

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 667 346**

51 Int. Cl.:

**C12Q 1/68** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **08.07.2011 PCT/NL2011/050504**

87 Fecha y número de publicación internacional: **12.01.2012 WO12005595**

96 Fecha de presentación y número de la solicitud europea: **08.07.2011 E 11736190 (7)**

97 Fecha y número de publicación de la concesión europea: **04.04.2018 EP 2591125**

54 Título: **Estrategias de secuenciación de región genómica 3-D de interés**

30 Prioridad:

**09.07.2010 US 362778 P**

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

**10.05.2018**

73 Titular/es:

**CERGENTIS B.V. (100.0%)  
Yalelaan 62  
3584 CM Utrecht, NL**

72 Inventor/es:

**VAN MIN, MAX, JAN y  
DE LAAT, WOUTER, LEONARD**

74 Agente/Representante:

**ELZABURU, S.L.P**

**Observaciones :**

**Véase nota informativa (Remarks, Remarques o Bemerkungen) en el folleto original publicado por la Oficina Europea de Patentes**

**ES 2 667 346 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

Estrategias de secuenciación de región genómica 3-D de interés

**Campo de la invención.**

La presente invención se refiere al campo de la biología molecular y, más concretamente, a la tecnología del ADN. La invención se refiere con más detalle a la secuenciación del ADN. La invención se refiere a estrategias para determinar una secuencia de ADN (o parte de ella) de una región genómica de interés. En particular, la invención se refiere a la determinación de la secuencia de partes de un genoma que están en una configuración espacial entre sí. La invención se refiere además a los usos de los métodos de la invención en el desarrollo de diagnósticos y tratamientos médicos personalizados, en el cribado de tejidos para detectar la presencia de tumores malignos y otras enfermedades.

**Antecedentes.**

Se ha dedicado un esfuerzo considerable al desarrollo de estrategias de "enriquecimiento de la diana" para la secuenciación, en las que regiones genómicas de una muestra de ADN son capturadas selectivamente y/o amplificadas selectivamente, y posteriormente secuenciadas (revisado en Mamanova et al., Nature Methods, 2010, (2): 111 - 118). Las estrategias de enriquecimiento genómico son importantes, ya que permiten centrarse en una región genómica concreta que, en comparación con el análisis completo del genoma, es más efectiva en tiempo y coste, y también mucho menos difícil de analizar. Existen diferentes estrategias de enriquecimiento genómico. Por ejemplo, la realización de una reacción de PCR usando un único par de cebadores amplificará una región genómica y, por tanto, enriquecerá esa región genómica. Sin embargo, el tamaño del producto de la PCR que se puede hacer es limitado. Los protocolos de PCR largos tienen en la actualidad un límite superior de 10 - 40 kB que pueden amplificarse (Cheng et al., Proc Natl Acad Sci USA, 1994; 91 (12): 5695 - 5699), pero estos planteamientos tienden a carecer de solidez y cada PCR requiere una optimización y una validación y, aun así, el límite de tamaño es limitado. Para aumentar el tamaño de las regiones que pueden amplificarse, así como la solidez del ensayo, se han desarrollado planteamientos en mosaico usando una multitud de pares de cebadores de PCR diseñados específicamente para una región genómica de interés. Estos cebadores se usan, por ejemplo, en un planteamiento de PCR múltiple o en una PCR RainDance. Diversos métodos enzimáticos, como la circularización de la diana, son compatibles con tales estrategias de amplificación dirigida. Otros métodos implican el uso de sondas de captura, en un *array* o en solución, en donde se usan sondas de 60 - 120 bases de longitud para capturar la región genómica de interés mediante hibridación.

Como resulta claro de los ejemplos anteriores, para enriquecer una región genómica de interés se requiere de antemano información de la secuencia a lo largo de la región genómica de interés, porque esto es necesario para el diseño de sondas y/o cebadores para capturar y/o amplificar la región genómica de interés. Por ejemplo, para enriquecer una secuencia de 30 Mb, se requerirían típicamente 6.000 PCRs distintas. Con las sondas de captura, se requiere aún más información de la secuencia, ya que se necesitarían al menos hasta 250.000 sondas de 120 bp y han de diseñarse para capturar una secuencia de 30 Mb. Estos ensayos son sesgados usando datos de secuencia para las sondas y/o cebadores que cubren ampliamente la región genómica de interés. No recogen secuencias que se desvían demasiado de las secuencias molde diseñadas y, por tanto, no detectarán, por ejemplo, las inserciones. Además, estos planteamientos requieren fragmentar el ADN en, típicamente, secuencias de unos pocos 100 de pares de bases antes del análisis. Esto significa que la región genómica de interés se fragmenta en muchas partes, lo que tiene como resultado la pérdida de información, entre otras cosas, con respecto a redistribuciones dentro de la región de interés. Por ello, existe la necesidad de estrategias mejoradas de enriquecimiento genómico que sean mucho menos sesgadas, que no requieran miles de secuencias cortas, y que permitan la secuenciación completa hipotéticamente neutral de la región de interés.

En el estudio de la arquitectura nuclear de mamíferos, se han desarrollado ensayos de captura de la conformación cromosómica (3C/4C), con los que puede analizarse la organización estructural de una región genómica (documentos WO 2007/004057, WO 2008/08845). Estas tecnologías implican la reticulación *in vivo* de células, p. ej. con formaldehído, de manera que la arquitectura de la cromatina incluyendo el ADN se fija en su arquitectura tridimensional. A continuación, la cromatina se fragmenta, p. ej. con una enzima de restricción, seguido de ligación de los fragmentos de ADN reticulados. El resultado es que los fragmentos de ADN que están cerca uno del otro se ligan. Los productos de ligación se amplifican luego mediante PCR y se analizan en relación con la frecuencia de interacción de los fragmentos de ADN ligados, lo que es indicativo de la proximidad de los fragmentos. La amplificación por PCR puede basarse en una secuencia diana dentro de la región genómica de interés. Una alta frecuencia de interacción con la región genómica de interés indica una proximidad cercana, una baja frecuencia de interacción indica una proximidad distante. Para identificar los fragmentos de ADN, se requiere información de la secuencia. Dicha información de la secuencia se puede proporcionar mediante la detección de fragmentos amplificados con un *microarray*, que comprende sondas, o mediante la secuenciación de una pequeña parte de fragmentos amplificados (típicamente, un mínimo de 20 a 30 pares de bases es suficiente para identificar la correspondiente posición en un genoma). En cualquier caso, el número de fragmentos de ADN identificados, es decir, la frecuencia de interacción, indica la proximidad del fragmento al punto de vista, información que puede usarse para determinar las interacciones intracromosómicas e intercromosómicas.

