de productos amplificados con una complejidad relativa reducida respecto a la muestra original. Las moléculas de ácido nucleico diana pueden amplificarse utilizando, por ejemplo, LM-PCR no específica a través de múltiples rondas de ciclos térmicos. Opcionalmente, los productos amplificados se pueden enriquecer aún más mediante una segunda selección con las sondas. Los productos de la segunda selección se pueden amplificar de nuevo antes de su uso tal como se ha descrito. Esta aproximación se resume gráficamente en la figura 1 y en un diagrama de flujo en la figura 2. Los enlazantes pueden proporcionarse en un tamaño arbitrario y con una secuencia de ácido nucleico arbitraria de acuerdo con lo que se requiere para aplicaciones analíticas posteriores al paso de reducción de la complejidad. Los enlazantes pueden variar entre aproximadamente 12 y aproximadamente 100 pares de bases, incluyendo un intervalo entre aproximadamente 18 y 100 pares de bases, y preferiblemente entre aproximadamente 20 y 24 pares de bases.

Alternativamente, las sondas de ácido nucleico para moléculas diana pueden sintetizarse sobre un soporte sólido, liberadas del soporte sólido como un conjunto de sondas y amplificadas como se ha descrito. El conjunto amplificado de sondas liberadas puede inmovilizarse covalentemente o no covalentemente sobre un soporte, tal como vidrio, metal, cuentas de cerámica o poliméricas u otro soporte sólido. Las sondas pueden diseñarse para la liberación conveniente desde el soporte sólido, proporcionando, por ejemplo, en o cerca del extremo de la sonda proximal al soporte una secuencia de ácido nucleico lábil al ácido o al álcali que libera las sondas bajo condiciones de pH alto o bajo, respectivamente. Se conocen varias reacciones químicas de enlace en la técnica. El soporte puede proporcionarse, por ejemplo, en una columna que tiene una entrada y salida de fluidos. La técnica está familiarizada con los métodos para la inmovilización de ácidos nucleicos sobre soportes, por ejemplo mediante la incorporación de un nucleótido biotinilado en las sondas y recubriendo el soporte con estreptavidina de tal manera que el soporte revestido atrae de forma no covalente e inmoviliza las sondas. La muestra o las muestras se pasan a través del soporte que contiene la sonda en condiciones de hibridación de tal forma que las moléculas de ácidos nucleicos que hibridan con el soporte inmovilizado pueden eluirse para el análisis posterior u otra utilización.

En un aspecto, la descripción permite la captura y enriquecimiento para las moléculas de ácido nucleico diana o región genómica diana a partir de una muestra biológica compleja mediante selección genómica directa. El método también es útil en la búsqueda de variantes y mutaciones genéticas, tales como polimorfismos de nucleótido único (SNP), o un conjunto de SNP, que subyacen en las enfermedades humanas. Se contempla que la captura y el enriquecimiento utilizando la tecnología de hibridación de microchips es mucho más flexible que otros métodos actualmente disponibles en el campo del enriquecimiento genómico, tales como el uso de BAC (cromosoma artificial bacteriano) para la selección genómica directa (ver Lovett et al., 1991).

El método permite métodos de secuenciación específicos basados en chip o, expansiva, capilar, u otros métodos conocidos en la técnica. En general, las estrategias para la secuenciación expansiva de fragmentos generados al azar son rentables y se integran fácilmente en un proyecto, pero la invención mejora la eficiencia del método de secuenciación expansiva presentando sólo fragmentos de una o más regiones genómicas de interés para la secuenciación. El método proporciona una capacidad de enfocar las estrategias de secuenciación de regiones genómicas específicas, tales como cromosomas individuales o exones para los propósitos médicos de secuenciación.

Las moléculas de ácido nucleico diana pueden enriquecerse a partir de una o más muestras que incluyen ácidos nucleicos de cualquier fuente, en forma purificada o no purificada. La fuente no necesita contener un complemento completo de moléculas de ácido nucleico genómico de un organismo. La muestra, preferiblemente de una fuente biológica, incluye, pero no se limita a los aislamientos agrupados de pacientes individuales, muestras de tejido, o cultivo celular. Tal como se utiliza aquí, el término "moléculas de ácido nucleico diana" se refiere a moléculas de una región genómica diana a ser estudiado. Las sondas preseleccionadas determinan el rango de moléculas de ácido nucleico diana. El experto en posesión de esta descripción apreciará el rango completo de posibles dianas y dianas asociadas.

La región diana puede ser uno o más bloques continuos de varias megabases (Mb), o varias regiones contiguas o no contiguas más pequeñas, tales como todos los exones de uno o más cromosomas, o de los sitios que se sabe que contienen SNP. Por ejemplo, el soporte sólido puede soportar un microchip de secuencia contigua diseñado para capturar uno o más cromosomas completos, partes de uno o más cromosomas, todos los exones, todos los exones de uno o más cromosomas, exones seleccionados, intrones y exones para uno o más genes, regiones reguladoras de genes, y así sucesivamente. Alternativamente, para aumentar la probabilidad de enriquecer las dianas no únicas deseadas o dianas de difícil captura, las sondas pueden dirigirse a secuencias asociadas con la secuencia diana real (por ejemplo, en el mismo fragmento, pero separada de ella), en dicho caso se capturarán fragmentos genómicos que contienen tanto el objetivo deseado como las secuencias asociadas y enriquecidas. Las secuencias asociadas pueden ser adyacentes o separadas de las secuencias diana, pero el experto en la materia

apreciará que cuanto más cerca están las dos partes entre sí, más probable será que los fragmentos genómicos contengan ambas porciones. Para reducir aún más el impacto limitado de la hibridación cruzada de las moléculas fuera de objetivo, se puede realizar ya sea mejorando la integridad del enriquecimiento, con rondas secuenciales de captura utilizando conjuntos de sonda de captura distintos pero relacionados dirigidos a la región diana. Las sondas relacionadas son sondas correspondientes a las regiones en estrecha proximidad entre sí en el genoma que pueden, por lo tanto, hibridar con el mismo fragmento de DNA genómico.

Los oligonucleótidos de microchips están diseñados para localizar la región diana o regiones del genoma. La longitud de las sondas individuales está normalmente entre 50 y 200 bases. Estas sondas pueden estar diseñadas para ser sondas superpuestas, lo que significa que los nucleótidos de partida de las sondas adyacentes son más pequeños que la longitud de una sonda o de sondas que no se solapan, donde la distancia entre sondas adyacentes son mayores que la longitud de una sonda. La distancia entre las sondas adyacentes en general se solapa, con una separación entre el nucleótido de partida de las dos sondas que varían entre 1 y 100 bases. Esta distancia puede variar para facilitar que algunas regiones genómicas sean localizadas por un mayor número de sondas que otras. Esta variación puede utilizarse para modular la eficiencia de regiones genómicas individuales, normalizando la captura. Las sondas se pueden probar por su exclusividad en el genoma. Para evitar la unión no específica de elementos genómicos para capturar chips, los elementos altamente repetitivos del genoma deben ser excluidos de los diseños de microchips de selección utilizando un nuevo método que utiliza una estrategia similar al programa WindowMasker desarrollado por Morgolis (2006) para identificar estas regiones y excluirlas de la selección de sondas. El proceso comparó el conjunto de sondas contra un histograma de frecuencias precalculado de todas las posibles sondas 15-mericas en el genoma humano. Para cada sonda, las frecuencias de los 15-meros que comprenden las sondas se utilizan entonces para calcular la frecuencia media del 15-mero de la sonda. Cuanto mayor es la frecuencia media de 15-mero, más probable es que la sonda esté dentro de los límites de una región repetitiva del genoma. Sólo se deben utilizar sondas con una frecuencia media de 15-mero inferior a 100.

La naturaleza y el rendimiento de las sondas puede variarse para normalizar ventajosamente o ajustar la distribución de las moléculas diana capturadas y enriquecidas de acuerdo con los métodos. Un objetivo de dicha normalización es entregar un gen expresado por lectura (ver Soares, et al., 1994). La normalización se puede aplicar, por ejemplo, a las poblaciones de moléculas de DNAC antes de la construcción de la biblioteca, porque la distribución de las moléculas en la población refleja los diferentes niveles de expresión de genes expresados a partir de los cuales se producen las poblaciones de moléculas de DNAC. Por ejemplo, el número de reacciones de secuenciación necesarios para analizar de forma eficaz cada región diana se puede reducir mediante la normalización del número de copias de cada secuencia diana en la población enriquecida de tal manera que a través de la serie de sondas, el rendimiento de captura de sondas distintas se normaliza, en base a una combinación del grado de adecuación ("fitness") y otros atributos de la sonda. El grado de adecuación, que se caracteriza por una "métrica de captura" puede determinarse informáticamente o empíricamente. En un enfoque, la capacidad de las moléculas diana para unirse se puede ajustar proporcionando las denominadas sondas de oligonucleótidos isotérmicas ( $T_m$  equilibrado), como se describe en la Publicación de la Solicitud de Patente US N ° 2005/0282209 (NimbleGen Systems, Madison, WI), que permiten un rendimiento de sonda uniforme, eliminar artefactos y / o sesgos de hibridación y proporcionar resultados de mayor calidad. Las longitudes de sonda se ajustan (típicamente, de aproximadamente 20 a aproximadamente 100 nucleótidos, preferiblemente de aproximadamente 40 a aproximadamente 85 nucleótidos, en particular de aproximadamente 45 a aproximadamente 75 nucleótidos, por ejemplo, 45 nucleótidos, pero opcionalmente también más de 100 nucleótidos hasta aproximadamente 250 nucleótidos) para igualar la temperatura fusión (por ejemplo,  $T_m = 76^\circ\text{C}$ , por lo general desde alrededor de  $55^\circ\text{C}$  a aproximadamente  $76^\circ\text{C}$ , en particular desde aproximadamente  $72^\circ\text{C}$  a aproximadamente  $76^\circ\text{C}$ ) en todo el conjunto. Por lo tanto, las sondas están optimizadas para realizar de manera equivalente a una rigurosidad determinada en las regiones genómicas de interés, incluyendo las regiones ricas en AT y GC. En relación con esto, la secuencia de las sondas individuales se puede ajustar, utilizando bases naturales o análogos de bases sintéticas tales como inositol, o una combinación de las mismas para lograr un grado de adecuación de captura deseada de aquellas sondas. Del mismo modo, pueden emplearse sondas de ácidos nucleicos bloqueadas, sondas de ácidos nucleicos peptídicos o similares que tienen estructuras que producen un rendimiento de captura deseado. El experto en la materia en posesión de la presente descripción apreciarán que la longitud de la sonda, la temperatura de fusión y la secuencia se puede ajustar de forma coordinada para cualquier sonda dada para llegar a un rendimiento de captura deseada para la sonda. Convenientemente, la temperatura de fusión ( $T_m$ ) de la sonda se puede calcular utilizando la fórmula:  $T_m = 5x (G_n + C_n) + 1x (A_n + T_n)$ , donde n es el número de cada base específica (A, T, G o C) presente en la sonda.

El rendimiento de captura también puede normalizarse mediante la determinación del grado de adecuación de captura de las sondas en el conjunto de la sonda, y después ajustando la cantidad de sondas individuales en el soporte sólido. Por ejemplo, si una primera sonda captura veinte veces más ácido nucleico que una segunda sonda, entonces el rendimiento de captura de ambas sondas puede ser igualado al proporcionar veinte veces más copias de la segunda sonda, por ejemplo mediante el aumento de veinte veces el número de características que muestran la segunda sonda. Si las sondas se preparan en serie y se aplican al soporte sólido, la concentración de las sondas individuales en el conjunto se puede variar de la misma manera.

Aún más, otra estrategia para la normalización de captura de ácidos nucleicos diana es someter a las moléculas diana eluidas a una segunda ronda de hibridación frente a las sondas bajo condiciones menos rigurosas que las que

se utilizaron para la primera ronda de hibridación. Aparte del enriquecimiento sustancial en la primera hibridación que reduce la complejidad relativa respecto al ácido nucleico genómico original, la segunda hibridación puede llevarse a cabo bajo condiciones de hibridación que saturan todas las sondas de captura. Suponiendo que se proporcionan sustancialmente las mismas cantidades de las sondas de captura en el soporte sólido, la saturación de las sondas asegurará que las mismas cantidades de cada diana se eluyan después de la segunda hibridación y lavado.

Otra estrategia de normalización sigue la elución y amplificación de las moléculas diana capturadas del soporte sólido. Las moléculas diana en el eluato se desnaturalizan usando, por ejemplo, un proceso de desnaturalización química o térmica, para un estado de una sola cadena y se vuelven a hibridar. Las consideraciones cinéticas dictan que las especies abundantes rehibridan antes que las especies menos abundantes. Como tal, mediante la eliminación de la fracción inicial de especies rehibridadas, las especies de cadena sencilla restantes estarán equilibradas en relación con la población inicial en el eluato. El tiempo requerido para la eliminación óptima de las especies abundantes se determina empíricamente.

En resumen, una realización proporciona un nuevo método de reducción de la complejidad genética de una población de moléculas de ácido nucleico. Este método comprende

(a) exponer las moléculas de ácido nucleico fragmentadas, desnaturalizadas de dicha población en múltiples sondas de oligonucleótidos diferentes que se unen a un soporte sólido en condiciones de hibridación para capturar moléculas de ácido nucleico que hibridan específicamente con dichas sondas, o exponiendo las moléculas de ácido nucleico fragmentadas, desnaturalizadas de dicha población en múltiples, sondas de oligonucleótidos diferentes en condiciones de hibridación, seguido por la unión de los complejos de moléculas hibridadas sobre un soporte sólido para capturar moléculas de ácido nucleico que hibridan específicamente con dichas sondas, en el que (en ambos casos) dichas moléculas de ácido nucleico fragmentadas, desnaturalizadas tienen un tamaño medio de aproximadamente 100 a aproximadamente 1000 residuos de nucleótidos, preferiblemente entre aproximadamente 250 a aproximadamente 800 residuos de nucleótidos y lo más preferiblemente entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos,

(b) separar los ácidos nucleicos no unidos e hibridados de forma inespecífica de las moléculas capturadas;

(c) eluir las moléculas capturadas del soporte sólido, preferiblemente en un grupo de eluatos con una complejidad genética reducida en relación a la muestra original, y

(d) opcionalmente repetir los pasos (a) a (c) durante al menos un ciclo adicional con las moléculas capturadas eluidas.

En la mayoría de los casos, la población de moléculas de ácidos nucleicos son moléculas originadas a partir de una muestra de DNA genómico (moléculas de ácido nucleico genómico). Sin embargo, también es posible comenzar con una muestra de DNAC o incluso RNA. La fragmentación puede realizarse en principio mediante cualquier método que se conoce en la técnica como ya se ha explicado anteriormente. Sin embargo, las moléculas de ácido nucleico desnaturalizadas fragmentadas deben tener un tamaño medio entre aproximadamente 100 a aproximadamente 1000 residuos de nucleótidos, preferiblemente entre aproximadamente 250 a aproximadamente 800 residuos de nucleótidos y lo más preferiblemente entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos. Por ejemplo, esto se puede lograr mediante la nebulización de DNA genómico (véase por ejemplo la solicitud de patente europea EP 0 552 290).

Los parámetros de reducción de la complejidad genética se pueden elegir casi arbitrariamente, dependiendo del deseo del usuario para la selección de secuencia, y se definen mediante las secuencias de las múltiples sondas de oligonucleótidos. En una realización, dichas múltiples sondas definen una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos. En otra realización, dichas sondas múltiples definen la secuencia completa de al menos un locus genético único, dicho locus tiene un tamaño de al menos 100 kb y preferiblemente al menos 1 Mb o un tamaño tal como se ha especificado anteriormente. En otra realización, dichas sondas múltiples definen sitios conocidos por contener SNP. En una realización adicional, dichas sondas múltiples definen un microchip de secuencia contigua. Dicho microchip de secuencia contigua en el contexto de la presente invención se ha definido como el diseño para capturar la secuencia completa de al menos un cromosoma completo. En este contexto, el término "definir" se entiende de tal manera que la población de sondas múltiples comprende al menos una sonda para cada secuencia diana que pasará a enriquecerse. Preferiblemente, la población de sondas múltiples comprende adicionalmente al menos una segunda sonda para cada secuencia diana que se enriquecerá, caracterizadas porque dicha segunda sonda posee una secuencia que es complementaria a dicha primera secuencia.

El soporte sólido de acuerdo con la presente invención es o bien un microchip de ácidos nucleicos o una población de cuentas. Dichas cuentas pueden ser de vidrio, metal, cerámica y cuentas poliméricas. Si dicho soporte sólido es un microchip, es posible sintetizar las sondas de captura de oligonucleótidos in situ directamente sobre dicho soporte sólido. Por ejemplo, las sondas pueden sintetizarse en el microchip utilizando un sintetizador de chip sin máscara

(US 6.375.903). Las longitudes de las múltiples sondas de oligonucleótidos pueden variar, son dependientes del diseño experimental y sólo están limitadas por la posibilidad de sintetizar tales sondas. Preferiblemente, la longitud media de la población de sondas múltiples es de aproximadamente 20 a aproximadamente 100 nucleótidos, preferiblemente entre aproximadamente 40 a aproximadamente 85 nucleótidos, en particular entre aproximadamente 45 a aproximadamente 75 nucleótidos, por ejemplo, 45 nucleótidos.

Si el soporte sólido es una población de cuentas, las sondas de captura pueden sintetizarse inicialmente en un microchip utilizando un sintetizador de chip sin máscara, a continuación, se libera o escinde de acuerdo con métodos estándar conocidos, se amplifica opcionalmente y después se inmoviliza en dicha población de cuentas de acuerdo con métodos conocidos en la técnica. Las cuentas pueden empaquetarse en una columna de modo que una muestra se carga y se pasa a través de la columna para reducir la complejidad genética. Alternativamente, con el fin de mejorar la cinética de hibridación, la hibridación puede tener lugar en una solución acuosa que comprende las cuentas con las múltiples moléculas de oligonucleótidos inmovilizadas en suspensión.