### Compendio de la invención.

Ahora se ha encontrado que el procedimiento de reticulación y fragmentación de ADN dentro de una célula y la ligación subsiguiente de fragmentos de ADN reticulados, puede proporcionar un punto de partida ideal para analizar una región genómica de interés que comprende una secuencia de nucleótidos diana, es decir, la plantilla cromosómica lineal que rodea una secuencia de nucleótidos diana. La invención se basa en el concepto de que la reticulación del ADN se producirá preferiblemente en aquellas secuencias que están próximas en el molde de cromosoma lineal a una secuencia de nucleótidos diana. El formaldehído, por ejemplo, puede usarse como agente de reticulación. Después de la reticulación, el ADN puede someterse a tratamientos (enzimáticos), es decir, fragmentación y ligación, mientras el ADN permanece en su estado reticulado. Solo pueden ligarse los fragmentos reticulados que están próximos entre sí. Los fragmentos de ADN que se unen al fragmento de ADN que comprende la secuencia de nucleótidos diana son de hecho representativos de la región genómica de interés que comprende la secuencia de nucleótidos diana. Esto es porque, como promedio, la posibilidad de reticulación intracromosómica es siempre mayor que las frecuencias de reticulación intercromosómica. En general, la posibilidad de que diferentes fragmentos se reticulen se correlaciona inversamente con la distancia lineal. Como estimación, y dependiendo de las condiciones reales de reticulación, el 20 - 30% de los fragmentos ligados con un nucleótido diana de interés se localizan dentro de 0,5 Mb de la secuencia de nucleótidos diana, mientras que el 50 - 80% de los fragmentos ligados con un nucleótido diana de interés se originan del cromosoma que comprende la secuencia de nucleótidos diana. Los fragmentos de ADN ligados que comprenden la secuencia de nucleótidos diana, y por tanto la región genómica de interés, pueden amplificarse, es decir enriquecerse, usando uno o más cebadores de oligonucleótidos que reconocen la secuencia de nucleótidos diana. La secuencia de la región genómica de interés puede determinarse a continuación usando tecnologías de secuenciación (de alto rendimiento) bien conocidas en la técnica. El método es poco sesgado, dado que no se requiere una extensa información de secuencia para enfocar la región genómica de interés. Por ejemplo, una región genómica de interés puede comprender un alelo de interés. Se puede seleccionar una secuencia de nucleótidos diana de manera que no esté dentro de la secuencia del alelo de interés. Después puede amplificarse una región genómica de interés usando una secuencia de nucleótidos diana, sin requerir información de secuencia del alelo de interés. Por tanto, el alelo de interés puede enriquecerse sin requerir ninguna secuencia de ese alelo. El efecto es que el método de enriquecimiento no está sesgado mediante el uso de oligonucleótidos y/o sondas que cubren la secuencia alélica de interés. Además, como la etapa de ligación implica la ligación de fragmentos que están en proximidad entre sí, el método también puede permitir el análisis de secuencias de alelos separados. Por ejemplo, cuando una muestra de ADN reticulado comprende múltiples alelos (por ejemplo, a causa de que la muestra de ADN proviene de una población de células heterogéneas, o porque la ploidía es mayor que uno), cada alelo puede tener un vecindario genómico diferente. Un fragmento de ADN, que comprende una secuencia de nucleótidos diana, solo interactuará con fragmentos de ADN que estén en el mismo espacio. Así pues, los fragmentos de ADN ligados son representativos del entorno genómico del que se originan los fragmentos. Determinando al menos parte de la secuencia de todos los diferentes fragmentos de ADN ligados, las secuencias de fragmentos de ADN pueden acoplarse a continuación usando la información de secuencia de los diferentes fragmentos de ADN ligados y puede construirse una secuencia para regiones genómicas distintas de interés.

### Definiciones.

En la descripción y ejemplos que siguen se usan diversos términos y expresiones. Para proporcionar una comprensión clara y coherente de la especificación y las reivindicaciones, incluido el alcance que deben darse a dichos términos, se proporcionan las definiciones que siguen. A menos que se defina otra cosa en el presente documento, todos los términos técnicos y científicos usados tienen el mismo significado que entiende comúnmente un experto habitual en la técnica a la que pertenece esta invención.

Los métodos para llevar a cabo las técnicas convencionales usadas en los métodos de la invención serán evidentes para un profesional experto. La práctica de técnicas convencionales en biología molecular, bioquímica, química computacional, cultivo celular, ADN recombinante, bioinformática, genómica, secuenciación y campos relacionados son bien conocidas por los expertos en la materia y se discuten, por ejemplo, en las referencias bibliográficas siguientes: Sambrook et al., *Molecular Cloning. A Laboratory Manual*, 2ª Edición, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989; Ausubel y col., *Current Protocols in Molecular Biology*, John Wiley & Sons, Nueva York, 1987 y actualizaciones periódicas; y la serie *Methods in Enzymology*, Academic Press, San Diego.

Como se usa en el presente texto, las formas singulares "un", "una", "el" y "la" incluyen referencias plurales a menos que el contexto indique claramente otra cosa. Por ejemplo, un método para aislar "una" molécula de ADN, como se usa anteriormente, incluye aislar una pluralidad de moléculas (por ejemplo, decenas, centenares, miles, decenas de miles, centenares de miles, millones, o más moléculas).

Una "región genómica de interés" es una secuencia de ADN de un organismo del cual es deseable determinar, al menos en parte, la secuencia de ADN. Por ejemplo, una región genómica que se sospecha que comprende un alelo asociado con una enfermedad puede ser una región genómica de interés. Como se usa en el presente texto, el término "alelo" o "alelos" significa cualquiera de una o más formas alternativas de un gen en un locus concreto. En una célula diploide de un organismo, los alelos de un gen dado se encuentran localizados en una ubicación concreta, o locus (loci en plural) en un cromosoma. Un alelo está presente en cada cromosoma del par de cromosomas homólogos. Por tanto, en una célula diploide pueden existir dos alelos y, por lo tanto, dos regiones de

genómicas separadas (distintas) de interés.

Un "ácido nucleico" como se define en el presente texto puede incluir cualquier polímero u oligómero de bases de pirimidina y purina, preferiblemente citosina, timina y uracilo, y adenina y guanina, respectivamente (Véase Albert L. Lehninger, Principles of Biochemistry, en 793 - 800 (Worth Pub. 1982).