En una realización, las múltiples sondas de oligonucleótidos diferentes llevan cada una un grupo químico o enlazante, es decir, una porción que permite la inmovilización sobre un soporte sólido, también denominado un grupo inmovilizable. A continuación, el paso de exponer las moléculas fragmentadas, desnaturalizadas de ácido nucleico de la muestra a las múltiples, diferentes sondas de oligonucleótidos en condiciones de hibridación se lleva a cabo en una solución acuosa y la inmovilización sobre un soporte sólido apropiado se lleva a cabo posteriormente. Por ejemplo, dicha porción puede ser biotina que puede utilizarse para la inmovilización en un soporte sólido recubierto con estreptavidina. En otra realización, dicha porción puede ser un hapteno como digoxigenina, que puede utilizarse para la inmovilización sobre un soporte sólido revestido con un anticuerpo reconocedor de hapteno, por ejemplo un anticuerpo de unión a digoxigenina.

En una realización específica, la pluralidad de sondas inmovilizadas se caracteriza por un rendimiento de captura normalizada. El rendimiento de captura normalizada se consigue generalmente mediante métodos como se ha descrito anteriormente, que comprende típicamente las etapas de a) determinar el grado de adecuación de captura de las sondas en el conjunto de sondas; y b) ajustar la cantidad de al menos una sonda en el soporte sólido. Alternativamente, el rendimiento de captura normalizada se consigue mediante un método que comprende los pasos de a) determinar el grado de adecuación de captura de las sondas en el conjunto de sondas; y b) ajustar al menos uno de la secuencia, la temperatura de fusión y la longitud de la sonda de al menos una sonda en el soporte sólido. Todavía alternativamente, el rendimiento de captura normalizada se consigue mediante un método que comprende los pasos de a) exponer las moléculas capturadas al menos a una sonda inmovilizada sobre el soporte sólido en condiciones menos rigurosas que en el primer paso de exposición de tal manera que al menos una sonda está saturada, b) lavar los ácidos nucleicos no unidos y los unidos inespecíficamente a partir del soporte sólido; y c) eluir los ácidos nucleicos diana unidos del soporte sólido. Todavía alternativamente, el rendimiento de captura normalizada se consigue mediante un método que comprende los pasos de a) la desnaturalizar las moléculas capturadas eluidas a un estado de cadena sencilla; b) rehibridar las moléculas de una sola cadena hasta que una porción de las moléculas sean de doble cadena; y descartar las moléculas de doble cadena y c) retener las moléculas de cadena sencilla.

Por lo general, al menos una sonda inmovilizada se hibrida a una región genómica de interés en fragmentos de ácido nucleico en la muestra. Alternativamente, la al menos una sonda inmovilizada puede hibridarse con secuencias en fragmentos de ácido nucleico diana que comprende una región genómica de interés, separando las secuencias de hibridación de la región genómica de interés. Además, está también dentro del alcance de la presente invención, la realización de al menos un segundo paso de hibridación utilizando al menos una sonda de oligonucleótidos relacionada con, pero distinta a la sonda utilizada en la hibridación inicial.

En particular, la presente descripción también está dirigida a un método para determinar la información de la secuencia de ácido nucleico de al menos una región de ácido nucleico genómico en una muestra, comprendiendo el método los pasos de:

- reducir la complejidad genética de una población de moléculas de ácido nucleico de acuerdo con cualquier método como se ha descrito aquí, y
- determinar la secuencia de ácido nucleico de las moléculas capturadas por ejemplo, mediante la realización de una reacción de secuenciación. Preferiblemente, dicha reacción de secuenciación es una secuenciación mediante reacción de síntesis.

De acuerdo con esta realización, el DNA genómico se fragmenta preferiblemente mediante estrés mecánico. El tamaño medio deseado de los fragmentos de DNA deberá ser pequeño ( $\leq 1000$  pb) y depende del método de secuenciación a aplicar.

La secuenciación por síntesis de acuerdo con la literatura en la técnica (véase, por ejemplo Hyman, E.D., 1988) se define como cualquier método de secuenciación que monitoriza la generación de productos secundarios tras la incorporación de un desoxinucleósido trifosfato específico durante la reacción de secuenciación (véase, por ejemplo Rhonaghi et al., 1998). Una forma de realización particular y más prominente de la secuenciación mediante reacción

de síntesis es el método de secuenciación de pirofosfato. En este caso, la generación de pirofosfato durante la incorporación de nucleótidos se controla mediante una cascada enzimática que finalmente resulta en la generación de una señal quimioluminiscente. Por ejemplo, el sistema secuenciador Genome 454 (Roche Applied Science N° Cat. 04 760 085 001) se basa en la tecnología de secuenciación de pirofosfato. Para la secuenciación en un instrumento 454 GS20 o 454 FLX, el tamaño medio del fragmento de DNA genómico debe estar en el rango de 200 o 600 pb, respectivamente.

Alternativamente, la secuenciación mediante la reacción de síntesis es una reacción de secuenciación de tipo terminador. En este caso, los bloques de construcción dNTP incorporados comprenden un marcador detectable, que es preferiblemente un marcador fluorescente que impide una mayor extensión de la cadena de DNA naciente. Luego se retira el marcaje y se detecta tras la incorporación del bloque de construcción de dNTP en el híbrido de molde / cebador de extensión, por ejemplo, mediante la utilización de una polimerasa de DNA que comprende una exonucleasa 3'-5' o actividad de corrección de galeradas.

Ventajosamente, el método de la primera reducción de la complejidad genómica y luego la determinación de múltiples secuencias comprende además el paso de ligación de moléculas adaptadoras para uno o ambos, preferiblemente ambos extremos de las moléculas de ácido nucleico fragmentadas. Las moléculas adaptadoras se definen preferentemente como oligonucleótidos de doble cadena con extremos romos. Además, el método puede comprender además el paso de amplificación de dichas moléculas de ácido nucleico con al menos un cebador, dicho cebador comprende una secuencia que corresponde a o se hibrida específicamente con la secuencia de dichas moléculas adaptadoras.

Con el fin de ligar moléculas adaptadoras en una molécula diana de doble hebra, se prefiere que esta misma molécula diana sea de extremo romo. Con el fin de lograr esto, las moléculas diana de doble cadena se someten a una reacción de relleno con una polimerasa de DNA, tal como una polimerasa de DNA T4 o la polimerasa Klenow en presencia de desoxinucleósidos trifosfatos, lo que resulta en moléculas diana de extremos romos. Además, por ejemplo la polinucleótido quinasa T4 se añade antes de la ligación con el fin de añadir grupos fosfato al extremo 5' para el posterior paso de ligación. Tras la ligación de los adaptadores (oligonucleótidos de DNA cortos de doble cadena de extremos romos con aproximadamente 3-20 pares de bases) en el DNA diana pulido, puede realizarse de acuerdo con cualquier método que se conoce en la materia, preferiblemente mediante una reacción de ligasa de DNA T4.

Dicha ligación se puede realizar antes o después del paso de exponer una muestra que comprende, moléculas genómicas desnaturalizadas fragmentadas de ácido nucleico a múltiples sondas de oligonucleótidos en condiciones de hibridación para la captura de moléculas de ácido nucleico diana que se hibridan a dichas sondas. En caso de que la ligación se realice posteriormente, los ácidos nucleicos enriquecidos que se liberan del soporte sólido en forma de cadena sencilla se deben rehibridar primero seguido por una reacción de extensión del cebador y una reacción de relleno de acuerdo con los métodos estándar conocidos en la materia.

La ligación de dichas moléculas adaptadoras permite un paso de amplificación posterior de las moléculas capturadas. Independiente de si la ligación tiene lugar antes o después del paso de captura, existen dos formas de realización alternativas. En la primera forma de realización, se utiliza un tipo de moléculas adaptadoras. Esto da lugar a la población de fragmentos con secuencias terminales idénticas en ambos extremos del fragmento. Como consecuencia de ello, es suficiente utilizar sólo un cebador en un paso de amplificación potencial posterior. En una realización alternativa, se utilizan dos tipos de moléculas adaptadoras A y B. Esto resulta en una población de moléculas enriquecidas compuesta de tres tipos diferentes: (i) fragmentos que tienen un adaptador (A) en un extremo y otro adaptador (B) en el otro extremo, (ii) fragmentos que tienen adaptadores de A en ambos extremos, y (iii) fragmentos que tienen adaptadores de B en ambos extremos.

La generación de moléculas enriquecidas de acuerdo al tipo (i) es una ventaja excepcional, si la amplificación y la secuenciación es por ejemplo, se realiza con el instrumento 454 GS20 y GSFLX de life science corporation (ver Library Prep Manual GS20, dic 2006, WO 2004/070007). Si uno de dichos adaptadores, por ejemplo el adaptador B lleva una biotina modificada, entonces las moléculas (i) y (iii) pueden, por ejemplo unirse a partículas magnéticas cubiertas de estreptavidina (SA) para un mayor aislamiento y los productos de (ii) se eliminan mediante lavado. En caso de que el DNA enriquecido e inmovilizado por SA sea monocatenario después de la elución del chip de captura / soporte sólido, es ventajoso hacer que el DNA sea de doble cadena. En este caso los cebadores complementarios al adaptador A se pueden añadir a los productos lavados de la SA. Ya que las porciones que son B-B (iii anteriormente) no tienen A o su complemento disponible, solo los productos capturados por SA y adaptados a A-B se harán de doble cadena tras la extensión por cebador a partir de un cebador complementario a A. Posteriormente, las moléculas de DNA de doble cadena que se han unido a dichas partículas magnéticas se desnaturalizan térmicamente o químicamente (por ejemplo, NaOH) de una forma tal que la cadena recién sintetizada se libera en la solución. Debido a la estrecha unión de biotina / estreptavidina, por ejemplo, las moléculas con sólo dos adaptadores B no se liberarán en la solución. La única cadena disponible para su liberación es la cadena sintetizada con extensión por cebador complementario a A hasta complementario a B. Dicha solución que comprende moléculas diana de cadena sencilla con un adaptador A en un extremo y un adaptador B en el otro extremo puede, por ejemplo

posteriormente unirse a otro tipo de cuentas que comprenden una secuencia de captura que es suficientemente complementaria a las secuencias adaptadoras de A o B para su posterior procesamiento.

En el caso que el flujo de trabajo del Genome Sequencer (Roche Applied Science N° de catálogo 04 896 548 001), en un primer paso, se lleva a cabo la amplificación (clonal) mediante PCR en emulsión. Por lo tanto, está también dentro del alcance de la presente invención, que el paso de amplificación se realice en forma de una PCR en emulsión. Las cuentas que llevan los ácidos nucleicos diana amplificados por clonación pueden entonces transferirse de forma arbitraria a una placa picotitulada de acuerdo con el protocolo del fabricante y someterse a una reacción de secuenciación pirofosfato para la determinación de la secuencia.

Por lo tanto, los métodos de acuerdo con la presente descripción permiten determinaciones de secuencia para una variedad de aplicaciones diferentes. Por ejemplo, la presente descripción proporciona también un método para la detección de la variación en la región codificante en relación a un genoma de referencia, preferiblemente en una muestra que comprende, moléculas de ácido nucleico genómico fragmentado desnaturalizado, comprendiendo el método los pasos de:

- realizar el(los) método(s) tal como se describió anteriormente,
- determinar la secuencia de ácido nucleico de las moléculas capturadas, y
- comparar la secuencia determinada con una base de datos, en particular con una base de datos de polimorfismos en el genoma de referencia para identificar las variantes del genoma de referencia.

En un aspecto adicional importante, la presente descripción también proporciona un equipo para realizar un método o una parte de un método de acuerdo con la presente invención como se describe aquí. Por lo tanto, la presente invención también se refiere a un equipo que comprende

- una (primera) molécula adaptadora de doble cadena, y
- un soporte sólido con múltiples sondas, en el que las múltiples sondas se seleccionan de:
  - una pluralidad de sondas que definen una pluralidad de exones, intrones o secuencias reguladoras a partir de una pluralidad de loci genéticos
  - una pluralidad de sondas que definen la secuencia completa de al menos un locus genético único, dicho locus posee un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb o un tamaño tal como se ha especificado en este documento,
  - una pluralidad de sondas que definen sitios conocidos por contener SNP, y
  - una pluralidad de sondas que definen un chip, en particular un microchip de secuencia contigua especialmente diseñado para capturar la secuencia completa de al menos un cromosoma completo.

Preferiblemente, el equipo contiene dos moléculas adaptadoras de doble cadena diferentes. El soporte sólido puede ser cualquiera de una pluralidad de cuentas o un microchip como se describe aquí.

En una realización, dicho equipo comprende además al menos uno o más compuestos a partir de un grupo que consiste en la polimerasa de DNA, polinucleótido quinasa de T4, DNA ligasa de T4, una solución de hibridación de chip, por ejemplo, como se describe aquí, una solución de lavado de chip, en particular, una solución de lavado con SSC, DTT y opcionalmente SDS, por ejemplo, tampón de lavado I (0,2 x SSC, 0,2% (v / v) SDS, DTT 0,1 mM), tampón de lavado II (0,2 x SSC, DTT 0,1 mM) y / o solución de lavado III (0,05x SSC, DTT 0,1 mM), y / o una solución de elución de chip, por ejemplo, agua o una solución que contiene tampón TRIS y / o EDTA.

En una realización específica adicional, sin ser mutuamente exclusivo a la realización descrita en el presente documento, el equipo comprende una segunda molécula adaptadora. Al menos una cadena de oligonucleótido de dicha primera o segunda molécula adaptadora puede llevar una modificación, que permite la inmovilización sobre un soporte sólido. Por ejemplo, tal modificación puede ser un marcador de biotina que se puede utilizar para la inmovilización en un soporte sólido recubierto con estreptavidina. Alternativamente, dicha modificación puede ser un hapteno como digoxigenina, que puede utilizarse para la inmovilización sobre un soporte sólido revestido con un anticuerpo que reconoce haptenos.

Tal como se utiliza aquí, el término "hibridación" se usa en referencia al emparejamiento de ácidos nucleicos complementarios. La hibridación y la fuerza de hibridación (es decir, la fuerza de la asociación entre los ácidos nucleicos) se ve afectada por factores tales como el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones implicadas, la Tm del híbrido formado, y la proporción G:C de los ácidos nucleicos. Aunque la invención no se limita a un conjunto particular de condiciones de hibridación, preferiblemente se emplean condiciones de hibridación rigurosas. Las condiciones de hibridación rigurosas dependen de la secuencia y serán diferentes al variar los parámetros ambientales (por ejemplo, concentraciones de sal, y la presencia de materia orgánica). Generalmente, se seleccionan las condiciones "rigurosas" en alrededor de 5 °C a 20 °C por debajo del punto de fusión térmico (Tm) para la secuencia de ácido nucleico específica a una fuerza iónica y pH definidos. Preferiblemente, las condiciones rigurosas están entre alrededor de 5 °C y 10 °C más bajas que el punto de fusión térmico para un ácido nucleico específico unido a un ácido nucleico complementario. La Tm es la temperatura (bajo

fuerza iónica y pH definidos) a la cual el 50% de un ácido nucleico (por ejemplo, marcaje de ácido nucleico) se hibrida a una sonda perfectamente coincidente.

De manera similar, las condiciones de lavado "rigurosas" se determinan habitualmente de manera empírica para la hibridación de cada conjunto de marcajes con el correspondiente chip de sondas. Los chips se hibridan primero (normalmente bajo condiciones de hibridación rigurosas) y luego se lavan con tampones que contienen concentraciones sucesivamente más bajas de sales, o mayores concentraciones de detergentes, o a temperaturas crecientes hasta que la relación de señal-ruido para la hibridación específico a no específica sea lo suficientemente alta para facilitar la detección de la hibridación específica. Las condiciones de temperatura rigurosas generalmente incluirán temperaturas en exceso de aproximadamente 30 °C, más usualmente en exceso de aproximadamente 37 °C, y ocasionalmente en exceso de aproximadamente 45 °C. Las condiciones de sal rigurosas serán habitualmente inferiores a aproximadamente 1000 mM, habitualmente inferiores a aproximadamente 500 mM, más habitualmente inferiores a aproximadamente 150 mM. Para más información véase, por ejemplo, Wetmur et al. (1966) y Wetmur (1991).

"Condiciones rigurosas" o "condiciones de alta rigurosidad", como se define en este documento, puede ser la hibridación en formamida al 50%, 5x SSC (NaCl 0,75 M, citrato sódico 0,075 M), fosfato de sodio 50 mM (pH 6,8), 0,1% pirofosfato de sodio, solución de Denhardt 5x, DNA de esperma de salmón sonicado (50 mg / ml), 0,1% de SDS, y sulfato de dextrano al 10% a 42 °C, con lavados a 42 °C en 0,2 x SSC (cloruro de sodio / citrato de sodio) y 50% de formamida a 55 °C, seguido de un lavado con 0,1 x SSC que contiene EDTA a 55 °C.

A modo de ejemplo, pero sin limitación, se contempla que los tampones que contienen 35% de formamida, 5x SSC, y 0,1% (p / v) de dodecil sulfato de sodio son adecuados para hibridar bajo condiciones moderadamente no rigurosas a 45 °C durante 16-72 horas. Además, se prevé que la concentración de formamida se puede ajustar adecuadamente entre un intervalo de 20-45% dependiendo de la longitud de la sonda y el nivel de rigurosidad deseado. También está incluido dentro del alcance de la invención que la optimización de la sonda se puede obtener para sondas más largas (>> 50-mero), mediante el aumento de la temperatura de hibridación o la concentración de formamida para compensar un cambio en la longitud de la sonda. Ejemplos adicionales de condiciones de hibridación se proporcionan en varias fuentes, incluyendo: "Direct selection of cDNAs with large genomic DNA clones", en Molecular Cloning: A Laboratory Manual (2001).

En otra realización, la presente descripción se refiere a un método para aislar y reducir la complejidad de una pluralidad de secuencias de ácido nucleico que comprende: proporcionar un soporte sólido en el que dicho soporte sólido comprende sondas de hibridación hibridables con secuencias de ácido nucleico diana, y una muestra de ácido nucleico fragmentado que comprende secuencias de ácido nucleico diana, amplificar dichas sondas de hibridación en el que los productos de amplificación comprenden una porción de unión y en el que dichos productos de amplificación se mantienen en solución, hibridar dicha muestra de ácido nucleico a dichos productos de amplificación en solución de forma que la hibridación entre dichos productos de amplificación y las secuencias de ácidos nucleicos diana se permite que se produzca, separar los complejos de hibridación del ácido nucleico diana / producto de amplificación de los ácidos nucleicos hibridados de forma inespecífica por dicha porción de unión, y eluyendo las secuencias de ácido nucleico diana hibridado del complejo, aislando de este modo y reduciendo la complejidad de una pluralidad de secuencias de ácidos nucleicos.