5 Como se define en el presente texto, se contempla cualquier componente de desoxirribonucleótido, ribonucleótido o ácido nucleico peptídico, y cualquier variante química del mismo, tales como formas metiladas, hidroximetiladas o glucosiladas de estas bases, y similares. Los polímeros u oligómeros pueden ser de composición heterogénea u homogénea, y pueden aislarse de fuentes de origen natural o pueden producirse artificial o sintéticamente. Además, los ácidos nucleicos pueden ser ADN o ARN, o una mezcla de los mismos, y pueden existir permanentemente o transitoriamente en forma monocatenaria o bicatenaria, incluyendo homodúplex, heterodúplex y estados híbridos.

10 Un "ADN de muestra" es una muestra que se obtiene de un organismo o de un tejido de un organismo, o cultivo de tejido y/o celular, que comprende ADN. Se puede obtener una muestra de ADN de un organismo a partir de cualquier tipo de organismo, p. ej. microorganismos, virus, plantas, hongos, animales, seres humanos y bacterias, o combinaciones de los mismos. Por ejemplo, una muestra de tejido de un paciente humano sospechoso de una infección bacteriana y/o viral puede comprender células humanas, pero también virus y/o bacterias. La muestra puede comprender células y/o núcleos celulares. El ADN de muestra puede ser de un paciente o de una persona que puede estar en riesgo o se sospecha que tiene una enfermedad en particular, por ejemplo cáncer o cualquier otra enfermedad que justifique la investigación del ADN del organismo.

20 Con "reticulación" de acuerdo con la invención se entiende hacer reaccionar ADN en dos posiciones diferentes, de manera que estas dos posiciones diferentes pueden estar conectadas. La conexión entre las dos posiciones diferentes puede ser directamente, formando un enlace covalente entre las cadenas de ADN. Dos cadenas de ADN pueden reticularse directamente usando irradiación con UV, formando enlaces covalentes directamente entre cadenas de ADN. La conexión entre las dos posiciones diferentes puede ser indirectamente, a través de un agente, p. ej. una molécula de reticulante. Una primera sección de ADN puede estar conectada a un primer grupo reactivo de una molécula reticulante que comprende dos grupos reactivos, ese segundo grupo reactivo de la molécula reticulante puede estar conectado a una segunda sección de ADN, reticulando la primera y la segunda secciones de ADN indirectamente a través de la molécula reticulante. Una reticulación también puede formarse indirectamente entre dos cadenas de ADN a través de más de una molécula. Por ejemplo, una molécula reticulante típica que se puede usar es formaldehído. El formaldehído induce reticulaciones proteína-proteína y ADN-proteína. Por lo tanto, el formaldehído puede reticular diferentes cadenas de ADN entre ellas a través de sus proteínas asociadas. Por ejemplo, el formaldehído puede reaccionar con una proteína y ADN, conectando una proteína y ADN a través de la molécula de reticulante. Por tanto, dos secciones de ADN pueden reticularse usando formaldehído formando una conexión entre una primera sección de ADN y una proteína, la proteína puede formar una segunda conexión con otra molécula de formaldehído que se conecta a una segunda sección de ADN, formando así una reticulación que puede representarse como DNA1-reticulante-proteína-reticulante-DNA2. En todo caso, se entiende que la reticulación como se define en el presente texto implica formar conexiones (directa o indirectamente) entre cadenas de ADN que están en proximidad física entre sí. Las cadenas de ADN pueden estar en proximidad física una de otra en la célula, ya que el ADN está altamente organizado, mientras que está separado de un punto de vista de la secuencia, p. ej. por 100 kb. Siempre que el método de reticulación sea compatible con las etapas subsiguientes de fragmentación y ligación, tal reticulación puede contemplarse con el propósito del método de la invención.

Una "muestra de ADN reticulado" es una muestra de ADN que ha sometida a reticulación. La reticulación del ADN de la muestra tiene el efecto de que el estado tridimensional del ADN dentro de la muestra permanece en gran parte intacto. De esta manera, las cadenas de ADN que están en proximidad física unas de otras permanecen en la vecindad de cada una.

45 La "reticulación inversa" como se define en el presente texto comprende romper los enlaces cruzados de manera que el ADN que ha sido reticulado ya no lo está y es adecuado para posteriores etapas de amplificación y/o secuenciación. Por ejemplo, la realización de un tratamiento con proteasa K en una muestra de ADN que ha sido reticulado con formaldehído digerirá la proteína presente en la muestra. Como el ADN reticulado está conectado indirectamente a través de proteína, el tratamiento con proteasa en sí mismo puede revertir la reticulación entre el ADN. Sin embargo, los fragmentos de proteína que permanecen conectados al ADN pueden obstaculizar la secuenciación y/o la amplificación posteriores. Por tanto, invertir las conexiones entre el ADN y la proteína también puede dar como resultado una "reticulación inversa". La conexión ADN-reticulante-proteína puede invertirse mediante una etapa de calentamiento, por ejemplo incubando a 70 °C. Como en una muestra de ADN hay presentes grandes cantidades de proteína, a menudo es deseable digerir la proteína además con una proteasa. Por tanto, se puede contemplar cualquier método de "reticulación inversa" en el que las hebras de ADN que están conectadas en una muestra reticulada se hacen adecuadas para la secuenciación y/o amplificación.

"Fragmentación de ADN" incluye cualquier técnica que, cuando se aplica al ADN, que puede ser ADN reticulado o no, o cualquier otro ADN, tiene como resultado fragmentos de ADN. Las técnicas bien conocidas en este campo son sonicación, cizallamiento y/o restricción enzimática, pero también se pueden concebir otras técnicas.

Una "endonucleasa de restricción" o "enzima de restricción" es una enzima que reconoce una secuencia de nucleótidos específica (sitio de reconocimiento) en una molécula de ADN bicatenario, y escindiría ambas cadenas de la molécula de ADN en cada sitio de reconocimiento o en sus proximidades, dejando un extremo romo o un borde colgante 3' o 5'. La secuencia de nucleótidos específica que se reconoce puede determinar la frecuencia de escisión, p. ej. una secuencia de nucleótidos de 6 nucleótidos tiene lugar como promedio cada 4096 nucleótidos, mientras que una secuencia de nucleótidos de 4 nucleótidos tiene lugar con mucha más frecuencia, como promedio cada 256 nucleótidos.

La "ligadura" como se define en el presente texto implica la reunión de fragmentos de ADN separados. Los fragmentos de ADN pueden tener extremos romos, o pueden tener colgantes compatibles (colgantes pegajosos) de manera que los salientes pueden hibridarse entre sí. La unión de los fragmentos de ADN puede ser enzimática, con una enzima ligasa, ADN ligasa. Sin embargo, también puede usarse una ligación no enzimática, siempre que los fragmentos de ADN se unan, es decir, siempre que formen un enlace covalente. Típicamente se forma un enlace fosfodiéster entre el grupo hidroxilo y el fosfato de las cadenas distintas.