En otra realización, la presente descripción comprende el método anterior, que comprende además la secuenciación de las secuencias de ácidos nucleicos diana eluidas.

Una realización adicional se refiere a un equipo que comprende: secuencias de la sonda de hibridación que comprende una porción de unión en el que dichas secuencias de sondas se diseñan para hibridarse a una o más secuencias diana de ácido nucleico y en el que dichas secuencias de sonda están en solución, un sustrato que comprende una pareja de unión para unirse a dicha porción de unión, y las instrucciones para llevar a cabo los métodos anteriores.

En algunas realizaciones, una muestra que contiene moléculas de ácido nucleico desnaturalizado (por ejemplo, de una sola cadena), preferiblemente moléculas de ácido nucleico genómico, que puede ser moléculas fragmentadas, se expone bajo condiciones de hibridación a una pluralidad de sondas de oligonucleótidos, en el que la pluralidad de sondas de oligonucleótidos o amplicones derivados de dichas sondas están en solución, para capturar a partir de las moléculas de ácido nucleico de la muestra secuencias de ácido nucleico diana y separar las regiones no hibridadas del genoma o cualquier otro ácido nucleico de la muestra a partir de las secuencias diana hibridadas, en el que dicha separación comprende la captura a través de una porción de unión (por ejemplo, asociada a la sonda o amplicón derivados de la sonda) los complejos de hibridación que se encuentran en solución y lavar los complejos unidos separando de este modo las secuencias diana hibridadas de las secuencias hibridadas no diana inespecíficas (Figura 6).

La presente descripción proporciona métodos y sistemas para aislar una pluralidad de secuencias de ácidos nucleicos y reducir la complejidad de una muestra de ácido nucleico grande, tal como DNA genómico o muestra de RNA, biblioteca de DNAC o biblioteca de RNAm para facilitar su posterior procesamiento y análisis genético. En

algunas realizaciones, los métodos y sistemas comprenden la amplificación in situ de sondas de ácido nucleico inmovilizadas (preseleccionadas) en el que los amplicones derivados de la sonda comprenden una porción de unión. La captura de amplicones etiquetados, en solución, secuencias de ácidos nucleicos diana a partir de una muestra mediante la hibridación de la muestra con los amplicones utilizando métodos basados en solución. El complejo híbrido amplicón marcado / ácido nucleico diana es capturado a través de la porción de unión, preferiblemente lavada y el ácido nucleico diana eluido. Las secuencias genómicas eluidas son más susceptibles de análisis genético detallado que una muestra genómica que no haya sido sometida a este procedimiento de enriquecimiento. En consecuencia, los métodos descritos proporcionan una aproximación rentable, flexible y eficiente para reducir la complejidad de una muestra genómica. A lo largo del resto de la descripción, las muestras genómicas se utilizan para propósitos descriptivos, pero se entiende que otras muestras no genómicas podrían someterse a los mismos procedimientos.

En algunas realizaciones, la invención proporciona un método para aislar una pluralidad de secuencias de ácidos nucleicos y reducir la complejidad de una muestra de ácido nucleico mediante la hibridación de la muestra con amplicones de sondas de ácido nucleico en solución bajo condiciones preferiblemente rigurosas suficientes para soportar la hibridación entre la amplicones de la sonda y las regiones complementarias de la muestra de ácido nucleico. Los complejos de amplicón de sonda / ácido nucleico diana se lavan bajo condiciones suficientes para eliminar los ácidos nucleicos unidos de forma inespecífica. Las secuencias de ácidos nucleicos diana hibridados se eluyen de los amplicones derivados de sonda y pueden opcionalmente amplificarse posteriormente (por ejemplo, mediante LM-PCR), por ejemplo para aplicaciones posteriores tales como resecuenciación.

La presente descripción proporciona un método para aislar una pluralidad de secuencias de ácidos nucleicos y reducir la complejidad genética de una población de moléculas de ácido nucleico, comprendiendo el método los pasos de exponer moléculas de ácido nucleico fragmentadas, desnaturalizadas de una población diana a múltiples, amplicones derivados de sondas de oligonucleótido diferentes en la que los amplicones están en solución y en el que los amplicones comprenden además una porción de unión, en condiciones de hibridación para capturar moléculas de ácido nucleico que hibridan específicamente con los amplicones de la sonda, uniendo o capturando los complejos de moléculas hibridadas mediante la unión de la porción de unión encontrada en el amplicón de sonda a su pareja de unión (por ejemplo, biotina / SA, digoxigenina / anti-digoxigenina, 6HIS / níquel, etc.), en el que las moléculas de ácido nucleico, desnaturalizadas fragmentadas tienen un tamaño medio de aproximadamente 100 a aproximadamente 1000 residuos de nucleótidos, preferiblemente entre aproximadamente 250 a aproximadamente 800 residuos de nucleótidos y lo más preferiblemente entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos, separando los ácidos nucleicos no unidos e inespecíficamente hibridados de los amplicones de sonda unidos, eluyendo las moléculas diana hibridadas de los amplicones, y secuenciando opcionalmente las moléculas diana.

Como tal, algunas realizaciones proporcionan métodos basados en solución y sistemas para aislar una pluralidad de secuencias de ácidos nucleicos y reducir la complejidad genética de una población de moléculas de ácido nucleico. Los métodos y sistemas comprenden exponer, secuencias de muestras de ácidos nucleicos desnaturalizados fragmentados, que pueden comprender o no uno o más adaptadores de ligación en uno o en ambos extremos de la muestra de ácido nucleico fragmentado antes de la desnaturalización, en múltiples, amplicones de sondas de hibridación diferentes en solución en el que dichos amplicones derivan de diferentes sondas de hibridación múltiples, prediseñadas en el que dichos amplicones comprenden una porción de unión o secuencia y, opcionalmente, un sitio de endonucleasa de restricción (RE), en condiciones de hibridación suficientes para hibridar las secuencias de ácido nucleico diana desnaturalizadas con los amplicones derivados de sonda (por ejemplo, en solución), en el que las secuencias de ácidos nucleicos fragmentadas, desnaturalizadas tienen un tamaño medio de aproximadamente 100 a aproximadamente 1000 residuos de nucleótidos, preferiblemente tienen entre aproximadamente 250 a aproximadamente 800 residuos de nucleótidos y lo más preferiblemente entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos, separando los ácidos nucleicos no unidos e hibridados de forma inespecífica de los amplicones derivados de sonda mediante la unión de los complejos amplicón / objetivo a través de la porción de unión y lavando los complejos unidos, eluyendo las secuencias de ácidos nucleicos diana del complejo unido en el que la diana secuenciada demuestra una reducción de la complejidad genética en relación a la muestra original y opcionalmente repetir los pasos de hibridación, lavado y elución utilizando las secuencias de ácido nucleico diana iniciales enriquecidas eluidas para enriquecer aún más las secuencias de ácido nucleico diana.

En algunas realizaciones, las sondas de captura de ácidos nucleicos diana se inmovilizan sobre un sustrato mediante una variedad de métodos. En una realización, las sondas se pueden colocar en portaobjetos (por ejemplo, patentes de EE.UU. núms. 6.375.903 y 5.143.854). En realizaciones preferidas, las sondas se sintetizan in situ sobre un sustrato mediante el uso de sintetizadores de chips sin máscara (MAS) como se describe en la Patente de Estados Unidos N° 6.375.903, 7.037.659, 7.083.975, 7.157.229 que permite la síntesis in situ de secuencias de oligonucleótidos directamente en el portaobjetos para su posterior amplificación in situ mediante reacción en cadena de la polimerasa (PCR).

En algunas realizaciones, un soporte sólido es una población de cuentas o partículas. Las sondas de captura se sintetizan inicialmente en un portaobjetos de microchips utilizando un sintetizador de chips sin máscara, se amplifican, liberan o cortan de acuerdo con los métodos estándar, opcionalmente se amplifican e inmovilizan en

dicha población de cuentas. Las cuentas pueden empaquetarse, por ejemplo, en una columna de modo que una muestra diana se carga y se pasa a través de la columna y la hibridación de sonda / muestra diana tiene lugar en la columna, seguido de un lavado y elución de las secuencias de la muestra diana para reducir la complejidad genética. En algunas realizaciones, una columna posee puertos de entrada y de salida de fluidos. En algunas realizaciones, con el fin de mejorar la cinética de hibridación, la hibridación se lleva a cabo en una solución acuosa que comprende las cuentas con las múltiples sondas inmovilizadas en suspensión en un medio acuoso.

En algunas realizaciones, las sondas de ácidos nucleicos para moléculas diana se sintetizan en un soporte sólido, liberadas del soporte sólido como un conjunto de sondas y amplificadas. El conjunto amplificado de sondas liberadas están inmovilizadas de forma covalente o no covalente en un soporte (por ejemplo, vidrio, metal, cuentas de cerámica, polímeros, partículas paramagnéticas, etc.). Las sondas están diseñadas para la liberación conveniente desde el soporte sólido proporcionando, por ejemplo, una secuencia de ácido nucleico lábil al ácido o al álcali en o cerca del extremo proximal de la sonda que libera las sondas bajo condiciones de pH bajo o alto, respectivamente. La técnica está familiarizada con los métodos para la inmovilización de ácidos nucleicos sobre soportes, por ejemplo mediante la incorporación de un nucleótido biotinilado en las sondas y recubriendo un soporte con estreptavidina de tal manera que el soporte recubierto atrae e inmoviliza las sondas en el conjunto. La muestra o muestras pasan a través del soporte que contiene la sonda (por ejemplo, el portaobjetos, columna, etc.) bajo condiciones de hibridación tales que las moléculas de ácido nucleico diana que hibridan con el soporte inmovilizado pueden eluirse para el análisis posterior u otra utilización.

En algunas realizaciones, las sondas de hibridación iniciales diseñadas para la amplificación posterior para su uso en los métodos de captura basadas en solución como se describe en el presente documento, se imprimen o se depositan sobre un soporte sólido tal como un portaobjetos con microchips de DNA, chip, micropocillo, columna, tubo, cuentas o partículas. Los sustratos pueden ser, por ejemplo, vidrio, metal, cuentas de cerámica, poliméricos, etc. En realizaciones preferidas, el soporte sólido es un microchip (por ejemplo, portaobjetos de vidrio), en el que las sondas se sintetizan en el microchip utilizando un sintetizador de matriz sin máscara. Las longitudes de las múltiples sondas de oligonucleótidos pueden variar y dependen del diseño experimental y se limitan sólo por la posibilidad de sintetizar tales sondas. En realizaciones preferidas, la longitud media de la población de múltiples sondas antes de la amplificación in situ es de aproximadamente 20 a aproximadamente 100 nucleótidos, preferiblemente entre aproximadamente 40 a aproximadamente 85 nucleótidos, en particular entre aproximadamente 45 a aproximadamente 75 nucleótidos. Las sondas de hibridación inmovilizadas se utilizan posteriormente como moldes para la amplificación de PCR in situ y opcionalmente amplificación por PCR asimétrica proporcionando de ese modo amplicones derivados de sonda para hibridación basada en solución y enriquecimiento de las moléculas de ácido nucleico diana de una muestra compleja.

En algunas realizaciones, las sondas de hibridación corresponden en secuencia para al menos una región de un genoma y se pueden proporcionar en un soporte sólido utilizando en paralelo, por ejemplo la tecnología de síntesis de chips sin máscara (MAS). Alternativamente, las sondas se pueden obtener en serie utilizando un sintetizador de DNA estándar y luego aplicarse al soporte sólido o pueden obtenerse a partir de un organismo y a continuación, inmovilizarse sobre el soporte sólido. En algunas realizaciones, se contempla que las sondas de hibridación, con independencia del método de síntesis, comprenden secuencias de cebadores de amplificación para su uso en técnicas de amplificación. En algunas realizaciones, las secuencias de cebadores de amplificación incorporadas en secuencias de sondas de hibridación comprenden además secuencias de endonucleasas de restricción (RE). En algunas realizaciones, las sondas de hibridación como las que se encuentran sobre un sustrato de microchips se amplifican in situ usando cebadores complementarios a las secuencias de cebadores en el que uno o ambos de los cebadores comprenden además el enlazante químico, tal como una porción de unión (por ejemplo, biotina, digoxigenina, etc.) de modo que los amplicones de PCR derivados de las sondas de hibridación están en solución.

La solución que comprende los amplicones derivados de las sondas se transfieren a, por ejemplo, un tubo, pocillo, u otro recipiente y se mantienen en solución. Se contempla la realización adicional de una o más rondas de amplificación para aumentar la producción de la cadena de amplicón que comprende la porción de unión, por ejemplo, mediante PCR asimétrica. Una muestra de ácido nucleico, preferiblemente fragmentado y desnaturalizado para producir secuencias diana de cadena sencilla fragmentadas, se añade a los amplicones en solución y se permite que se produzca la hibridación entre los amplicones derivados de la sonda y la muestra de ácido nucleico diana monocatenario fragmentado. Después de la hibridación, los ácidos nucleicos que no hibridan o que se hibridan de forma inespecífica, se separan del complejo amplicón / diana mediante la captura del complejo amplicón / diana a través de la porción de unión y lavando el complejo amplicón / diana. Por ejemplo, si la porción de unión es biotina, un sustrato recubierto con estreptavidina se utiliza para capturar el complejo. El complejo unido se lava, por ejemplo con una o más soluciones de lavado. Los ácidos nucleicos restantes (por ejemplo, unidos específicamente a los amplicones) se eluyen a partir del complejo, por ejemplo, mediante el uso de agua o un tampón de elución (por ejemplo, que comprende tampón TRIS y / o EDTA) para producir un eluato enriquecido para las secuencias de ácido nucleico diana.

Los oligonucleótidos basados en microchips proporcionados para los métodos y sistemas de captura basados en la amplificación en solución, como se describe en el presente documento, están diseñados para dirigirse a una región o regiones de un genoma. La longitud de las sondas individuales está normalmente entre 50 y 200 bases. Estas

sondas pueden diseñarse para ser sondas solapantes, lo que significa que los nucleótidos de partida de sondas adyacentes están separados en el genoma por menos de la longitud de una sonda, o sondas no solapadas, en la que la distancia entre las sondas adyacentes es mayor que la longitud de una sonda. La distancia entre las sondas adyacentes se solapa generalmente, con una separación entre el nucleótido de partida de dos sondas que varían entre 1 y 100 bases. La distancia se varía para causar que algunas regiones genómicas sean localizadas por un mayor número de sondas que otras. Esta variación se utiliza, por ejemplo, para modular la eficiencia de la captura de regiones genómicas individuales, normalizando la captura. Las sondas pueden analizarse por su exclusividad en el genoma. En realizaciones preferidas de la presente invención, para evitar la unión inespecífica de elementos genómicos con amplicones derivados de sondas, los elementos altamente repetitivos del genoma se excluyen de los diseños de selección de sondas usando un método que utiliza una estrategia similar para el programa WindowMasker desarrollado por, por ejemplo, Morgolis (2006, Bioinformatics 15: 134-141,) para identificar estas regiones y excluirlas del diseño de las sondas.

La naturaleza y el rendimiento de las sondas diseñadas para la amplificación para los métodos de captura basados en solución de la presente invención se pueden variar para normalizar ventajosamente o ajustar la distribución de las moléculas diana capturadas y enriquecidas de acuerdo con los métodos de la presente invención. Un objetivo de dicha normalización es entregar un gen expresado por lectura (por ejemplo, Soares, et al., 1994, Proc. Natl. Acad. Sci. 91: 9228-9232). La normalización se aplica, por ejemplo, a las poblaciones de moléculas de DNAC antes de la construcción de la biblioteca ya que generalmente la distribución de las moléculas en la población refleja los diferentes niveles de expresión de genes expresados a partir de los cuales se producen las poblaciones de moléculas de DNAC. Por ejemplo, el número de reacciones de secuenciación necesarias para analizar de forma eficaz cada región diana se reduce al normalizar el número de copias de cada secuencia diana en la población enriquecida de tal manera que a través del grupo de sondas el rendimiento de captura de diferentes sondas se normaliza, sobre la base de una combinación del grado de adecuación y otros atributos de la sonda.

El grado de adecuación, caracterizado por una métrica de captura, se determina ya sea informáticamente o empíricamente. En una aproximación, la capacidad de las moléculas diana para unirse se ajusta proporcionando los denominados sondas de oligonucleótidos isotérmicos ( $T_m$ -equilibrado), como se describe en la Publicación de Patente N ° 2005/10282209, que permite un rendimiento uniforme de la sonda, elimina artefactos de hibridación y / o sesgos y proporciona resultados de mayor calidad. Las longitudes de sonda se ajustan (normalmente, entre aproximadamente 20 a aproximadamente 100 nucleótidos, preferiblemente entre aproximadamente 40 a aproximadamente 85 nucleótidos, en particular entre aproximadamente 45 a aproximadamente 75 nucleótidos, pero opcionalmente también entre más de 100 nucleótidos hasta aproximadamente 250 nucleótidos) para igualar la temperatura de fusión (por ejemplo,  $T_m = 76$  °C, típicamente entre aproximadamente 55 °C a aproximadamente 76 °C, en particular entre aproximadamente 72 °C a aproximadamente 76 °C) a través de todo el conjunto de sondas antes de la amplificación. Por lo tanto, las sondas están optimizadas para actuar de manera equivalente a una rigurosidad determinada en las regiones genómicas de interés, incluyendo las regiones ricas en AT y GC. El experto en la materia apreciará que la longitud de la sonda, la temperatura de fusión y la secuencia se puede ajustar de forma coordinada para cualquier amplicón derivado de sonda determinado para llegar a un rendimiento de hibridación deseado para el amplicón de la sonda. Por ejemplo, la temperatura de fusión ( $T_m$ ) del amplicón derivado de la sonda se puede calcular utilizando la fórmula:  $T_m = S_x (G_n + C_n) + I_x (A_n + T_n)$ , donde n es el número de cada base específica (A, T, G o C) presentes en el amplicón de la sonda.

El rendimiento de captura también se puede normalizar mediante la determinación del grado de adecuación de captura de los amplicones de la sonda en el conjunto de sondas, y en consecuencia ajustar después la cantidad de sondas individuales en el soporte sólido para los propósitos de amplificación. Por ejemplo, si los amplicones de sonda derivados de una primera sonda se prevé que capturen veinte veces más ácido nucleico que un segundo conjunto de amplicones derivados de sonda, entonces el rendimiento de captura de ambos conjuntos de amplicones de sonda pueden igualarse al proporcionar veinte veces más copias de la segunda sonda con propósitos de amplificación, por ejemplo, aumentando veinte veces el número mostrado de la segunda sonda en los microchips de sondas antes de la amplificación.

En otras realizaciones, una estrategia adicional para la normalización de captura de ácidos nucleicos diana es someter a las moléculas diana eluidas a una segunda ronda de hibridación basada en solución contra los amplicones derivados de sonda en condiciones menos rigurosas que las que se utilizaron para la primera ronda de hibridación. Aparte de el enriquecimiento sustancial en la primera hibridación que reduce la complejidad relativa del ácido nucleico genómico original, la segunda hibridación puede llevarse a cabo bajo condiciones de hibridación que saturan todas las sondas de captura. Suponiendo que se proporcionan en solución sustancialmente la misma cantidad amplicones derivados de la sonda, la saturación de los amplicones asegurará que cantidades sustancialmente iguales de cada diana se eluyan tras la segunda hibridación y el lavado.

En algunas realizaciones, los cebadores de amplificación utilizados para la amplificación in situ de las sondas de hibridación para los posteriores métodos y sistemas de enriquecimiento y captura basados en solución descritos en el presente documento, comprenden la químico enlazante tales como las porciones de unión. Las porciones de unión comprenden cualquier porción que está unida o incorporada en el extremo 5' de un cebador de amplificación útil en la posterior captura del complejo de hibridación de amplicón de sonda / diana de ácido nucleico. Una porción

de unión es cualquier secuencia que está diseñada en el extremo 5' de una secuencia de cebador, tal como una secuencia de 6 histidinas (6HIS) que es capturable. Por ejemplo, un cebador que comprende una secuencia de 6HIS es capturable por el níquel, por ejemplo en un tubo, micropocillo, o columna e purificación que está recubierto con níquel o contiene cuentas recubiertas de níquel, partículas, etc., donde las cuentas están empaquetadas en una columna y se carga una muestra y se pasa a través de la columna para la captura del complejo para reducir la complejidad (por ejemplo, y la posterior elución de la diana). Otro ejemplo de una porción de unión útil en realizaciones de la presente invención incluye un hapteno, por ejemplo digoxigenina que está, por ejemplo, unida al extremo 5' de un cebador de amplificación. La digoxigenina es capturable mediante el uso de un anticuerpo dirigido contra digoxigenina, por ejemplo un sustrato que está revestido o contiene un anticuerpo anti-digoxigenina.

En realizaciones preferidas, un cebador de amplificación utilizado en los métodos y sistemas de la presente invención contiene una porción de biotina unida al extremo 5' del cebador y el posterior amplicón derivado de sonda. La biotina es capturable por estreptavidina (SA), como tal, el amplicón marcado con biotina puede capturarse sobre un sustrato o columna que está revestido o contiene SA. En realizaciones preferidas, la estreptavidina se recubre sobre partículas paramagnéticas que pueden a su vez capturarse magnéticamente para facilitar el lavado y elución de los ácidos nucleicos enriquecidos de diana. La presente descripción no está limitada por el tipo de enlazante químico usado, y un experto en la materia conocerá otras opciones que son igualmente modificables para los métodos y sistemas de la presente descripción.

En algunas realizaciones, los métodos y sistemas comprenden determinar la información de la secuencia de ácido nucleico de al menos una región de ácido nucleico, en particular el ácido nucleico genómico, (todo el genoma o al menos un cromosoma entero o parcial) en una muestra, comprendiendo el método los pasos de realizar los métodos como se ha descrito anteriormente seguido por la determinación de la secuencia de ácido nucleico de las moléculas capturadas, en particular, mediante la realización de secuenciación por reacciones de síntesis.

En algunas realizaciones los ácidos nucleicos diana son normalmente ácidos desoxirribonucleicos o ácidos ribonucleicos, e incluyen productos sintetizados in vitro mediante la conversión de un tipo de molécula de ácido nucleico (por ejemplo, DNA, RNA y DNAc) a otro, así como moléculas sintéticas que contienen análogos de nucleótidos. Las moléculas de DNA genómico desnaturalizadas son en particular, moléculas que son más cortas que las moléculas de ácido nucleico genómico que aparecen de forma natural. Un experto en la materia puede producir moléculas de tamaño aleatorio o no aleatorio a partir de moléculas más grandes mediante fragmentación química, física o escisión enzimática o utilizando protocolos bien conocidos. Por ejemplo, la fragmentación química puede emplear metales ferrosos (por ejemplo, Fe-EDTA), los métodos físicos pueden incluir sonicación, fuerza hidrodinámica o nebulización (por ejemplo, véase la solicitud de patente europea EP 0 552 290) y los protocolos enzimáticos puede emplear nucleasas tales como nucleasa micrococcal (Mnasa) o exo-nucleasas (tales como Bal31 o Exo1) o endonucleasas de restricción.

La presente descripción no se limita al método en el que se generan fragmentos y contempla cualquier método útil en la fragmentación de ácidos nucleicos. En realizaciones de la presente invención, se prefieren los fragmentos enriquecidos en un intervalo de tamaño compatible con la tecnología post-enriquecimiento. Por ejemplo, las realizaciones de la presente invención contemplan tamaños de fragmentos de ácido nucleico en el intervalo de entre aproximadamente 100 y aproximadamente 1000 residuos de nucleótidos o pares de bases, o entre aproximadamente 250 y aproximadamente 800 residuos de nucleótidos o pares de bases, o entre aproximadamente 400 a aproximadamente 600 residuos de nucleótidos o pares de bases, en particular alrededor de 500 residuos de nucleótidos o pares de bases.

La población de moléculas de ácido nucleico que puede comprender las secuencias de ácido nucleico diana contiene preferiblemente todo el genoma o al menos, un cromosoma de un organismo, o al menos una molécula de ácido nucleico con al menos alrededor de 100 kb. En particular, el tamaño(s) de la molécula(s) de ácido nucleico es/son, al menos, de alrededor de 200 kb, al menos alrededor de 500 kb, al menos alrededor de 1 Mb, al menos alrededor de 2 Mb o al menos alrededor de 5 Mb, especialmente con un tamaño de entre aproximadamente 100 kb y aproximadamente 5 Mb, de entre aproximadamente 200 kb y aproximadamente 5 Mb, de entre aproximadamente 500 kb y aproximadamente 5 Mb, de entre alrededor de 1 Mb y alrededor de 2 Mb, o entre aproximadamente 2 Mb y aproximadamente 5 Mb. En algunas realizaciones, las moléculas de ácido nucleico son DNA genómico, mientras que en otras realizaciones las moléculas de ácido nucleico son cDNA, o especies de RNA (por ejemplo, tRNA, mRNA, miRNA).

En algunas realizaciones, las moléculas de ácido nucleico que pueden o no comprender las secuencias de ácido nucleico diana se pueden seleccionar de un animal, una planta o un microorganismo, y en realizaciones particulares, las moléculas de ácido nucleico son de un primate, preferiblemente un ser humano. En algunas realizaciones, si hay muestras limitadas de las moléculas de ácido nucleico, se amplifican los ácidos nucleicos (por ejemplo, por amplificación del genoma completo) antes de poner en práctica el método de la presente invención. Por ejemplo, puede ser necesaria una amplificación previa para llevar a cabo realizaciones de la presente invención con fines forenses (por ejemplo, en medicina forense, etc).

Se contempla que en las realizaciones preferibles, la población de moléculas de ácido nucleico es una población de moléculas de DNA genómico. Las sondas de hibridación y los posteriores amplicones pueden comprender una o más secuencias dirigidas frente a una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos, la secuencia completa de al menos un locus genético único, dicho locus con un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb, o al menos uno de los tamaños especificados anteriormente, puntos en los que se conoce contienen SNP, o secuencias que definen un chip (del inglés "array"), en particular un chip de secuencias contiguas, diseñada para capturar la secuencia completa de al menos un cromosoma completo.

Se contempla que las secuencias de ácidos nucleicos diana se enriquezcan a partir de una o más muestras que incluyan ácidos nucleicos de cualquier fuente, de forma purificada o no purificada. La fuente no necesariamente debe contener el complemento completo de moléculas de ácido nucleico genómico de un organismo. La muestra, preferiblemente de origen biológico, incluye, pero no se limita a, grupos de aislamientos de pacientes individuales, muestras de tejido o cultivos celulares. La región diana puede ser uno o más bloques continuos de varias megabases, o varias regiones contiguas o no contiguas de menor tamaño, como todos los exones de uno o más cromosomas, o de los puntos que se conoce contienen SNP. Por ejemplo, las sondas de hibridación y los amplicones derivados de la sonda a continuación pueden complementar un chip de secuencias contiguas diseñado para capturar uno o más cromosomas completos, partes de uno o más cromosomas, todos los exones, todos los exones de uno o más cromosomas, exones seleccionados, intrones y exones para uno o más genes, regiones reguladoras de genes, y así sucesivamente.

Alternativamente, para aumentar la probabilidad de que las dianas deseadas no únicas o de difícil de captura se enriquecen, las sondas pueden ser dirigidas a secuencias asociadas con la secuencia diana real (por ejemplo, en el mismo fragmento, pero separada), en cuyo caso se capturarán y se enriquecerán fragmentos genómicos que contienen tanto el objetivo deseado como las secuencias asociadas. Las secuencias asociadas pueden ser adyacentes o estar separadas de las secuencias diana, pero un experto en la materia apreciará que cuanto más cerca estén ambas partes entre sí, más probable es que los fragmentos genómicos contengan ambas porciones. Para reducir el impacto limitado de la hibridación cruzada con moléculas no diana, mejorando así la integridad del enriquecimiento, se realizan rondas de captura secuencial utilizando conjuntos de sondas de captura, y por lo tanto amplicones derivados de las sondas, distintos pero relacionados, dirigidos a la región diana. Las sondas relacionadas son sondas correspondientes a regiones de gran proximidad entre sí en el genoma y que se hibridan con el mismo fragmento de DNA genómico.

En algunas realizaciones, los métodos comprenden el paso de ligación de moléculas adaptador o enlazantes en uno o ambos extremos de las moléculas de ácido nucleico antes de la desnaturalización y la hibridación con los amplicones sonda en solución.

En algunas realizaciones, los métodos comprenden además amplificar dichas moléculas de ácido nucleico adaptador modificadas con al menos un cebador, y dicho cebador comprende una secuencia que se hibrida específicamente con la secuencia de dicha(s) molécula(s) adaptador.

En algunas realizaciones se proporcionan enlazantes bicatenarios en uno o ambos extremos de las moléculas de ácido nucleico fragmentado antes de la desnaturalización de la muestra y la hibridación a los amplicones derivados de la sonda en solución. En tales realizaciones, se amplifican moléculas de ácido nucleico diana después de la elución para producir un conjunto de productos amplificados con una complejidad menor respecto a la muestra original. Las moléculas de ácidos nucleicos diana pueden amplificarse utilizando, por ejemplo, PCR mediada por ligación no específica (LM-PCR) mediante múltiples rondas de amplificación, y los productos pueden enriquecerse aún más, si es necesario, mediante una o más rondas de selección con las sondas derivadas del amplicón. Los enlazantes o adaptadores se proporcionan, por ejemplo, de un tamaño arbitrario y con una secuencia de ácido nucleico arbitraria de acuerdo con las aplicaciones analíticas posteriores que se desee realizar tras el paso de reducción de la complejidad. Los enlazantes pueden variar entre aproximadamente 12 y aproximadamente 100 pares de bases, incluyendo un intervalo de aproximadamente 18 y 100 pares de bases, y preferiblemente entre aproximadamente 20 y 24 pares de bases. Las moléculas adaptadoras en el contexto de la presente invención se definen preferiblemente como oligonucleótidos de doble cadena de extremos romos.

Con el fin de ligar las moléculas adaptadoras a una molécula diana de doble cadena, es preferible que esta molécula diana tenga los extremos romos. Para lograrlo, las moléculas diana de doble cadena son sometidas a, por ejemplo, una reacción de rellenado con una polimerasa de DNA, como la polimerasa de DNA de T4 o la polimerasa Klenow en presencia de dNTP, lo que resulta en moléculas diana de extremos romos. Además, los extremos de los fragmentos se fosforilan utilizando la quinasa de polinucleótidos de T4 y los métodos conocidos por los expertos en la materia (por ejemplo, véase *Molecular Cloning: A Laboratory Manual*, Ed. Sambrook et al., Cold Spring Harbour Press); para añadir grupos fosfato en los extremos 5' de los fragmentos antes de la ligación de los adaptadores. La ligación de los adaptadores (por ejemplo, oligonucleótidos cortos de doble hebra de DNA de extremos romos con alrededor de 3-20 pares de bases) en el DNA diana fosforilado y pulido puede realizarse a continuación de acuerdo con cualquier método conocido en la materia, por ejemplo mediante la reacción de la ligasa de DNA de T4.

La ligación de los adaptadores a las moléculas de ácidos nucleicos diana fragmentados se pueden realizar antes o después de la exposición de una muestra que comprende moléculas de ácido nucleico genómico fragmentado desnaturizado a múltiples amplicones sonda de oligonucleótidos en solución, bajo condiciones de hibridación, para capturar las moléculas de ácido nucleico diana. Cuando la ligación se realiza después de la hibridación, los ácidos nucleicos enriquecidos que se liberan de los amplicones en forma de cadena sencilla inicialmente se rehibridan, seguido de una reacción de extensión del cebador y una reacción de relleno de acuerdo con los métodos estándar conocidos en la materia.

La ligación de moléculas adaptadoras permite un paso de posterior amplificación de las moléculas capturadas. Independientemente de si la ligación tiene lugar antes o después del paso de captura, existen varias formas de realización alternativas. En una realización, un tipo de molécula adaptadora (por ejemplo, una molécula adaptadora A) se liga de forma que resulta en una población de fragmentos con secuencias terminales idénticas en ambos extremos del fragmento. Como consecuencia, es suficiente utilizar sólo un cebador en un potencial paso de amplificación subsiguiente. En una realización alternativa, se utilizan dos tipos de moléculas adaptadoras A y B. Esto resulta en una población de moléculas enriquecidas compuesta de tres tipos diferentes: (i) fragmentos que tienen un adaptador (A) en un extremo y otro adaptador (B) en el otro extremo, (ii) fragmentos que tienen adaptadores A en ambos extremos, y (iii) fragmentos que tienen adaptadores B en ambos extremos. La generación de moléculas enriquecidas con adaptadores es una ventaja excepcional, si la amplificación y secuenciación se va a realizar, por ejemplo, utilizando el instrumento 454 Life Sciences Corporation GS20 y GSFLX (por ejemplo, véase el Manual GS20 Library Prep, dic 2006, WO 2004/070007).

La presente descripción se refiere a un método para detectar las variaciones de la región(es) codificante(s) de una muestra de genoma a ensayar, con respecto a una muestra de referencia del genoma, en particular, en relación con un genoma de referencia que comprende, moléculas de ácido nucleico genómico fragmentado desnaturizado, el método que comprende las etapas descritas anteriormente tanto en el genoma de prueba como de referencia, comparando también las secuencias con una secuencia en una base de datos, en particular, con una secuencia de una base de datos de polimorfismos en una muestra de genoma de referencia para identificar variantes en una muestra del genoma de prueba. La descripción es, por lo tanto, útil para la búsqueda de variantes y mutaciones genéticas, tales como polimorfismos de nucleótido único (SNP), o un conjunto de SNP, inserciones y/ o deleciones genómicas, translocaciones, etc., que pueden ser la base enfermedades humanas. Se contempla que la tecnología de de captura y enriquecimiento basada en la hibridación en solución utilizada como aquí se describe es más flexible que otros métodos disponibles actualmente en el área del enriquecimiento genómico.

En algunas realizaciones, las secuencias de ácidos nucleicos diana eluidas se pueden secuenciar, hibridarse a un chip de re-secuenciación o de llamada de SNP, y la secuencia o genotipo pueden analizarse posteriormente. El enriquecimiento en solución como se proporcionan por las formas de realización de la presente invención permite los métodos de secuenciación dirigida en chip, expansiva (del inglés *shotgun*), capilar u otros métodos de secuenciación conocidos en la materia. En general, las estrategias de secuenciación expansiva de fragmentos generados aleatoriamente son rentables y se integran fácilmente en un proyecto. La presente descripción mejora la eficiencia de la aproximación de secuenciación expansiva al presentar sólo fragmentos de una o más regiones genómicas de interés para la secuenciación. La descripción proporciona una capacidad de enfocar las estrategias de secuenciación sobre regiones genómicas específicas, tales como cromosomas individuales o exones, con propósitos de secuenciación médicos. Como tal, se realiza un enfoque más centrado en el descubrimiento de enfermedades.

En algunas realizaciones, las secuencias de ácidos nucleicos diana eluidas resultantes de los métodos de enriquecimiento en solución como los que se describen en este documento, son posteriormente secuenciadas. La secuenciación se puede realizar mediante una serie de métodos diferentes, como con el empleo de secuenciación mediante tecnología de síntesis. La secuenciación por síntesis de acuerdo con el conocimiento previo se define como cualquier método de secuenciación que monitoriza la generación de productos secundarios tras la incorporación de un desoxinucleósido trifosfato específico durante la reacción de secuenciación (Hyman, 1988, Anal Biochem 174: 423-436; Rhonaghi et al, 1998, Science 281: 363-365). Una realización importante de la secuenciación mediante reacción de síntesis es el método de secuenciación con pirofosfato. En este caso, la generación de pirofosfato durante la incorporación de nucleótidos se monitoriza mediante una cascada enzimática que resulta en la generación de una señal quimioluminiscente. El secuenciador 454 Genome Sequencer System (Roche Applied Science nº cat. 04 760 085 001), un ejemplo de secuenciación por síntesis, se basa en la tecnología de secuenciación con pirofosfato. Para la secuenciación en un instrumento FLX 454 GS20 o 454, el tamaño medio de fragmento de DNA genómico se encuentra en el rango entre 200 o 600 pb, respectivamente, como se describe en la literatura del producto.