"Cebadores de oligonucleótidos", en general, se refiere a cadenas de nucleótidos que pueden estimular la síntesis de ADN. La ADN polimerasa no puede sintetizar ADN *de novo* sin cebadores. Un cebador hibrida con el ADN, es decir, se forman pares de bases. Los nucleótidos que pueden formar pares de bases, que son complementarios entre sí, son p. ej. citosina y guanina, timina y adenina, adenina y uracilo, y guanina y uracilo. La complementariedad entre el cebador y la cadena de ADN existente no tiene que ser del 100%, es decir, no todas las bases de un cebador necesitan un par de bases con la cadena de ADN existente. Desde el extremo 3' de un cebador hibridado con la cadena de ADN existente, los nucleótidos se incorporan usando la cadena existente como plantilla (síntesis de ADN dirigida por plantilla). Se puede uno referir a las moléculas de oligonucleótidos sintéticos que se usan en una reacción de amplificación como "cebadores".

"Amplificar" se refiere a una reacción de amplificación o multiplicación de polinucleótidos, concretamente una población de polinucleótidos que se replican a partir de una o más secuencias de partida. La amplificación puede referirse a una variedad de reacciones de amplificación, que incluyen, pero sin limitarse a ellas, la reacción en cadena de la polimerasa (PCR), reacciones de polimerasa lineal, amplificación basada en la secuencia de ácido nucleico, amplificación de círculo rodante y reacciones similares.

"Secuenciación" se refiere a la determinación del orden de los nucleótidos (secuencias de bases) en una muestra de ácido nucleico, p. ej. ADN o ARN. Se dispone de muchas técnicas, como la secuenciación de Sanger y las tecnologías de secuenciación de alto rendimiento, como las ofrecidas por Roche, Illumina y Applied Biosystems.

El término "cóntigo" ("*contig*") se usa en relación con el análisis de la secuencia de ADN, y se refiere a los tramos de ADN contiguos reensamblados derivados de dos o más fragmentos de ADN que tienen secuencias de nucleótidos contiguas. Por lo tanto, un cóntigo puede ser un conjunto de fragmentos de ADN solapantes que proporciona una secuencia contigua (parcial) de una región genómica de interés. Un cóntigo también puede ser un conjunto de fragmentos de ADN que, cuando se alinean con una secuencia de referencia, pueden formar una secuencia de nucleótidos contigua. Por ejemplo, el término "cóntigo" abarca un conjunto de fragmentos de ADN (ligado) que se ordenan de forma tal que tienen un solapamiento de secuencia de cada fragmento de ADN (ligado) con al menos uno de sus vecinos. El fragmento o fragmentos de ADN enlazado o acoplado (ligado) puede ordenarse manualmente o, preferiblemente, utilizando programas informáticos apropiados tales como FPC, PHRAP, CAP3, etc., y también se pueden agrupar en cóntigos distintos.

Un "adaptador" es una molécula de oligonucleótido bicatenario corta, con un número limitado de pares de bases, p. ej. de aproximadamente 10 a aproximadamente 30 pares de bases de longitud, que están diseñados de manera que puedan ligarse a los extremos de los fragmentos. Los adaptadores están compuestos generalmente por dos oligonucleótidos sintéticos que tienen secuencias de nucleótidos que son parcialmente complementarias entre sí. Cuando se mezclan los dos oligonucleótidos sintéticos en solución bajo condiciones apropiadas, se emparejarán entre sí formando una estructura bicatenaria. Después de la hibridación, se puede diseñar un extremo de la molécula adaptadora de manera que sea compatible con el extremo de un fragmento de restricción y se pueda ligar al mismo; el otro extremo del adaptador puede diseñarse de forma que no se pueda ligar, pero no es necesario que sea este el caso, por ejemplo, cuando un adaptador se va a ligar entre fragmentos de ADN.

Un "identificador" es una secuencia corta que puede añadirse a un adaptador o un cebador o incluirse en su secuencia o utilizarse como marcador para proporcionar un identificador único. Tal identificador de secuencia (o etiqueta) puede ser una secuencia de bases única de longitud variable pero definida, típicamente de 4 - 16 pb usada para identificar una muestra concreta de ácido nucleico. Por ejemplo, los marcadores de 4 pb permiten  $4(\text{exp}4) = 256$  etiquetas diferentes. Los ejemplos típicos son secuencias ZIP, conocidas en la técnica como marcadores usados normalmente para detección única por hibridación (Iannone y col., Cytometry 39: 131 - 140, 2000). Los identificadores son útiles en un método de la invención ya que, al usar dicho identificador, el origen de una muestra (PCR) puede determinarse con un procesamiento posterior. En el caso de combinar productos procesados que se originan de diferentes muestras de ácidos nucleicos, las diferentes muestras de ácidos nucleicos pueden identificarse usando identificadores diferentes. Por ejemplo, dado que de acuerdo con la invención la secuenciación puede realizarse usando una secuenciación de alto rendimiento, pueden combinarse múltiples muestras. Los

identificadores pueden ayudar a identificar las secuencias correspondientes a las diferentes muestras. Los identificadores también se pueden incluir en adaptadores para la ligación a fragmentos de ADN que ayudan en la identificación de secuencias de fragmentos de ADN. Los identificadores preferiblemente difieren entre sí en al menos dos pares de bases y preferiblemente no contienen dos bases consecutivas idénticas para evitar errores de lectura. La función del identificador puede combinarse a veces con otras funcionalidades, como adaptadores o cebadores.

La "selección de tamaño" como se define en el presente texto implica técnicas con las cuales se seleccionan los rangos de tamaño particulares de las moléculas, p. ej. fragmentos de ADN (ligado) o fragmentos de ADN (ligado) amplificados. Las técnicas que se pueden usar son, por ejemplo, electroforesis en gel, exclusión por tamaño, cromatografía de extracción en gel, pero no están limitadas a ellas, siempre y cuando se puedan seleccionar moléculas con un tamaño particular, tal técnica será suficiente.

Con los términos "alineación" y "alineamiento" se entiende la comparación de dos o más secuencias de nucleótidos basándose en la presencia de tramos cortos o largos de nucleótidos idénticos o similares. Los métodos y programas informáticos para la alineación son bien conocidos en la técnica. Un programa informático que se puede usar o adaptar para la alineación es "Align 2", creado por Genentech, Inc., que se archivó con la documentación del usuario en la Oficina de Derechos de Autor de los Estados Unidos, Washington, DC 20559, el 10 de diciembre de 1991.