Una secuenciación mediante reacción de síntesis, alternativamente, puede estar basada en un tipo de colorante de terminación de la reacción de secuenciación. En este caso, los bloques de construcción desoxinucleotрифосфatos (ddNTP) colorantes incorporados comprenden un marcador detectable, que es preferiblemente un marcador fluorescente que impide una mayor extensión de la cadena de DNA naciente. Luego se retira la señal y se detecta tras la incorporación del bloque de construcción ddNTP en el híbrido de extensión de molde/ cebador, por ejemplo, mediante el uso de una polimerasa de DNA que comprende una actividad exonucleasa 3'-5' o de corrección de pruebas.

En el caso del flujo de trabajo del Genome Sequencer (Roche Applied Science N° Catálogo 04 896 548 001), en una primera etapa (clonal) se lleva a cabo la amplificación mediante una PCR en emulsión. Por lo tanto, esto también está dentro del alcance de la presente invención, porque el paso de amplificación se realiza por métodos de PCR en emulsión. Las cuentas que son portadoras de los ácidos nucleicos diana amplificados de forma clonal puede entonces transferirse arbitrariamente a una placa picotitulada de acuerdo con el protocolo del fabricante y someterse a una reacción de secuenciación pirofosfato para determinación de la secuencia.

En algunas realizaciones, la presente descripción comprende un equipo que comprende reactivos y materiales para llevar a cabo los métodos de acuerdo con la presente invención. Tal equipo puede incluir uno o más sustratos de microchip sobre los que se inmovilizan una pluralidad de sondas de hibridación específicas de una o más secuencias de ácidos nucleicos diana a partir de uno o más loci genéticos de destino (por ejemplo, específicos de exones, intrones, secuencias de SNP, etc.), una pluralidad de sondas que definen un chip de secuencias contiguas diseñado para capturar la secuencia completa de al menos un cromosoma completo, los cebadores de amplificación, los reactivos para realizar los métodos de reacción en cadena de la polimerasa (por ejemplo, soluciones salinas, polimerasas, dNTP, tampones de amplificación, etc.), los reactivos para realizar las reacciones de ligación (por ejemplo, adaptadores de la ligación, quinasa de polinucleótidos de T4, ligasa, tampones, etc.), los sustratos que comprenden una porción pareja de unión, tubos, soluciones de hibridación, soluciones de lavado, soluciones de elución, imán (s), y el tubo titulares.

En algunas realizaciones, la presente descripción proporciona un sistema (por ejemplo, un equipo) para realizar un método o una parte de un método de acuerdo con la presente invención como se describe aquí. Por lo tanto, la presente descripción es un equipo que comprende una (primera) molécula adaptadora bicatenaria y múltiples amplicones derivados de la sonda en solución, en el que los amplicones derivados de la sonda se amplifican a partir de una pluralidad de sondas que define una pluralidad de exones, intrones y/ o secuencias reguladoras a partir de una pluralidad de loci genéticos; y/ o una pluralidad de amplicones derivados de sonda en solución que define la secuencia completa de al menos un locus genético único, y dicho locus que tiene un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb o un tamaño como el que se especifica en el presente documento, y/ o una pluralidad de amplicones derivados de sondas que definen puntos que se conoce contienen SNP, y/ o una pluralidad de amplicones derivados de una sonda que define un chip, en particular un chip de secuencias contiguas especialmente diseñado para capturar la secuencia completa de al menos un cromosoma completo. En algunas formas de realización, un equipo comprende además dos moléculas adaptadoras diferentes de doble cadena.

En algunas realizaciones, un equipo comprende una o más moléculas o compuestos de captura. Por ejemplo, al menos una sonda de oligonucleótido comprende una modificación que permite la inmovilización sobre un soporte sólido. Por ejemplo, una sonda comprende una porción biotina para su inmovilización sobre una partícula paramagnética recubierta con estreptavidina. Otro ejemplo es un hapteno, tal como la digoxigenina, que se asocia con una sonda para la inmovilización sobre un soporte sólido utilizando un anticuerpo que reconozca el hapteno (por ejemplo, anti-digoxigenina).

En algunas realizaciones, un equipo comprende además al menos uno o más compuestos a partir de un grupo que consiste en una polimerasa de DNA, una quinasa de polinucleótidos de T4, ligasa de DNA de T4, una o más soluciones de hibridación del chip, y/ o uno o más soluciones de lavado del chip. En realizaciones preferibles, tres soluciones de lavado se incluyen en un equipo de la presente invención, y estas soluciones de lavado comprenden SSC, DTT y opcionalmente SDS. Por ejemplo, los equipos de la presente invención comprenden el tampón de lavado I (SSC al 0,2%, SDS al 0,2% (v/v), DTT 0.1 mM), Tampón de lavado II (SSC al 0,2%, DTT 0,1 mM,) y/ o tampón de lavado III (SSC al 0,05%, DTT 0,1 mM). En algunas realizaciones, los sistemas de la presente invención comprenden además una solución de elución, por ejemplo agua o una solución que contenga tampón TRIS y/ o EDTA.

Los siguientes ejemplos se proporcionan como ilustraciones adicionales no limitantes de las realizaciones particulares de la invención.

## EXPERIMENTACIÓN

Los siguientes ejemplos se proporcionan para demostrar e ilustrar adicionalmente ciertas realizaciones y aspectos de la presente invención, y no deben interpretarse como limitantes del alcance de la misma.

Cuando se proporcione un rango de valores, se entiende que está englobado específicamente dentro de la invención cada valor intermedio, hasta la décima parte de la unidad del límite inferior a menos que el contexto indique claramente lo contrario, y también entre los límites superior e inferior de ese rango. Cada uno de los rangos menores entre cualquier valor establecido o valor que interviene en un intervalo indicado, y cualquier otro valor establecido o valor intermedio en ese intervalo establecido están incluidos en la invención. Los límites superior e inferior de estos intervalos menores pueden independientemente ser incluidos o excluidos del rango, ya cada campo donde sea, ni bien ambos límites se incluyen en los rangos más pequeños también se engloban dentro de la invención, sujetos a

cualquier límite específicamente excluido en el intervalo indicado. Cuando el intervalo indicado incluye uno o ambos de los límites, los intervalos que excluyen uno o ambos de esos límites incluidos también se incluyen en la invención.

Ejemplo 1 - Descubrimiento de nuevos polimorfismos y mutaciones en las grandes regiones genómicas

Este ejemplo genérico describe cómo llevar a cabo la selección que permite el descubrimiento rápido y eficiente de nuevos polimorfismos y mutaciones en grandes regiones genómicas. Se utilizan microchips que tienen sondas inmovilizadas, en una o múltiples rondas de hibridación de selección con una diana de DNA genómico total, y las secuencias seleccionadas se amplifican mediante LM-PCR (véanse las Figs. y 2).

a) Preparación del DNA genómico y de los enlazantes de doble cadena

Se fragmenta el DNA mediante sonicación hasta un tamaño medio de 500 pares de bases. Una reacción para pulir los extremos de los fragmentos de DNA sonicados se realizó con:

fragmentos de DNA	41 µl
polimerasa de DNA de T4	20 µl
mezcla de reacción de la polimerasa de DNA de T4	20 µl
agua	10 µl

La reacción se incubó a 11°C durante 30 min. La reacción se sometió entonces a procedimientos de extracción con fenol/cloroformo y el DNA se recuperó mediante una precipitación con etanol. El sedimento precipitado se disolvió en 10 µl de agua (para proporcionar una concentración final de 2 µg/ µl).

Dos oligonucleótidos complementarios se hibridan para crear un enlazante de doble cadena, mezclando lo siguiente:

oligonucleótido 1 (1 µg/µl)	22,5 µl
(5'-CTCGAGAATTCTGGATCCTC-3 ') (Id. de Sec. N°: 1)	
oligonucleótido 2 (1 µg/µl)	22,5 µl
(5'-GAGGATCCAGAATTCTCGAGTT-3 ') (Id. de Sec. N°: 2)	
tampón de rehibridación 10x	5 µl
agua hasta	50 µl

La reacción se calentó a 65°C durante 10 min; luego se dejó enfriar a 15-25°C durante 2 horas. La longitud de los dos oligonucleótidos complementarios 1 y 2 fue de entre 12 y 24 nucleótidos, y la secuencia se seleccionó dependiendo de la funcionalidad deseada por el usuario. El enlazante de doble cadena se purificó a continuación por cromatografía en columna a través de una columna de centrifugación de Sephadex G-50. La solución de enlazante purificada se concentró a continuación mediante liofilización a una concentración de 2 µg/ µl.

b) ligación de enlazantes a los fragmentos de DNA genómico

Se configuró la siguiente reacción para ligar los enlazantes a los fragmentos de DNA genómico. La reacción se incubó a 14°C durante toda la noche.

Enlazantes hibridados del paso a) (20 µg)	10 µl
DNA genómico del paso a) (10 µl)	5 µl
DNA ligasa de T4	10 U
Tampón de ligación 10x	2 µl
Agua hasta	20 µl

El volumen de reacción se ajustó a 500 µl con agua y se ligó el DNA genómico se purificó usando un equipo de purificación de PCR QIAquick. El DNA purificado se almacenó a una concentración de 1 µg/ µl.

c) selección primaria y captura de híbridos

Para preparar la muestra de DNA genómico para su hibridación con el microchip, se resuspendió el DNA genómico modificado enlazante (10 µg) en 3,5 µl de agua libre de nucleasas y se combinó con 31,5 µl de tampón de hibridación NimbleGen (Roche NimbleGen, Inc., Madison, WI), 9 µl de aditivo de hibridación (Roche NimbleGen, Inc), en un volumen final de 45 µl. Las muestras se desnaturalizaron con calor a 95°C durante 5 minutos y se transfirieron a un bloque de calor a 42°C.

Para capturar el DNA genómico diana en el microchip, las muestras se hibridaron con chips de CGH de NimbleGen, fabricados como se describe en la US 6.375.903 (Roche NimbleGen, Inc.). La fabricación sin máscara de oligonucleótidos de captura en los microchips se realizó mediante síntesis de oligonucleótidos dirigida por luz utilizando un microespejo digital como se describe en Singh-Gasson et al. (1999, Nat Biotech 17: 974 a 978, tal y como se realiza con un sintetizador de chips sin máscara. El análisis de la expresión génica usando chips de

## ES 2 522 912 T3

5 oligonucleótidos producidos mediante fotolitografía sin máscara se describe en Nuwaysir et al (2002, Genome Res 12: 1749-1755, que se incorpora aquí como referencia en su totalidad). La hibridación se realizó en un sistema de hibridación MAUI (BioMicro Systems, Inc., Salt Lake City, UT) de acuerdo con las instrucciones del fabricante durante 16 horas a 42°C utilizando el modo de mezcla B. Tras la hibridación, los chips se lavaron dos veces con tampón de lavado I (0,2 x SSC, SDS al 0,2% (v/v), DTT 0,1 mM, NimbleGen Systems) un total de 2,5 minutos. Los chips se lavaron a continuación durante 1 minuto, en tampón de lavado II (0,2 x SSC, DTT 0,1 mM, NimbleGen Systems) seguido de un lavado de 15 segundos en tampón de lavado III (SSC 0,05 x, DTT 0,1 mM, Roche NimbleGen, Inc.).

10 Para eluir el DNA genómico hibridado al microchip, los chips se incubaron dos veces durante 5 minutos en agua a 95°C. El DNA eluido se secó y separó usando centrifugación al vacío.

d) amplificación del DNA seleccionado inicialmente

15 El DNA genómico seleccionado inicialmente se amplificó tal como se describe a continuación. Se realizaron diez réplicas independientes de la reacción de amplificación en tubos de PCR de 200 µl. Sólo se requiere un cebador oligonucleótido, ya que cada fragmento tiene el mismo enlazante ligado a cada extremo:

Reactivos de reacción:

20 molde: selección primaria	
material	5 µl
oligonucleótido 1 (200 ng/µl)	1 µl
(5'-CTCGAGAATTCTGGATCCTC-3') (Id. de Sec. N°: 1)	
dNTP (25 mM cada uno)	0,4 µl
25 Tampón de reacción de la polimerasa de DNA Pfu HF Ultra 10x	5 µl
polimerasa de DNA Pfu HF Ultra 10x	2,5 U
agua hasta	50 µl

30 Las reacciones se amplificaron de acuerdo con el siguiente programa:

Número de ciclo	Desnaturalización	Hibridación	Polimerización
1	2 min. a 95°C		
2-31	30 s a 95°C	30 s a 55°C	1 min. a 72°C

35 Los productos de reacción se analizaron por electroforesis en gel de agarosa. Los productos de amplificación se purificaron usando un equipo de purificación de PCR QIAquick. Las muestras eluidas se agruparon y la concentración del DNA seleccionado primario amplificado se determinó por espectrofotometría. Un volumen de DNA en las muestras agrupadas equivalente a 1 µg se reduce a 5 µl en un concentrador de velocidad en vacío. Se reservó 1 µl (al menos 200 ng) de la materia prima seleccionada para la comparación con los productos de selección secundaria. Si es necesario, se realizan subsiguientes rondas de enriquecimiento mediante adicionales de hibridación del chip y amplificación de la muestra eluida.

45 e) preparación de sondas de oligonucleótidos diana para su liberación de los microchips y su inmovilización en un soporte

Las sondas se sintetizaron sobre un microchip, luego se liberaron mediante un grupo base lábil Fmoc (9-fluorenilmetiloxycarbonilo). Las sondas se marcaron con biotina y después se inmovilizaron sobre la superficie de un soporte sólido de estreptavidina utilizando métodos conocidos para la unión covalente o no covalente.

50 Opcionalmente, antes de la inmovilización sobre el soporte sólido, las sondas sintetizadas se amplificaron usando LM-PCR, Phi29 u otra estrategia de amplificación para aumentar la cantidad de las sondas sintetizadas mediante la inserción de secuencias entre ellos que facilitan su amplificación. Este material se puede utilizar entonces para la secuenciación directa, resecuenciación basada en chips, genotipificación o cualquier otro análisis genético dirigido a la región del genoma enriquecida mediante el empleo de una hibridación en fase de solución y una captura mediada por SA de los productos de hibridación.

Ejemplo 2 - Resecuenciación dirigida por un chip

60 Una serie de microchips de oligonucleótidos de alta densidad que capturan segmentos cortos que corresponden a 6.726 regiones exónicas de genes individuales de al menos 500 pares de bases se escogieron a partir de 660 genes distribuidos a lo largo del genoma humano (construcción de secuencia HG17) (aproximadamente 5 Mb de secuencia total) y se sintetizaron de acuerdo con los protocolos de fabricación de microchips habituales de Roche Nimble-Gen, Inc. La superposición de sondas de microchip de más de 60 bases cada una en el chip cubren cada región diana del genoma, con una sonda posicionada cada 10 bases en la cadena directa del genoma.

Las regiones genómicas altamente repetitivas se excluyeron en el diseño de los microchips de captura, para reducir la probabilidad de la unión no específica entre los microchips y las moléculas de ácido nucleico genómico. La estrategia para identificar y excluir regiones genómicas altamente repetitivas fue similar a la del programa WindowMasker (Morgulis et al.). La frecuencia promedio de los 15-meros de cada sonda se calculó mediante la comparación de las frecuencias de todos los 15-meros presentes en la sonda frente un histograma de frecuencias precalculado de todas las posibles sondas 15-meros en el genoma humano. La probabilidad de que la sonda represente una región repetitiva del genoma aumenta a medida que aumenta la frecuencia promedio de 15 unidades. Sólo las sondas que tienen una frecuencia promedio del 15-mero por debajo de 100 se incluyeron en los microchips de captura.

Para probar la reproducibilidad del sistema de captura, se utilizó primero el diseño exónico para capturar el DNA genómico fragmentado de una línea celular humana (linfoma de Burkitt, NA04671 (Coriell)) con el método que se muestra esquemáticamente en la Figura 2. Brevemente, el DNA genómico (20 µg) se sometió a la amplificación de todo el genoma (WGA; utilizando el servicio de Qiagen (Hilden, Alemania)). 20 µg del producto de amplificación de todo el genoma (WGA) se trataron con el fragmento Klenow de la DNA polimerasa I (NEB, Beverly MA) para generar extremos romos. Los fragmentos con extremos romos se sometieron a ultrasonidos para generar fragmentos de aproximadamente 500 pares de bases y luego se fosforilaron en 5' con la quinasa de polinucleótido (NEB). Se hibridaron y se ligaron a los extremos de los fragmentos fosforilados en 5' los enlazantes de oligonucleótidos 5'-Pi-GAGGATCCAGAATTCTCGAGTT-3' (Id. de Sec. N°: 2) y 5'-CTCGAGA-ATTCTGGATCCTC-3' (Id. de Sec. N°: 1):

Los fragmentos terminados en el enlazante se desnaturalizaron para producir productos de cadena sencilla que fueron expuestos a los microchips de captura en condiciones de hibridación y en presencia de tampón de hibridación 1X (Roche NimbleGen, Inc.) durante aproximadamente 65 horas a 42°C con mezclado activo usando una estación de hibridación MAUI (Roche NimbleGen, Inc.). Las moléculas monocatenarias que no se hibridaron se lavaron de la superficie de los microchips en condiciones de lavado astringentes, 3 x 5 minutos con tampón de lavado astringente (Roche NimbleGen, Inc.) y enjuague con los tampones de lavado I, II, y III (Roche NimbleGen, Inc.). Los fragmentos capturados en los microchips se eluyeron inmediatamente con 2 x 250 mL de agua a 95°C, se secaron y se resuspendieron para la amplificación por LM-PCR usando un cebador complementario de los oligonucleótidos enlazantes ligados previamente.

Para cuantificar el enriquecimiento de las regiones exónicas, se seleccionaron ocho regiones aleatorias para una PCR cuantitativa (qPCR). Estas regiones se amplificaron utilizando los siguientes cebadores:

Región 1 F: 5'-CTACCACGGCCCTTTCATAAAG-3' (Id. de Sec. N°: 3)  
R: 5'-AGGGAGCATTCCAGGAGAGAA-3' (Id. de Sec. N°: 4)

Región 2 F: 5'-GGCCAGGGCTGTGTACAGTT-3' (Id. de Sec. N°: 5)  
R: 5'-CCGTATAGAAGAGAAGACTCAATGGA-3' (Id. de Sec. N°: 6)

Región 3 F: 5'-TGCCCCACGGTAACAGATG-3' (Id. de Sec. N°: 7)  
R: 5'-CCACGCTGGTGATGAAGATG-3' (Id. de Sec. N°: 8)

Región 4 F: 5'-TGCAGGGCCTGGGTTCT-3' (Id. de Sec. N°: 9)  
R: 5'-GCGGAGGGAGAGCTCCTT-3' (Id. de Sec. N°: 10)

Región 5 F: 5'-GTCTCTTTCTCTCTTGTCCAGTTTT-3' (Id. de Sec. N°: 11)  
R: 5'-CACTGTCTTCTCCCGACATG-3' (Id. de Sec. N°: 12)

Región 6 F: 5'-AGCCAGAAGATGGAGGAAGCT-3' (Id. de Sec. N°: 13)  
R: 5'-TTAAAGCGCTTGCTTGGGA-3' (Id. de Sec. N°: 14)

Región 7 F: 5'-TCTTTTGAAGAAGGTATAGGTGTGGAA-3' (Id. de Sec. N°: 15)  
R: 5'-CAGGCCAGGCCACACT-3' (Id. de Sec. N°: 16)

Región 8 F: 5'-CGAGGCCTGCACAGTATGC-3' (Id. de Sec. N°: 17)  
R: 5'-GCGGGCTCAGCTTCTTAGTG-3' (Id. de Sec. N°: 18)

Después de una sola ronda de captura en el microchips, las muestras amplificadas enriquecidas y el DNA genómico control, que se fragmentó, se ligó al enlazante y se amplificó mediante LM-PCR, pero no se hibridó en un chip de captura, se compararon usando un sistema de PCR en tiempo real ABI 7300 (Applied Biosystems, Foster City, CA) midiendo la fluorescencia verde de SYBR de acuerdo con los protocolos del fabricante. Se logró un promedio de enriquecimiento de 378 veces en tres productos exónicos de captura replicados. El nivel de enriquecimiento máximo teórico era de 600 veces (3.000 Mb en el genoma y 5 Mb de secuencia total).

Las muestras eluidas de los microchips de captura se ligaron a enlazantes compatibles con la secuenciación en un 454, se amplificaron usando una PCR en emulsión sobre cuentas y se secuenciaron utilizando el instrumento de secuenciación FLX 454 (454, Branford CT). Debido a que cada fragmento secuenciado también contenía el enlazante de LM-PCR de 20 pb utilizado inmediatamente después de la elución de los microchips, la mayoría de lecturas de secuenciación del 454 contenían secuencia enlazante. La secuenciación del DNA de los tres replicados en el instrumento 454 FLX generaron 63 Mb, 115 Mb, y 93 Mb de secuencia total. Tras la eliminación *in silico* de la secuencia enlazante, cada lectura de secuenciación se comparó con la versión apropiada completa del genoma humano utilizando un análisis BLAST (Altschul, et al, 1990, J. Mol Biol 215: 403-410) con una puntuación de corte de  $e = 10^{-48}$ , afinado para maximizar el número de hallazgos únicos. Las lecturas que no mapaban únicamente en el genoma (entre el 10 y el 20%) fueron descartadas. El resto se consideraron secuencias capturadas. Las secuencias capturadas que, de acuerdo con la comparación de BLAST original, se mapan de forma única en regiones dentro de la región diana se consideraron éxitos de secuenciación. Estas se utilizaron para calcular el % de lecturas que se encontraban en las regiones diana, y la cobertura de secuenciación de la región diana completa. Los datos se visualizaron usando el programa SignalMap (Roche NimbleGen, Inc.).

El análisis BLAST mostró que el 91%, 89% y 91% de lecturas, respectivamente, se mapaban de forma única en el genoma; el 75%, 65%, y 77% eran de las regiones seleccionadas y el 96%, 93% y 95% de las secuencias diana contenían al menos una lectura de secuencia (Tabla 1, tres filas superiores) que representan un enriquecimiento promedio de alrededor de 400 veces. La cobertura mediana por base en cada muestra fue de 5, 7 y 7 veces la cobertura, respectivamente.

Tabla 1:

Muestra de DNA	Enriquecimiento en la qPCR	Rendimiento de FLX (Mb)	Porcentaje de lecturas mapadas de forma única en el genoma	Porcentaje de las lecturas totales mapadas en dianas de selección	Cobertura mediana de las regiones diana
NA04671	318	63,1	91%	75%	5
NA04671	399	115	89%	65 %	7
NA04671	418	93,0	91%	76%	7
HapMap CEPH	217	77,6	88%	74%	7
HapMap JPT	153	96,7	84%	66%	8
HapMap CHB	240	52,8	83%	59%	4
HapMap YRL	363	81,3	53%	38%	4

Ejemplo 3 - Variación de la secuencia capturada mediante enriquecimiento genómico y resecuenciación

Para determinar la capacidad de discernir la variación en el genoma humano, se capturaron muestras de DNA genómico a partir de cuatro tipos de células de la colección HapMap humana (CEPH/ NA 11839, CHB/ NA 18573, JPT/ NA18942, YRI/ NA18861, Coriell) sobre los chips exónicos de los ejemplos anteriores, se eluyeron y secuenciaron, como se describe aquí, con la excepción de que los DNA genómicos no eran todo el genoma amplificado antes de la captura. Los resultados de captura (que se muestran en la Tabla 1, filas 4-7) fueron similares a los anteriores, excepto que la cobertura de secuencia fue consistentemente más uniforme que en el caso anterior, lo que sugiere un sesgo introducido durante la AGC.

La secuencia de las cuatro muestras de HapMap se ensambló y se identificaron las mutaciones y se compararon con los datos de los SNP de HapMap para cada muestra (Tablas 1 y 2). El número total de posiciones en las regiones diana que se genotipó en el proyecto HapMap fue de 8103 (CEU), 8134 (CHB), 8134 (JPT), 8071 (YRI) para cada uno de los cuatro genomas. De éstos, la mayoría (~6000) de los sitios eran homocigotos para el alelo del genoma de referencia. El número de alelos variantes conocidos (homocigotos o heterocigotos) aparece en la segunda fila de la Tabla 2. Estas posiciones se analizaron para conocer su cobertura y determinar si el alelo(s) se encontraban en el DNA capturado.

Tabla 2:

Pop/ Indiv	CEPH/NA11839	CHB/ NA18573	JPT/ NA18942	CEPH/NA11839
Nº alelos variante conocidos	2235	2257	2206	2334
Rigurosidad de al menos una lectura por variante conocida de alelo HapMap				
Posiciones con $\geq 1$ lectura	2176 (97,3%)	2104 (93,2 %)	2168 (98,2%)	2.133 (91,3%)
Alelos variantes encontrados en $\geq 1$ lectura	2071 (92,6%)	1922 (85,1%)	2080 (94,2%)	1848 (79,1%)
Porcentaje de falsos negativos	7,4%	14,9%	5,8%	20,9%
Rigurosidad de al menos una lectura por variante conocida de alelo HapMap				

Posiciones con $\geq$ 1 lectura	2176 (97,3%)	2.104 (93,2%)	2.168 (98,2%)	2133 (91,3%)
Alelos variantes encontrados en $\geq$ 2 lecturas	1907 (85,3%)	1569 (69,5%)	1939 (87,8%)	1469 (62,9%)
Porcentaje de falsos negativos	14,7%	30,5%	12,2%	37,1%

Entre el 94% y 79% de las posiciones variantes conocidas entre las muestras de HapMap se identificaron en al menos una secuencia de lectura, lo que se esperaba, en base a la cobertura de secuencia general. No había ningún sesgo aparente en contra de alelos no presentes en el chip de captura cuando se comparó la cobertura de dianas que contenían 0, 1 o  $>1$  variantes conocidas (7,95, 8,48 y 8,82 veces la cobertura, respectivamente).

Hay un gran interés en el análisis de grandes regiones genómicas contiguas. Una serie de microchips de captura dirigidos a segmentos largos individuales de 200 kb – 5 Mb que rodean el gen BRCA 1 humano se probaron con el DNA NA04671. Para la serie de chips utilizados para capturar el locus del gen BRCA, se seleccionaron cinco regiones genómicas de tamaño creciente (200 kb, 500 kb, 1 Mb, 2 Mb y 5 Mb) que rodean al locus del gen BRCA1 a partir de la secuencia del genoma humano (bloque HG18). Las características de los chips de captura de locus se muestran en la Tabla 3. La densidad media de sondas contiguas es la distancia media entre el inicio de una sonda y el inicio de la siguiente sonda.

Tabla 3:

Tamaño de la región de BRCA1	Densidad promedio de las sondas contiguas de selección (pares de bases)	Coordenadas del cromosoma 17 (HG 18)
200 kb	1 pb	38.390.417 – 38.590.417
500 kb	1 pb	38.240.417 – 38.740.417
1 Mb	2 pb	37.990.417 – 38.990.417
2 Mb	3 pb	37.490.417 – 39.490.417
5 Mb	7 pb	35.990.417 – 40.990.417

La tabla 4 muestra que todas las dianas de captura funcionaron bien, con hasta 140 Mb de secuencia bruta generada en un solo ensayo del instrumento de secuenciación, generando una cobertura de unas 18 veces, a partir de una región de captura de 5 Mb. La Fig. 4b proporciona detalles del mapa de lectura de secuencia para la captura y secuenciación específica de locus. La línea 1 representa la posición cromosómica de 2.000 bases en el cromosoma humano 17, la línea 2 muestra la ubicación de las sondas, espaciadas cada 10 pares de bases y escalonados a lo largo del eje Y, el gráfico en 3 muestra la cobertura en la secuencia por bases, que oscila entre 0 y 100 por ciento, y el artículo 4 representa el mapa de lectura de las puntuaciones más altas de BLAST en las lecturas de secuenciación del 454. La Fig. 5 muestra la cobertura de secuencia acumulativa por base (Fig. 5a) y un histograma de la cobertura de secuencia (Fig. 4c) para la región de 2 Mb de BRCA1. El porcentaje de lecturas que se mapan en la secuencia diana aumenta con el tamaño de la región diana.

Tabla 4:

Tamaño de contigüidad (kb)	Selección Promedio sonda Vidrieros Densidad	Rendimiento del FLX (Mb)	Porcentaje de lecturas que mapan de forma única en el genoma	Porcentaje de lecturas totales que mapan en las dianas de selección	Mediana de la cobertura de porción única de la región
200	1 pb	102	55%	14%	79
500	1 pb	85,0	61%	36%	93
1000	2 pb	96,7	56%	35%	38
2000	3 pb	112,6	81%	60%	37
5000	7 pb	140	81%	64%	18

Estos datos ilustran el poder de los métodos de selección directa basada en microchips para el enriquecimiento de secuencias específicas. El inventor utilizó una plataforma de chips de alta densidad programable con 385.000 sondas que eran capaces de capturar fácilmente al menos 5 Mb de la secuencia total. Además de la especificidad del ensayo, los altos rendimientos de los pasos posteriores de secuenciación de DNA fueron superiores de forma consistente al funcionamiento promedio de rutina utilizando fuentes de DNA no capturadas. Esto se atribuye al proceso de enriquecimiento por captura que proporciona una purificación útil de secuencias únicas lejos de las repeticiones y otras impurezas que pueden confundir, por ejemplo, en la primera etapa de PCR en emulsión del proceso de secuenciación en el 454.

Ejemplo 4 - Captura y resecuenciación en fase de solución

La muestra de los Ejemplos 2 y 3 se analizó utilizando sondas de captura sintetizadas tras liberarse de un soporte sólido, de forma que el enriquecimiento fue ejecutado de forma ventajosa en fase de solución. Los diseños de microchips estándar (por ejemplo, el chip de secuencia contigua BRCA1 200K y los chips de captura exónica humanos de los ejemplos anteriores) fueron modificados mediante la adición de secuencias de cebadores 15meros terminales que contienen un sitio de reconocimiento MlyI, lo que facilita la eliminación enzimática del cebador, dejando la secuencia de oligonucleótidos de captura intacta.

Los chips se sintetizaron mediante la adición de un reactivo de fosforilación química (Glen Research) tras el enlazante T5 inicial y antes de la secuencia del cebador 3'. Se llevaron a cabo tres acoplamientos individuales para maximizar la posterior escisión de las sondas de captura de los chips.

Las sondas de captura inmovilizadas en el chip se trataron con hidróxido de amonio al 30% (NH<sub>4</sub>OH, Aldrich). Después de la síntesis, los chips se colocaron en una cámara húmeda y se añadieron aproximadamente 700 µl de NH<sub>4</sub>OH en la zona de síntesis a temperatura ambiente durante 20 minutos para escindir las sondas del chip. El NH<sub>4</sub>OH se mantuvo en gran medida dentro del área de la zona de síntesis debido a las diferencias de hidrofobicidad entre la zona de reacción y el vidrio circundante. La solución se retiró con una pipeta y se retuvo. Se añadieron 700 µl adicionales de NH<sub>4</sub>OH fresco a la superficie. El proceso se repitió un total de 3 veces (60 min. y 2,1 ml en total). Las sondas de captura de oligonucleótidos escindidos se secaron entonces mediante centrifugación al vacío en condiciones estándar conocidas en la técnica.

Las sondas de captura escindidas se amplificaron bajo condiciones estándar. Las sondas secas se volvieron a resuspender en 30 µl de agua desionizada (diH<sub>2</sub>O) y se alicuotaron en 30 reacciones de PCR individuales como sigue:

Reactivos de reacción:

Tampón 10X	2,5 µl
dNTP 25 mM	0,125 µl
Cebador 1A 20 µM	1,25 µl
Cebador 1b 20 µM biotinilado	1,25 µl
HotStart Taq	0,25 µl
MgCl	1 µl
Muestra	1 µl
H <sub>2</sub> O	17,625 µl
Volumen total	25 µl

Cebador 1a:  
5'-TGCCGGAGTCAGCGT-3 '(Id. de Sec. N°: 19)

Cebador 1b:  
5'-biotina-AGTCAGAGTCGCCAC-3 '(Id. de Sec. N°: 20)

Las reacciones se amplificaron de acuerdo con el siguiente programa:

<u>Número de ciclo</u>	<u>Desnaturalización</u>	<u>Hibridación</u>	<u>Polimerización</u>
1	15 min. a 95 ° C		
2-31	20 s a 95°C	45 s a 48°C	20 s a 72°C

Los productos de PCR se purificaron de los componentes de reacción utilizando el equipo de eliminación de nucleótidos QiaQuick (Qiagen), se secaron y se resuspendieron en 20 µl de diH<sub>2</sub>O. El rendimiento típico después de la purificación fue de aproximadamente 400-700 ng/ reacción por Nanodrop. Los amplicones pueden comprobarse en un gel de agarosa al 3%. Dependiendo de los requisitos de cantidad de sondas de captura, se realizaron rondas adicionales de PCR como las anteriores produciendo aproximadamente 200 ng de muestra por reacción. Los amplicones se purificaron y se caracterizaron como se ha indicado anteriormente.

La ronda final de amplificación de las sondas de captura se realizó mediante una PCR asimétrica. El protocolo fue como el anterior, excepto que mientras que concentración de cebador biotinilado sigue siendo la misma, la concentración de cebador no biotinilado se redujo a 0,001x de la concentración original. El protocolo se extendió a 35 ciclos para permitir una amplificación no exponencial. Los amplicones se secaron, se resuspendieron en 20 µl de diH<sub>2</sub>O y se caracterizaron.

La muestra de DNA genómico se preparó según el protocolo estándar; 20 µg de la muestra de AGC con enlazantes se secó con 100 µg de DNA Cot-1 y se resuspendió en 7,5 µl de tampón de hibridación y 3 µl de formamida. Una alícuota de 2 µg de las sondas de captura se secó y se resuspendió en 4,5 µl de diH<sub>2</sub>O. La solución de muestra se mezcló con la solución de sondas de captura y se incubó a 95°C durante 10 minutos. A continuación, la mezcla se

## ES 2 522 912 T3

transfirió a un tubo de PCR y se colocó en un termociclador durante 3 días a 42°C para su hibridación y formar dúplex.

5 Después de la hibridación, los dúplex se unieron a cuentas paramagnéticas (Dyna). 25 µl de cuentas se lavaron tres veces en tampón BW 2x (Tris HCl 10 mM, EDTA 1 mM, NaCl 2 M), y las cuentas se resuspendieron en la mezcla de hibridación. La unión se produjo a lo largo de 45 minutos a 42°C con mezclado suave ocasional.

10 Las cuentas unidas se aislaron utilizando un imán y se lavaron brevemente con 40 µl de tampón de lavado I, se incubaron durante 2 x 5 minutos en tampón de lavado riguroso a 47°C, se lavó con tampón de lavado I durante aproximadamente 2 minutos a temperatura ambiente, con tampón de lavado II durante aproximadamente 1 minuto, y con tampón de lavado III durante aproximadamente 30 segundos.

15 Para eluir los fragmentos capturados, la solución que contenía cuentas en tampón de lavado III se transfirió a un tubo Eppendorf de 1,5 ml. Las cuentas se aislaron con un imán. El tampón de lavado se eliminó y se añadieron ~100 µl de diH<sub>2</sub>O a 95°C. La solución se incubó a 95°C durante 5 minutos, después de lo cual las cuentas se unieron a un imán y se lavaron suavemente con diH<sub>2</sub>O a 95°C. A continuación, el líquido de lavado se eliminó y se conservó, y se reemplazó con diH<sub>2</sub>O fresca a 95°C. La incubación y lavado se repitió durante un total de 3 veces (15 minutos, aproximadamente 300 µl de eluido). Después del lavado final, el tubo Eppendorf que contenía el eluido se colocó en un soporte magnético durante aproximadamente 5 minutos para aislar cualquier cuenta aspirada durante la elución. 20 La solución se secó a temperatura elevada en un tubo Eppendorf. Los fragmentos capturados eluidos se volvieron a resuspender en 263 µl de diH<sub>2</sub>O antes de la LM-PCR estándar.

25 Después de LM-PCR, los fragmentos capturados fueron sometidos a secuenciación estándar ultra-profunda utilizando la plataforma 454 FLX, como anteriormente. Alternativamente, se puede evitar la LM-PCR mediante la ligación de secuencias adaptadoras de secuenciación en 454 en la muestra de pre-enriquecimiento. En ese caso, las secuencias enriquecidas eluidas se pueden añadir directamente en la PCR de emulsión para la secuenciación ultra-profunda.

30 Los datos indicaron que el 83,8% de las lecturas mapaban en las regiones diana, lo que es comparable e indistinguible de los resultados obtenidos mediante protocolos de captura basados en chips.