### Figuras.

La Figura 1 muestra un esquema de un método para determinar la secuencia de una región genómica de interés de acuerdo con la invención. El método implica:

- (a) reticulación, en donde p. ej. la fijación de formaldehído reticula espacialmente las secuencias de ADN cercanas en el núcleo (N) (siendo a menudo secuencias próximas al cromosoma (Ch), p. ej. secuencias del mismo gen) a través de sus proteínas asociadas (por ejemplo, histonas). Se indican 5 fragmentos hipotéticos de una región genómica de interés A, B, C, D y E.
- (b) a continuación, el ADN de la muestra reticulado se fragmenta, p. ej. llevando a cabo una digestión con una enzima de restricción (p. ej. un cortador frecuente (cuatro) (p. ej. NlaIII);
- (c) fragmentos de restricción reticulados se ligan para formar círculos de ADN;
- (d) después de la reversión de la reticulación, se lleva a cabo una etapa de amplificación, con un *primerset* de PCR (inverso) para un punto de vista cercano o dentro de la región genómica de interés. Los fragmentos (A, B, C, D y E) reticulados a este punto de vista se amplifican y se enriquecen sobre el resto del genoma. Los fragmentos amplificados se secuencian, por ejemplo, mediante secuenciación en círculos enteros (lecturas largas), el material amplificado por PCR también puede fragmentarse primero para crear una biblioteca de secuenciación compatible, por ejemplo, para la secuenciación Illumina o SOLiD.
- (e) a continuación se construye un *cóntigo* a partir de las lecturas, las secuencias puede compararse con un genoma de referencia para identificar la variación genética.

La Figura 2 muestra un esquema del gen BRCA1 con 5 puntos de vista diferentes (A, B, C, D y E). La flecha negra indica la dirección del sentido. Los números en los círculos con las flechas indican la posición en la secuencia del gen. El punto de vista E está al inicio del gen y el punto de vista A al final. Los puntos de vista están separados por aproximadamente 15 - 25 kB.

La Figura 3 muestra la electroforesis en gel de las muestras de ADN tomadas durante la preparación de un ADN de muestra reticulado para la secuenciación del gen BRCA1 como se describe en los ejemplos.

(A) El carril M indica un ADN marcador lambda ADN PstI, el carril 1 muestra un control no digerido, el carril 2 muestra un primer control digerido con NlaIII, el carril 3 es un control de ligación después de la ligación de la primera muestra digerida NlaIII, el carril 4 muestra una segunda digestión con Nspl.

(B) El carril M muestra un marcador lambda DNA PstI. Los carriles A, B, C, D y E muestran los productos de amplificación de las diferentes amplificaciones de ADN, correspondientes a las muestras de la etapa 67 como se describe en la sección de ejemplo, y correspondientes a los puntos de vista descritos en la figura 2.

### Descripción detallada de la invención.

De acuerdo con un aspecto de la invención, se proporciona un método para construir un *cóntigo* de una región genómica de interés que comprende una secuencia de nucleótidos diana, el cual método comprende fragmentar un ADN reticulado, ligar el ADN reticulado fragmentado, revertir la reticulación y determinar al menos parte de la secuencias de fragmentos de ADN ligados, que comprenden una secuencia de nucleótidos diana, y usar las secuencias determinadas para construir una secuencia de la región genómica de interés.

Una muestra de ADN reticulado comprende un ADN de muestra que se ha sometido a reticulación. La reticulación del ADN de muestra tal como está presente en la muestra da como resultado el mantenimiento en gran parte de la

arquitectura tridimensional del ADN. Por ejemplo, un agente de reticulación estándar que puede usarse es el formaldehído. Se pueden tomar muestras de un paciente y/o de tejido enfermo, y también se pueden derivar de otros organismos o de secciones distintas del mismo organismo, tales como muestras de un paciente, una muestra de tejido sano y una muestra de tejido enfermo. Entonces las muestras pueden analizarse de acuerdo con el método de la invención y compararse con una muestra de referencia, o pueden analizarse diferentes muestras y compararse entre ellas. Por ejemplo, si se sospecha que un paciente tiene cáncer de mama, se puede obtener una biopsia del tumor sospechoso. Se puede obtener otra biopsia de tejido no enfermo. Ambas biopsias de tejido se pueden analizar de acuerdo con la invención. Las regiones genómicas de interés pueden ser el gen BRCA1 y BRCA2, genes que tienen 83 y 86 kb de longitud (revisado en Mazoyer, 2005, Human Mutation 25: 415 - 422). Determinando la secuencia de interés de la región genómica de acuerdo con la invención y comparando las secuencias de la región genómica de las diferentes biopsias entre sí y/o con una secuencia del gen BRCA de referencia, se pueden encontrar mutaciones genéticas que ayudarán a diagnosticar al paciente y/o a determinar el tratamiento del paciente y/o a predecir el pronóstico del progreso de la enfermedad.

Fragmentando una muestra de ADN reticulado, los fragmentos de ADN que se originan a partir de una región genómica de interés permanecen próximos entre sí porque están reticulados. Cuando estos fragmentos de ADN reticulados se ligan posteriormente, se ligan fragmentos de ADN de la región genómica de interés, que están en proximidad entre ellos debido a los enlaces cruzados. Este tipo de ligadura puede también denominarse ligadura de proximidad. Los fragmentos de ADN que comprenden la secuencia de nucleótidos diana se pueden ligar con fragmentos de ADN dentro de una gran distancia lineal en el nivel de la secuencia. Determinando la secuencia (al menos parte de ella) de fragmentos ligados que comprenden el fragmento que comprende la secuencia de nucleótidos diana, se obtienen secuencias de fragmentos de ADN dentro del entorno espacial de la región genómica de interés. Es probable que cada secuencia de nucleótidos diana individuales esté reticulada con otros muchos fragmentos de ADN. Como consecuencia, a menudo puede ligarse más de un fragmento de ADN a un fragmento que comprende la secuencia de nucleótidos diana. Combinando secuencias (parciales) de los fragmentos de ADN ligados (amplificados) que se ligaron con un fragmento que comprende la secuencia de nucleótidos diana, se puede construir una secuencia de la región genómica de interés. Un fragmento de ADN ligado con el fragmento que comprende la secuencia de nucleótidos diana incluye cualquier fragmento que pueda estar presente en los fragmentos de ADN ligados.