Ejemplo 5 - Captura en la fase solución utilizando amplificación in situ de sondas de captura

35 Un diseño estándar de microchips se modificó mediante la adición de una secuencia de cebador 15-mero terminal que contenía un sitio de reconocimiento MlyI (GAGTC(5/5)). La incorporación de una diana MlyI en la secuencia del cebador facilita la eliminación enzimática del cebador dejando intactas las secuencias de oligonucleótidos de captura. Los chips se sintetizaron mediante métodos de síntesis de chips sin máscara estándar conocidos por los expertos en la materia.

40 Las sondas de captura se amplificaron utilizando la reacción en cadena de la polimerasa in situ (PCR) sobre un chip en un termociclador usando una cámara de hibridación de sellado (Grace Bio-Labs, Inc., Bend, OR) y Adaptador de plancha deslizando (Bio-Rad Laboratories, Hercules, CA). Se añadieron los constituyentes de la reacción de PCR (25 µl de tampón 10X de polimerasa, 1,25 µl de cada dNTP 25 mM, 12,5 µl de cada cebador 20 µM 1a y 1b, 2,5 µl de polimerasa Taq Hotstart, 10 µl de MgCl<sub>2</sub> 25 mM y 176,5 µl de diH<sub>2</sub>O, el volumen total de reacción es de 250 µL) a las 45 cámaras de hibridación de microchips y la PCR se realizó usando las condiciones; 100 °C durante 30 segundos, 97 °C durante 15 min., 30 ciclos de 100 °C durante 30 segundos, 47,5 °C durante 45 segundos, 78 °C durante 30 segundos, seguido por el enfriamiento de las reacciones hasta 1 °C durante 30 segundos y a 3,5 °C para mantener. Las secuencias de cebadores fueron para el cebador 1a 5'-TGCCGGAGTCAGCGT-3' (Id. de Sec. N°: 19) y el cebador 1b 5'-Biotina-AGTCAGAGTCGCCAC-3' (Id. de Sec. N°: 20), lo que refleja sitios de unión a cebador que 50 fueron incorporados a las secuencias de la sonda.

Los amplicones de sonda de captura de la reacción en cadena de la polimerasa se purificaron a partir de los componentes de reacción utilizando el equipo de eliminación de nucleótidos QIAquick® (Qiagen, Inc., Valencia, CA), se secaron y se resuspendieron en 20 µl de diH<sub>2</sub>O. El rendimiento de amplificación fue de aproximadamente 5 µg en 55 total, medido mediante espectrofotometría NanoDrop® (Thermo Fisher Scientific). Se pueden realizar rondas de amplificación adicionales, siguiendo el protocolo anterior, si se necesita una cantidad de amplicón adicional (por ejemplo, utilizando el protocolo anterior y 100 ng de muestra por reacción).

60 La ronda final de la amplificación de las sondas de captura se realizó mediante PCR asimétrica; 2,5 µl de tampón polimerasa 10X, 0,125 µl de dNTPs 25 mM, 0,0125 µl de cebador 1a 20µM, 1,25 µl de cebador 1b 20 µM, 0,25 µl de Taq Hotstart, 1 µl MgCl<sub>2</sub> 25 mM y 18,86 µl de diH<sub>2</sub>O (volumen de reacción total de 25 µl). Los amplicones se purificaron de los componentes de reacción utilizando las columnas Qiagen MinElute™ y se cuantificaron como se describe anteriormente.

65 Una muestra de DNA genómico se preparó según el protocolo estándar. Veinte µg de la muestra con enlazantes unidos se secaron con 100 µg de DNA Cot-1 y se resuspendieron en 7,5 µl de tampón de hibridación (Roche

NimbleGen, Madison, WI) y 3 µL de formamida. Una alícuota de 1 µg de las sondas de captura se secó y se resuspendió en 4,5 µ de diH<sub>2</sub>O. La solución de muestra se incubó a 95 °C durante 10 min. para desnaturalizar el DNA y se añadió a la solución de sonda de captura. La mezcla se transfirió a un tubo de PCR y se colocó en un ciclador térmico a 42 °C durante 3 días para permitir que se produzca la formación de dúplex.

5 Después de la hibridación, los dúplex estaban unidos a cuentas paramagnéticas recubiertas con estreptavidina (Dyna<sup>®</sup>, Invitrogen, Carlsbad, CA). Cien microlitros de cuentas se lavaron tres veces con tampón 2X BW (Tris HCl 10 mM, EDTA 1 mM, NaCl 2 M) y se resuspendieron en la mezcla de dúplex de hibridación. Se permitió que se produjera la unión entre las cuentas y el dúplex durante 45 min. a 42 °C con agitación suave ocasional. Las cuentas unidas se aislaron utilizando un imán y se lavaron brevemente en tampón de lavado I (0,2 X SSC, 0,2% (v / v) SDS, DTT 0,1 mM) a temperatura ambiente, seguido de dos lavados (cada lavado durante 5 min. a 47 °C) en 200 µl de tampón de lavado riguroso (MES 0,1 M pH 6,65, NaCl 0,1 M, 0,1% de Tween 20), un lavado adicional en el tampón de lavado I durante 2 min. a temperatura ambiente, una vez con tampón de lavado II (SSC 0,2x, DTT 0,1 mM) durante 1 min. a temperatura ambiente y finalmente durante 30 seg. en tampón de lavado III (SSC 0,05x, DTT 0,1 mM) a temperatura ambiente.

Los fragmentos capturados se eluyeron de las cuentas. La solución de cuentas lavadas en tampón de lavado III se transfirió a un tubo Eppendorf de 1,5 ml, las cuentas se aislaron con un imán, el tampón de lavado eliminado y reemplazado con 100 µl de diH<sub>2</sub>O a 95 °C y las cuentas se liberaron del imán. Las cuentas suspendidas se incubaron a 95 °C durante 5 min. después de lo cual se capturaron las cuentas y se lavaron suavemente con diH<sub>2</sub>O a 95 °C para eluir los fragmentos capturados. El eluato se eliminó y las cuentas se lavaron de nuevo, para un total de tres lavados con agua; total de 10 min. con volumen final del conjunto de eluatos de aproximadamente 300 µl. Después del lavado final, las cuentas magnéticas residuales se eliminaron a partir del conjunto de eluatos mediante una captura magnética adicional y se transfirió el eluato a un tubo nuevo. La solución se secó y los fragmentos eluidos capturados, se resuspendieron en 263 µl de diH<sub>2</sub>O en preparación para la posterior LM-PCR. La ligación se realizó mediante protocolos establecidos conocidos por los expertos en la materia, usando un ligador de secuencia 5'-CTCGAGAATTCTGGATCC-3 '(Id. de Sec. N°: 21).

Después de LM-PCR, los fragmentos capturados se sometieron a secuenciación ultraprofunda utilizando la plataforma 454 FLX (454 Life Sciences, Branford, CT). Alternativamente, la LM-PCR se puede evitar mediante la ligación de las secuencias adaptadoras 454 de la secuenciación con la muestra de pre-enriquecimiento. En este último caso, las secuencias enriquecidas eluidas se pueden añadir directamente en la emulsión de PCR de la plataforma 454 FLX.

35 La Figura 7 ilustra un experimento de resecuenciación a partir de fragmentos capturados en solución usando los métodos descritos anteriormente. Los controles de qPCR que utilizan secuencias de cebador control de PCR indican una media de enriquecimiento de 2600 veces a lo largo de los cuatro loci control. Secuencias de cebador control de qPCR:

40 qPCR gSel-0210F GACCCTTTACCTTGGCATTCTC (Id. de Sec. N°: 22)  
qPCR gSel-0210R GCTGGTACCCATTGGCAACT (Id. de Sec. N°: 23)

qPCR gSel-0271F GGAGTGAGTGGTTTTTCTTCATTTTT (Id. de Sec. N°: 24)  
45 qPCR gSel-0271R GCGCCACAAAGAGACATTCA (Id. de Sec. N°: 25)

qPCR gSel-0266F AAGGCCATACTTGGGTGAACTG (Id. de Sec. N°: 26)  
qPCR gSel-0266R GCTCTGATTGGTGGCTTCGT (Id. de Sec. N°: 27)

qPCR gSel-0283F TGCTTGCAGGTGTCTCTCAGA (Id. de Sec. N°: 28)  
50 qPCR gSel-0283R CAGTGAGATATTTGGTACCATGGTGTA (Id. de Sec. N°: 29)

De hecho, la conformación en la que el porcentaje de lo que se espera tras la secuenciación con lo que se realiza tras la resecuenciación es de aproximadamente el 100% para casi todas las regiones resecuenciadas.

55 Listado de referencias

- Altschul, S.F. et al. (1990) J. Mol. Biol. 215, 403-410  
Hyman, E. D. (1988), Anal. Biochem. 174, 423-436  
Lovett et al. (1991) PNAS USA, 88, 9628-9632  
60 Morgulis, A. et al. (2006) Bioinformatics, 15, 134-41  
Nuwaysir, E.F., et al., (2002) Genome Res. 12, 1749-1755  
Rhonaghi et al. (1998), Science 281, 363-365  
Soares, et al. (1994) PNAS, 91, 9228-9232  
Singh-Gasson, S., et al. (1999) Nat. Biotechnol. 17, 974-978  
65 Wetmur (1991) Critical Reviews in Biochemistry and Molecular Biology, 26(34):227-59  
Wetmur et al. (1966) J. Mol. Biol., 31, 349-70

"Direct selection of cDNAs with large genomic DNA clones," en Molecular Cloning: A Laboratory Manual (Ed. Sambrook, J. & Russell, D.W.) Capítulo 11 Protocolo 4, págs. 11.98-11.106 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, USA, 2001)

PE 0 552 290

5 US 2005/0282209

US 5.143.854

US 6.013.440

US 6.375.903

WO 2004/ 070007

10

LISTADO DE SECUENCIAS

<110> Roche Diagnostics GmbH/Hoffman-La Roche AG

15 <120> Métodos y sistemas para el enriquecimiento de secuencias basados en solución y análisis de regiones genómicas

<130> R65003PC

20 <150> EP07020660.2

<151> 2007-10-23

<150> US 12/194,574

25

<151> 2008-8-20

<160> 29

30 <170> PatentIn versión 3.5

<210> 1

<211> 20

35

<212> DNA

<213> Secuencia artificial

40

<220>

<223> Oligonucleótido sintético 1

<400> 1

45 ctcgagaatt ctggatcctc 20

<210> 2

<211> 22

50

<212> DNA

<213> Secuencia artificial

55

<220>

<223> Oligonucleótido sintético 2

<400> 2

60 gaggatccag aattctcgag tt 22

<210> 3

<211> 22

65

<212> DNA

<213> Secuencia artificial  
<220>  
5 <223> Región 1 Cebador directo  
<400> 3  
ctaccacggc ccttcataa ag 22  
10 <210> 4  
<211> 21  
15 <212> DNA  
<213> Secuencia artificial  
<220>  
20 <223> Región 1 Cebador reverso .  
<400> 4  
agggagcatt ccaggagaga a 21  
25 <210> 5  
<211> 20  
30 <212> DNA  
<213> Secuencia artificial  
<220>  
35 <223> Región 2 Cebador directo  
<400> 5  
ggccagggct gtgtacagtt 20  
40 <210> 6  
<211> 26  
45 <212> DNA  
<213> Secuencia artificial  
<220>  
50 <223> Región 2 Cebador reverso  
<400> 6  
ccgtatagaa gagaagactc aatgga 26  
55 <210> 7  
<211> 19  
60 <212> DNA  
<213> Secuencia artificial  
<220>  
65 <223> Región 3 Cebador directo

<400> 7  
tgccccacgg taacagatg 19

5 <210> 8  
<211> 20  
<212> DNA

10 <213> Secuencia artificial  
<220>

15 <223> Región 3 Cebador reverso  
<400> 8  
ccacgctggt gatgaagatg 20

20 <210> 9  
<211> 17  
<212> DNA

25 <213> Secuencia artificial  
<220>

30 <223> Región 4 Cebador directo  
<400> 9  
tgcagggcct gggttct 17

35 <210> 10  
<211> 18  
<212> DNA

40 <213> Secuencia artificial  
<220>

45 <223> Región 4 Cebador reverso  
<400> 10  
gcggagggag agctcctt 18

50 <210> 11  
<211> 27  
<212> DNA

55 <213> Secuencia artificial  
<220>

60 <223> Región 5 Cebador directo  
<400> 11  
gtctcttct ctctctgtc cagtttt 27

65 <210> 12

<211> 21  
<212> DNA  
5 <213> Secuencia artificial  
<220>  
<223> Región 5 Cebador reverso  
10 <400> 12  
cactgtcttc tcccgacat g 21  
<210> 13  
15 <211> 21  
<212> DNA  
20 <213> Secuencia artificial  
<220>  
<223> Reverse 6 Cebador directo  
25 <400> 13  
agccagaaga tggaggaagc t 21  
<210> 14  
30 <211> 19  
<212> DNA  
35 <213> Secuencia artificial  
<220>  
<223> Reverse 6 Cebador reverso  
40 <400> 14  
ttaaagcgct tggcttgga 19  
<210> 15  
45 <211> 26  
<212> DNA  
50 <213> Secuencia artificial  
<220>  
<223> Región 7 Cebador directo  
55 <400> 15  
tcttttgaga aggtataggt gtggaa 26  
<210> 16  
60 <211> 17  
<212> DNA  
65 <213> Secuencia artificial

<220>  
<223> Región 7 Cebador reverso  
5 <400> 16  
caggcccagg ccacact 17  
<210> 17  
10 <211> 19  
<212> DNA  
<213> Secuencia artificial  
15 <220>  
<223> Región 8 Cebador directo  
20 <400> 17  
cgaggcctgc acagtatgc 19  
<210> 18  
25 <211> 20  
<212> DNA  
<213> Secuencia artificial  
30 <220>  
<223> Región 8 Cebador reverso  
35 <400> 18  
gcgggctcag ctcttagtg 20  
<210> 19  
40 <211> 15  
<212> DNA  
<213> Secuencia artificial  
45 <220>  
<223> Cebador de PCR 1a  
50 <400> 19  
tgccggagtc agcgt 15  
<210> 20  
55 <211> 15  
<212> DNA  
<213> Secuencia artificial  
60 <220>  
<223> Cebador de PCR biotinilado 1b  
65 <400> 20  
agtcagagtc gccac 15

<210> 21  
 <211> 18  
 5 <212> DNA  
 <213> Secuencia artificial  
 10 <220>  
 <223> Enlazante de oligonucleótido sintético  
 <400> 21  
 15 ctcgagaatt ctggatcc 18  
 <210> 22  
 <211> 23  
 20 <212> DNA  
 <213> Secuencia artificial  
 25 <220>  
 <223> qPCR 0210 Cebador directo  
 <400> 22  
 30 gaccctcta ccttggcatt ctc 23  
 <210> 23  
 <211> 20  
 35 <212> DNA  
 <213> Secuencia artificial  
 40 <220>  
 <223> qPCR 0210 Cebador reverso  
 <400> 23  
 45 gctgtacc attgcaact 20  
 <210> 24  
 <211> 26  
 50 <212> DNA  
 <213> Secuencia artificial  
 55 <220>  
 <223> qPCR 0271 Cebador directo  
 <400> 24  
 60 ggagtgagtg gttttcttc atttt 26  
 <210> 25  
 <211> 20  
 65 <212> DNA

<213> Secuencia artificial  
 <220>  
 5 <223> qPCR 0271 Cebador reverso  
 <400> 25  
 gcgccacaaa gagacattca 20  
 10 <210> 26  
 <211> 22  
 15 <212> DNA  
 <213> Secuencia artificial  
 <220>  
 20 <223> qPCR 0266 Cebador directo  
 <400> 26  
 aaggccatac ttgggtgaac tg 22  
 25 <210> 27  
 <211> 20  
 30 <212> DNA  
 <213> Secuencia artificial  
 <220>  
 35 <223> qPCR 0266 Cebador reverso  
 <400> 27  
 gctctgattg gttgcttcgt 20  
 40 <210> 28  
 <211> 21  
 45 <212> DNA  
 <213> Secuencia artificial  
 <220>  
 50 <223> qPCR 0283 Cebador directo  
 <400> 28  
 tgcttcagg tgctctcag a 21  
 55 <210> 29  
 <211> 27  
 60 <212> DNA  
 <213> Secuencia artificial  
 <220>  
 65 <223> qPCR 0283 Cebador reverso

<400> 29  
cagtgagata ttggtacca tgggta 27

**REIVINDICACIONES**

1. Un método para aislar y reducir la complejidad de una pluralidad de secuencias de ácido nucleico que comprende:
- 5 a) proporcionar:
- i) un soporte sólido que comprende sondas de hibridación que pueden hibridarse con secuencias de ácido nucleico diana,
- 10 ii) una muestra de ácido nucleico que comprende las secuencias de ácidos nucleicos diana,
- b) amplificar dichas sondas de hibridación de forma que los productos de amplificación comprendan una porción de unión y en el que dichos productos de amplificación se mantengan en solución,
- 15 c) hibridar dicha muestra de ácido nucleico a dichos productos de amplificación en solución de forma que se permita que se produzca la hibridación entre dichos productos de amplificación y las secuencias de ácidos nucleicos diana,
- d) separar los complejos de hibridación de los ácidos nucleicos diana / producto de amplificación de los ácidos nucleicos hibridados no específicamente mediante dicha matriz de unión, y
- 20 e) eluir las secuencias de ácido nucleico diana hibridadas del complejo, aislandolo de este modo y reduciendo la complejidad de una pluralidad de secuencias de ácido nucleico.
2. El método según la reivindicación 1, que comprende además la secuenciación de las secuencias de ácidos nucleicos diana eluidas.
- 25 3. El método de acuerdo con cualquiera de las reivindicaciones 1 o 2, en el que dicho soporte sólido es un portaobjetos para microchip.
4. El método de acuerdo con una cualquiera de las reivindicaciones de 1 a 3, en el que dicha muestra de ácido nucleico es una muestra de DNA genómico fragmentado.
- 30 5. El método según la reivindicación 4, en el que dicha muestra de DNA genómico fragmentado comprende además moléculas adaptadoras en uno o ambos extremos de los fragmentos de la muestra de DNA genómico.
- 35 6. El método de acuerdo con cualquiera de las reivindicaciones de 1 a 5, en el que dichas sondas de hibridación comprenden además secuencias de unión de cebadores en uno o ambos extremos de dichas sondas.
7. El método según la reivindicación 6, en el que dichas secuencias de unión cebador, cuando están presentes en ambos extremos de las sondas, son iguales.
- 40 8. El método según la reivindicación 7, en el que dichas secuencias de unión cebador, cuando están presentes en ambos extremos de las sondas, son diferentes.
9. El método de acuerdo con cualquiera de las reivindicaciones de 1 a 8, en el que dicha amplificación comprende una reacción en cadena de polimerasa exponencial.
- 45 10. El método según la reivindicación 9, en el que dicha amplificación comprende, además, una reacción en cadena de la polimerasa asimétrica.
- 50 11. El método de acuerdo con cualquiera de las reivindicaciones de 1 a 10, en el que dicha porción de unión es una porción de unión a biotina.
12. El método según la reivindicación 11, en el que dicha separación comprende la unión de dicha porción de unión a biotina a un sustrato recubierto con estreptavidina.
- 55 13. El método según la reivindicación 12, en el que dicho sustrato recubierto con estreptavidina es una partícula paramagnética recubierta con estreptavidina.
- 60 14. El método de acuerdo con cualquiera de las reivindicaciones de 1 a 13, que comprende además el lavado de dichos complejos de hibridación del producto de amplificación / ácidos nucleicos diana separados antes de la elución.
- 65 15. El método de acuerdo con una cualquiera de las reivindicaciones de 1 a 14, en el que dicha población de moléculas de ácido nucleico o dicha pluralidad de secuencias de ácido nucleico contiene todo el genoma o al menos un cromosoma de un organismo, o de al menos una molécula de ácido nucleico con un tamaño de al menos aproximadamente 200 kb, al menos aproximadamente 500 kb, al menos aproximadamente 1 Mb, al menos

aproximadamente 2 Mb o al menos aproximadamente 5 Mb, especialmente un tamaño entre aproximadamente 100 kb y alrededor de 5 Mb, entre aproximadamente 200 kb y alrededor de 5 Mb, entre alrededor de 500 kb y aproximadamente 5 Mb, entre alrededor de 1 Mb y 2 Mb, o aproximadamente entre 2 Mb y unos 5 Mb.