Se conocen métodos en la técnica que implican reticular ADN, así como fragmentar y ligar los fragmentos de ADN (por ejemplo, documentos WO 2007/004057 o WO 2008/08845). Tales métodos están destinados a identificar frecuencias de interacción entre diferentes fragmentos de ADN, no a identificar la secuencia primaria de nucleótidos de los fragmentos que se encuentran próximos a una secuencia de nucleótidos diana. La idea original de utilizar 4C para la detección de frecuencias de interacción solo requería una lectura de secuencia corta. La frecuencia de las lecturas de secuencias cortas interactuantes se representa gráficamente frente a las posiciones cromosómicas de las lecturas. El patrón de tal gráfica es indicativo de si una región genómica particular de interés puede interactuar con una región en otra parte del genoma, o que, p. ej. se han producido translocaciones entre cromosomas. Por ejemplo, en el caso de que se observe una alta frecuencia de lecturas en un cromosoma distinto del que contiene la secuencia de nucleótidos diana, esto es indicativo de una translocación. En la presente invención, la frecuencia de interacción no está determinada. En la presente descripción se cayó ahora en la cuenta de que fragmentando el ADN reticulado y ligando posteriormente los fragmentos de ADN se captura de hecho una región genómica que rodea la secuencia de nucleótidos diana, que cuando se secuencia permite la reconstrucción de cóntigos de la región genómica. Mientras que en los métodos conocidos en la técnica, el foco ha sido determinar la frecuencia de interacción de lecturas de secuencia cortas con una secuencia de nucleótidos diana, el foco de la presente descripción se dirige a determinar la secuencia completa, o al menos una gran parte de ella, de fragmentos de ADN ligados (que comprende el fragmento de ADN con el nucleótido diana) de manera que a partir de las secuencias de fragmentos de ADN y el acoplamiento de fragmentos de ADN ligados, pueden construirse cóntigos para una región genómica de interés.

#### **Fragmentos ligados linealizados.**

De acuerdo con un segundo aspecto de la invención, se proporciona un método para edificar un cóntigo de una región genómica de interés que comprende una secuencia de nucleótidos diana, que comprende las etapas de:

- a) proporcionar una muestra de ADN reticulado;
- b) fragmentar el ADN reticulado;
- c) ligar el ADN fragmentado reticulado;
- d) revertir la reticulación;
- e) fragmentar opcionalmente el ADN de la etapa d), preferiblemente con una enzima de restricción;
- f) opcionalmente, ligar el ADN fragmentado de la etapa d) o e) a al menos un adaptador;
- g) opcionalmente, amplificar el ADN de la etapa d) o e) que comprende la secuencia de nucleótidos diana

usando al menos un cebador de oligonucleótido que se hibrida con la secuencia de nucleótidos diana, o amplificar el ADN de la etapa f) usando al menos un cebador adicional que se hibrida con el al menos un adaptador;

- 5 h) determinar al menos parte de la secuencia de los fragmentos de ADN ligados (amplificados) de las etapas d), e), f) o g) que comprende la secuencia de nucleótidos diana preferiblemente usando secuenciación de alto rendimiento;
- i) construir un cóntigo de la región genómica de interés a partir de las secuencias determinadas.

En la etapa a) se proporciona una muestra de ADN reticulado como se bosqueja en otra parte en el presente documento. La muestra de ADN reticulado es fragmentada en la etapa b). Por la fragmentación del ADN reticulado se producen fragmentos de ADN que se mantienen unidos por los enlaces cruzados. La etapa de fragmentación b) puede comprender sonicación, y puede ir seguida de reparación enzimática del extremo del ADN. Los resultados de la sonicación en la fragmentación de ADN en sitios aleatorios, que pueden ser de extremos romos o pueden tener colgajos 3' o 5', ya que estos puntos de rotura de ADN se presentan de forma aleatoria, el ADN puede repararse (enzimáticamente), rellenando posibles colgajos 3' o 5', de manera que se obtienen fragmentos de ADN que tienen extremos romos que permiten la ligación de los fragmentos a los adaptadores y/o entre sí en la etapa c) posterior. Alternativamente, pueden también hacerse colgajos de extremos romos mediante la eliminación de nucleótidos colgantes, usando p. ej. exonucleasas. La etapa de fragmentación b) también puede comprender la fragmentación con una o más enzimas de restricción, o combinaciones de las mismas. La fragmentación con una enzima de restricción es ventajosa ya que puede permitir el control del tamaño medio de fragmento. Los fragmentos que se forman pueden tener colgajos compatibles o extremos romos que permiten la ligadura de los fragmentos en el paso c) siguiente. Además, cuando se divide una muestra de ADN reticulado en una pluralidad de submuestras, para cada submuestra se pueden usar enzimas de restricción con diferentes sitios de reconocimiento. Esto es ventajoso porque al usar diferentes enzimas de restricción que tienen diferentes sitios de reconocimiento, se pueden obtener diferentes fragmentos de ADN de cada submuestra.

En la etapa c) siguiente, los fragmentos se ligan. Dado que un fragmento que comprende una secuencia de nucleótidos diana se puede reticular con otros múltiples fragmentos de ADN, se puede ligar más de un fragmento de ADN al fragmento que comprende la secuencia de nucleótidos diana. Esto puede dar como resultado combinaciones de fragmentos de ADN que están próximos entre sí, ya que se mantienen unidos por los enlaces cruzados. Se pueden formar diferentes combinaciones y/o diferente orden de los fragmentos de ADN en los fragmentos de ADN ligados. En caso de que los fragmentos de ADN se obtengan mediante restricción enzimática, se conoce el sitio de reconocimiento de la enzima de restricción, lo que hace que sea posible identificar los fragmentos como restos o sitios de reconocimiento de enzimas de restricción reconstituidos que pueden indicar la separación entre diferentes fragmentos de ADN. En el caso de que los fragmentos de ADN se obtuvieran mediante fragmentación aleatoria, como sonicación y posterior reparación enzimática del extremo del ADN, puede ser más difícil distinguir un fragmento de otro. Al margen del método de fragmentación que se use, la etapa de ligamiento c) puede realizarse en presencia de un adaptador, ligando secuencias adaptadoras entre los fragmentos. Alternativamente, el adaptador puede ligarse en una etapa distinta. Esto es ventajoso porque los diferentes fragmentos pueden identificarse fácilmente identificando las secuencias adaptadoras que están situadas entre los fragmentos. Por ejemplo, en el caso de que los extremos del fragmento de ADN fueran extremos romos, la secuencia adaptadora sería adyacente a cada uno de los extremos del fragmento de ADN, indicando el límite entre los fragmentos de ADN distintos. A continuación, la reticulación se revierte en la etapa d), lo que da como resultado un conjunto de fragmentos de ADN ligados que comprenden dos o más fragmentos. Una subpoblación del conjunto de fragmentos de ADN ligados comprende un fragmento de ADN que comprende la secuencia de nucleótidos diana. Al revertir la reticulación, se libera la fijación estructural o espacial del ADN y la secuencia de ADN queda disponible para las etapas posteriores, p. ej. amplificación y/o secuenciación, ya que el ADN reticulado puede no ser un sustrato adecuado para tales etapas. Las etapas siguientes e) y/o f) pueden realizarse después de la reversión de la reticulación; sin embargo, las etapas e) y/o f) también pueden realizarse mientras los fragmentos de ADN ligados todavía están en estado reticulado.

Los fragmentos de ADN ligados pueden ser fragmentados opcionalmente en la etapa e), preferiblemente con una enzima de restricción. La primera etapa de fragmentación y la segunda etapa de fragmentación opcional pueden dirigirse a obtener fragmentos de ADN ligados de un tamaño que sea compatible con la etapa posterior de amplificación y/o la etapa de determinación de la secuencia. Además, una segunda etapa de fragmentación, preferiblemente con una enzima, puede dar como resultado extremos de fragmentos ligados que son compatibles con la ligación opcional de un adaptador en la etapa f). La segunda etapa de fragmentación se puede realizar después de invertir la reticulación, si bien es también posible realizar la segunda etapa de fragmentación e) y/o la etapa de ligación f) mientras los fragmentos de ADN están aún reticulados.

En caso de que la etapa de fragmentación b) y e) comprenda enzimas de restricción, se prefiere que el sitio de reconocimiento de la enzima de restricción de la etapa e) sea más largo que el sitio de reconocimiento de la etapa b). La enzima de e) corta así a una frecuencia más baja que la etapa b). Esto significa que el tamaño del fragmento de ADN promedio de la etapa b) es menor que el tamaño promedio del fragmento de la etapa e) obtenido después de la restricción del ADN. De esta manera, en la primera etapa de fragmentación se forman fragmentos relativamente pequeños, que posteriormente se ligan. Como la segunda enzima de restricción de la etapa e) corta



menos frecuentemente que la etapa b), la mayoría de los fragmentos de ADN pueden no comprender el sitio de reconocimiento de restricción de la etapa e). Por tanto, cuando los fragmentos de ADN ligados se fragmentan posteriormente en la segunda etapa, muchos de los fragmentos de ADN de la etapa b) pueden permanecer intactos. Esto es útil porque las secuencias combinadas de los fragmentos de ADN de la etapa b) se pueden usar para construir un cóntigo para la región genómica de interés. Si la fragmentación de la etapa b) es menos frecuente que la fragmentación de la etapa c), el resultado sería que los fragmentos de la etapa b) se fragmentan, lo que puede dar como resultado la pérdida de secuencias de ADN relativamente grandes que son útiles para construir un cóntigo. Así pues, independientemente del método que se usase para fragmentar en la etapa b) y e), se prefiere que la fragmentación de la etapa b) sea más frecuente en comparación con la etapa e), de manera que los fragmentos de ADN de la etapa b) puedan quedar en gran parte intactos, es decir, en gran parte no están fragmentados por la etapa e).

A los fragmentos de ADN ligados obtenidos de la etapa d) o e) se liga opcionalmente al menos un adaptador. Los extremos de los fragmentos de ADN ligados necesitan ser compatibles con la ligación de dicho adaptador. Como los fragmentos de ADN ligados de la etapa d) o e) pueden ser ADN lineal, la ligación de un adaptador puede proporcionar una secuencia de hibridación de cebador. La secuencia adaptadora ligada con fragmentos de ADN ligados que comprenden la secuencia de nucleótidos diana, proporcionará moléculas de ADN que pueden amplificarse usando PCR.

En la siguiente etapa g), el ADN de la etapa f) que comprende la secuencia de nucleótidos diana puede amplificarse usando al menos un cebador oligonucleótido que se hibrida con la secuencia de nucleótidos diana, y al menos un cebador adicional que se hibrida con el al menos un adaptador. Como la etapa f) de ligación de un adaptador es opcional, el ADN de la etapa d) o e) que comprende el nucleótido diana también se puede amplificar en la etapa g) usando al menos un cebador oligonucleótido que se hibrida con la secuencia de nucleótidos diana.

A continuación se determina la secuencia de los fragmentos de ADN ligados (amplificados) obtenidos en las etapas d), e), f) o g) que comprenden la secuencia de nucleótidos diana. La determinación de la secuencia se realiza preferiblemente usando tecnología de secuenciación de alto rendimiento, ya que esto es más conveniente y permite determinar un gran número de secuencias para cubrir la región genómica completa de interés. A partir de estas secuencias determinadas, se puede construir un cóntigo de la región genómica de interés. Cuando se determinan las secuencias de los fragmentos de ADN, se pueden obtener lecturas solapantes a partir de las cuales se puede construir la región genómica de interés. En caso de que los fragmentos de ADN se obtuviesen por fragmentación aleatoria, la naturaleza aleatoria de la etapa de fragmentación puede ya dar como resultado fragmentos de ADN que, cuando se secuencian, dan como resultado lecturas superpuestas. Al aumentar el tamaño de la muestra, p. ej. aumentando el número de células analizadas, puede aumentar la fiabilidad de la región genómica de interés que se construye. Alternativamente, cuando en la etapa b) se analiza una pluralidad de submuestras usando diferentes enzimas de restricción, también se obtendrán lecturas solapantes. Aumentando la pluralidad de submuestras, aumentará el número de fragmentos superpuestos, lo que puede aumentar la fiabilidad del cóntigo de la región genómica de interés que se construye. A partir de estas secuencias determinadas que pueden solaparse, se puede construir un cóntigo. Alternativamente, si las secuencias no se solapan, p. ej. cuando puede haber sido usada una única enzima de restricción en la etapa b), la alineación de los fragmentos de ADN (ligados) con una secuencia de referencia puede permitir construir un cóntigo de la región genómica de interés.

Fragmentos ligados circularizados.

En un tercer aspecto de la invención, se proporciona un método para construir un cóntigo de una región genómica de interés que comprende una secuencia de nucleótidos diana, que comprende las etapas de:

- a) proporcionar una muestra de ADN reticulado;
- b) fragmentar el ADN reticulado;
- c) ligar el ADN reticulado fragmentado;
- d) revertir la reticulación;
- e) opcionalmente fragmentar el ADN de la etapa d), preferiblemente con una enzima de restricción;
- f) circularizar el ADN de la etapa d) o e);
- g) opcional y preferiblemente, amplificar el ADN circularizado que comprende la secuencia de nucleótidos diana usando preferiblemente al menos un cebador que se hibrida con la secuencia de nucleótidos diana;
- h) determinar al menos parte de la secuencia de los fragmentos de ADN ligados (amplificados) que comprenden el nucleótido diana usando secuenciación de alto rendimiento;
- i) construir un cóntigo de la región genómica de interés a partir de las secuencias determinadas.