5 16. El método de acuerdo con cualquiera de las reivindicaciones de 4 a 15, en el que dichas sondas se seleccionan de:

- una pluralidad de sondas que definen una pluralidad de exones, intrones o secuencias reguladoras de una pluralidad de loci genéticos,

10 - una pluralidad de sondas que definen la secuencia completa de al menos un locus genético único, y dicho locus tiene un tamaño de al menos 100 kb, preferiblemente al menos 1 Mb, o al menos uno de los tamaños que se especifican en la reivindicación 3,

15 - una pluralidad de sondas que definen puntos que se conoce contienen polimorfismos de un único nucleótido (SNP), o

- una pluralidad de sondas que definen un chip, en particular un chip de secuencias contiguas, diseñado para capturar la secuencia completa de al menos un cromosoma completo.

20 17. Un método para determinar la información de una secuencia de ácido nucleico de al menos una región de ácido nucleico, en particular de ácido nucleico genómico, y el método comprende los pasos de:

1. reducir la complejidad genética de una población de moléculas de ácido nucleico de acuerdo con un método de acuerdo con una de las reivindicaciones de 1 a 16, y

25 2. determinar la secuencia de ácido nucleico de las moléculas capturadas, en particular mediante la realización de una secuenciación mediante reacciones de síntesis.

30 18. Un método para detectar la variación en la región de codificación en relación a un genoma de referencia, que comprende el método las etapas:

1. reducir la complejidad genética de una población de moléculas de ácido nucleico de acuerdo con un método de acuerdo con una de las reivindicaciones de 1 a 16,

2. la determinación de la secuencia de ácido nucleico de las moléculas capturadas, y

35 3. comparar la secuencia determinada con las secuencias en una base de datos del genoma de referencia, en particular, con las secuencias en una base de datos de polimorfismos en el genoma de referencia para identificar las variantes del genoma de referencia.

Fig. 1

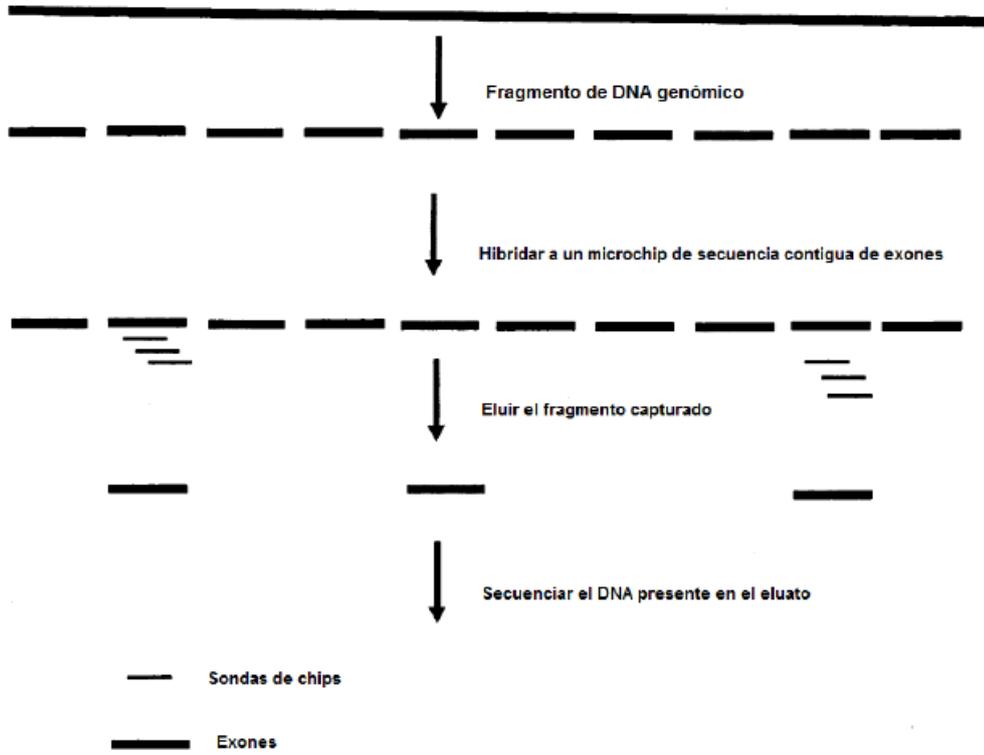


Fig. 2

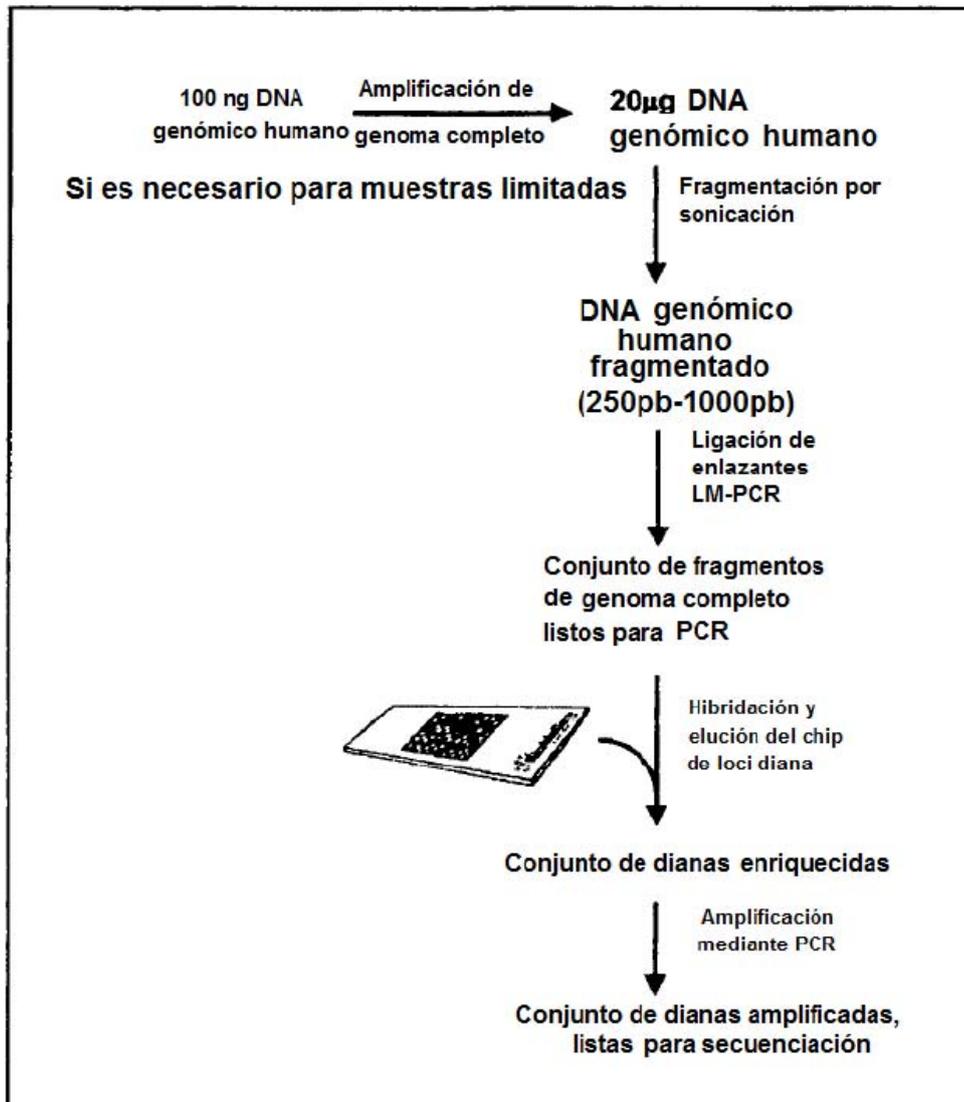
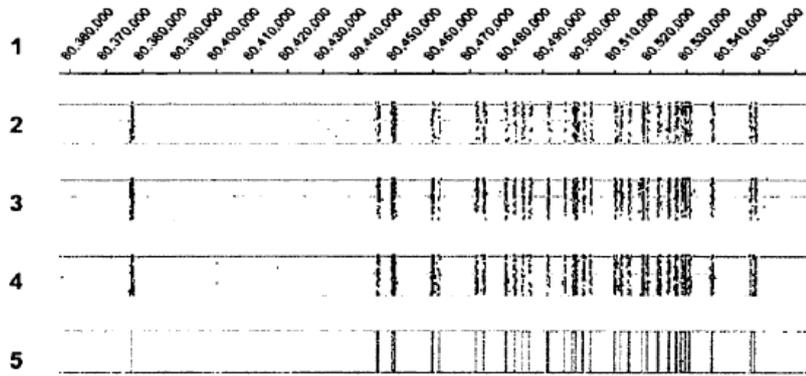


Fig. 3 (a-b)

a)



b)

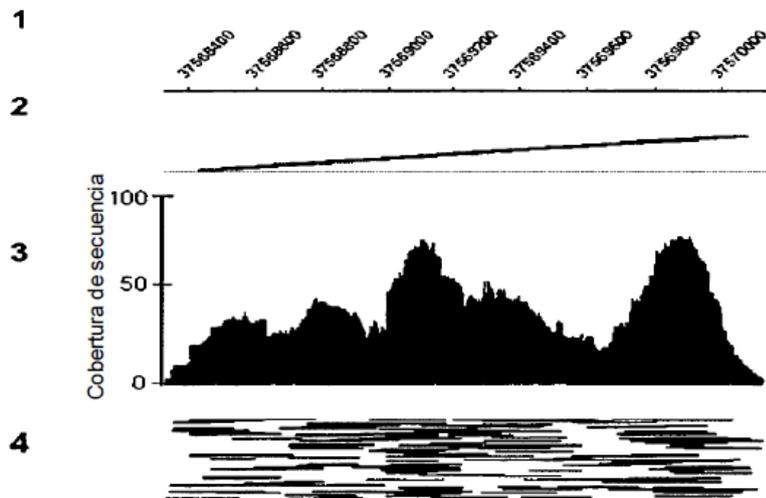
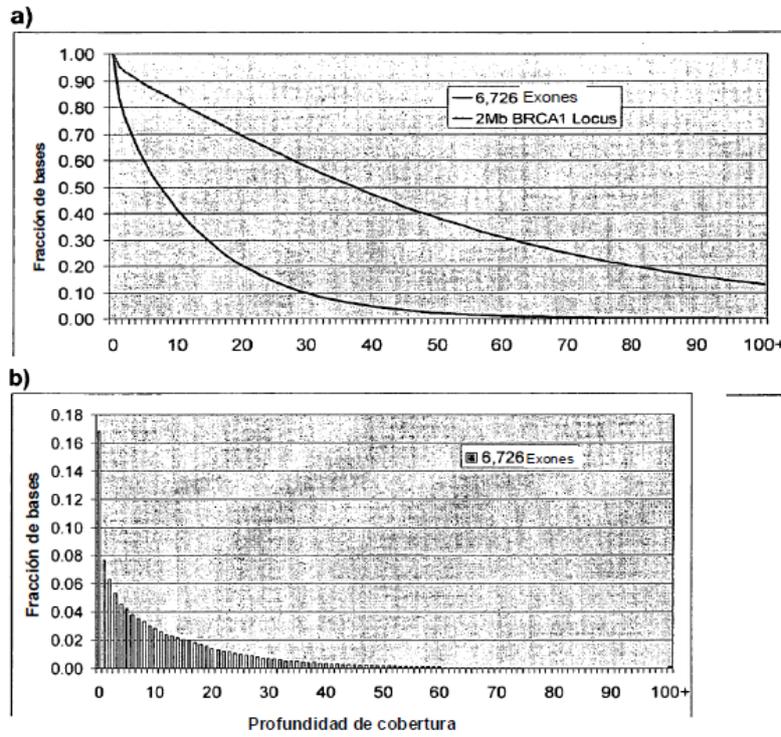


Fig. 4 (a-c)



c)

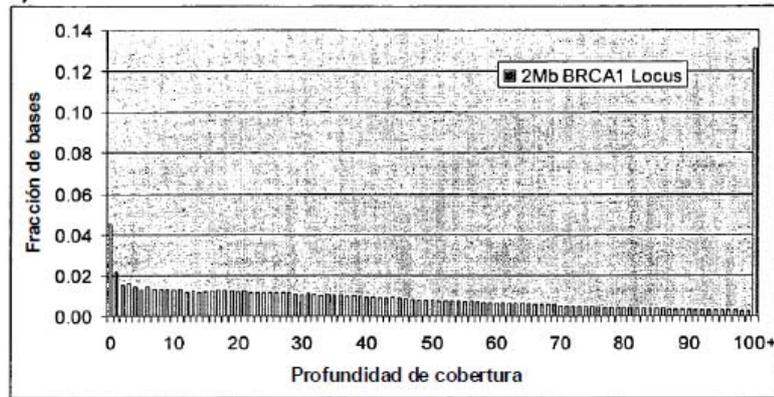


Fig. 5:

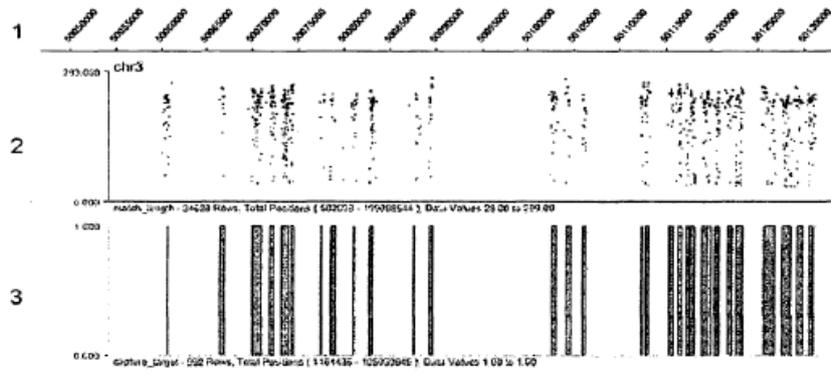


FIG. 6

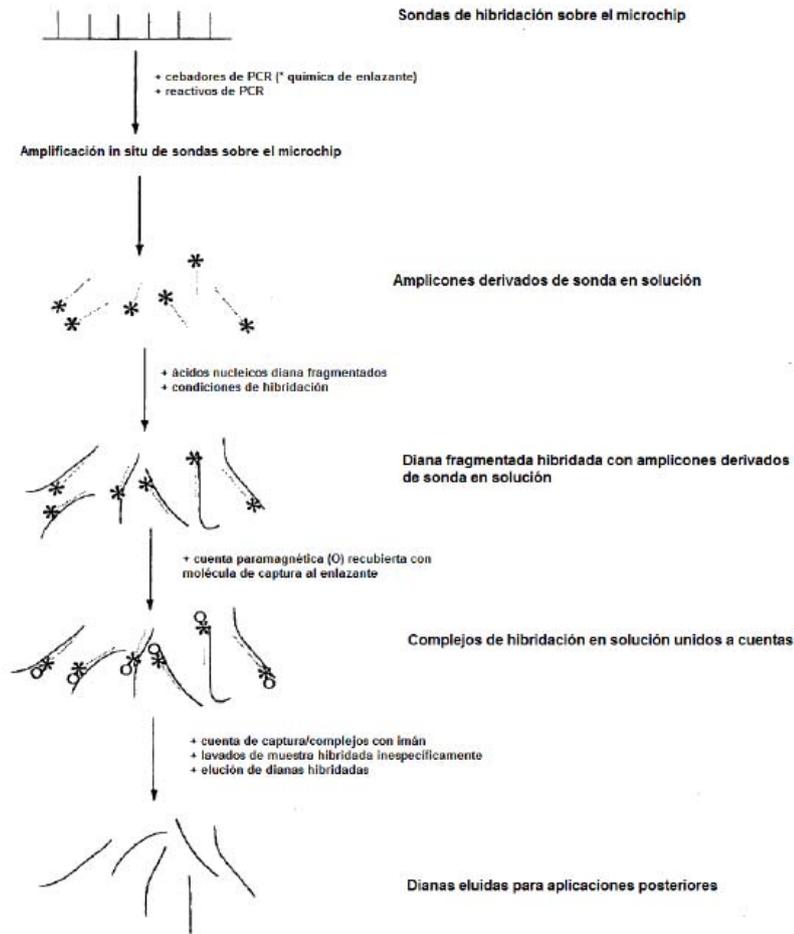


FIG. 7