En la etapa a) se proporciona una muestra de ADN reticulado como se describe en otra parte en el presente

documento. La muestra de ADN reticulado se fragmenta en la etapa b). Mediante la fragmentación del ADN reticulado se producen fragmentos de ADN que se mantienen unidos por los enlaces cruzados. La etapa de fragmentación b) puede comprender sonicación, y puede ir seguida de reparación enzimática del extremo del ADN. Los resultados de la sonicación en la fragmentación de ADN en sitios aleatorios, que pueden ser de extremos romos o pueden tener colgajos de 3' o 5', ya que estos puntos de rotura de ADN se presentan de forma aleatoria, el ADN puede repararse (enzimáticamente), rellenando posibles colgajos 3' o 5', de manera que se obtienen fragmentos de ADN que tienen extremos romos que permiten la ligación de los fragmentos con adaptadores o entre sí en la etapa c) siguiente. Alternativamente, los colgajos pueden también hacerse de extremos romos mediante la eliminación de nucleótidos colgantes, usando p. ej. exonucleasas. La etapa de fragmentación b) puede comprender también la fragmentación con una enzima de restricción, o sus combinaciones. La fragmentación con una enzima de restricción es ventajosa ya que permite el control del tamaño de fragmento promedio. Además, los fragmentos que se forman tendrán colgajos compatibles o extremos romos que permiten la ligación de los fragmentos en la etapa c) posterior sin requerir más modificaciones. Además, cuando se divide una muestra de ADN reticulado en una pluralidad de submuestras, para cada submuestra se pueden usar enzimas de restricción con diferentes sitios de reconocimiento. Esto es ventajoso porque usando diferentes enzimas de restricción que tienen sitios de reconocimiento diferentes se pueden obtener diferentes fragmentos de ADN de cada submuestra.

En la siguiente etapa c), los fragmentos se ligan. En caso de que los fragmentos de ADN fueran obtenidos mediante restricción enzimática, se conoce el sitio de reconocimiento de la enzima de restricción, lo que hace posible identificar los fragmentos ya que los restos de sitios de reconocimiento de enzimas de restricción reconstituidos pueden indicar la separación entre diferentes fragmentos de ADN. En el caso de que los fragmentos de ADN fueran obtenidos mediante fragmentación aleatoria, como la sonicación y la subsiguiente reparación enzimática del ADN final, puede ser más difícil distinguir un fragmento de otro. Independientemente de qué método de fragmentación se use, la etapa de ligación c) puede realizarse en presencia de un adaptador, ligando secuencias adaptadoras entre los fragmentos. Alternativamente, el adaptador puede ligarse en una etapa distinta. Esto es ventajoso porque los diferentes fragmentos pueden identificarse fácilmente identificando las secuencias adaptadoras que están entre los fragmentos. Por ejemplo, en el caso de que los extremos del fragmento de ADN fueran extremos romos, la secuencia adaptadora sería adyacente a los extremos del fragmento de ADN, indicando los fragmentos de ADN separados.

A continuación, la reticulación se revierte en la etapa d), lo que da como resultado un conjunto de fragmentos de ADN ligados que comprenden dos o más fragmentos. Una subpoblación del conjunto de fragmentos de ADN ligados comprende un fragmento de ADN que comprende la secuencia de nucleótidos diana. Al revertir la reticulación, se libera la fijación estructural/espacial del ADN y la secuencia de ADN queda disponible para los pasos posteriores, p. ej. amplificación y/o secuenciación, ya que el ADN reticulado puede no ser un sustrato adecuado para tales etapas. Las siguientes etapas e) y/o f) pueden realizarse después de la reversión de la reticulación, si bien las etapas e) y/o f) pueden también realizarse mientras los fragmentos de ADN ligados están todavía en estado reticulado.

Los fragmentos de ADN ligados pueden opcionalmente ser fragmentados en la etapa e), preferiblemente con una enzima de restricción. La fragmentación se puede realizar después de revertir la reticulación, pero también se prevé que la segunda fragmentación se realice antes de que se revierta la reticulación. Se prefiere usar una enzima de restricción para la fragmentación, ya que una enzima de restricción permite el control de la etapa de fragmentación y tiene por resultado, si se elige una enzima de restricción apropiada, extremos compatibles de los fragmentos de ADN ligados que son favorables para la ligación de los extremos compatibles, resultando fragmentos de ADN ligados circularizados, como se obtiene en la etapa f). Sin embargo, la fragmentación usando otros métodos, p. ej. cizallamiento y/o sonicación y subsiguiente reparación enzimática del extremo del ADN, de manera que se forma un ADN de doble hebra de extremos romos, también se puede ligar para formar ADN circularizado.

La primera etapa de fragmentación y la segunda etapa de fragmentación opcional están destinadas a obtener fragmentos de ADN ligados que son compatibles con la etapa siguiente de circularización, etapa de amplificación y/o etapa de determinación de la secuencia. En el caso de que las etapas de fragmentación b) y e) comprendan enzimas de restricción, se prefiere que la etapa de fragmentación e) tenga como resultado, como promedio, fragmentos más largos que los que se obtendrían en la etapa de fragmentación b). En caso de que la etapa de fragmentación b) y e) comprenda enzimas de restricción, se prefiere que el sitio de reconocimiento de la enzima de restricción de la etapa e) sea más largo que el sitio de reconocimiento de la etapa b). La enzima de e) corta así a una frecuencia más baja que la etapa b). Esto significa que el tamaño promedio del fragmento de ADN de la etapa b) es menor que el tamaño de fragmento promedio de la etapa e) obtenido después de restringir el ADN. De esta manera, en la primera etapa de fragmentación se forman fragmentos relativamente pequeños que posteriormente son ligados. Como la segunda enzima de restricción de la etapa e) corta menos frecuentemente que la etapa b), la mayoría de los fragmentos de ADN pueden no comprender el sitio de reconocimiento de restricción de la etapa e). Por tanto, cuando los fragmentos de ADN ligados se fragmentan posteriormente en la segunda etapa, muchos de los fragmentos de ADN de la etapa b) pueden permanecer intactos. Esto es útil porque las secuencias combinadas de los fragmentos de ADN de la etapa b) se pueden usar para construir un cóntigo para la región genómica de interés. Si la fragmentación de la etapa b) es menos frecuente que la fragmentación de la etapa c), el resultado sería que los fragmentos de la etapa b) están fragmentados, lo que puede dar como resultado la pérdida de secuencias de ADN relativamente grandes que son útiles para construir un cóntigo. Por lo tanto, independientemente del método que se usase para fragmentar en la etapa b) y e), se prefiere que la fragmentación de la etapa b) sea más frecuente en

comparación con la etapa e), de manera que los fragmentos de ADN de la etapa b) puedan permanecer en gran parte intactos, es decir, en gran parte no sean fragmentados por la etapa e).

Los fragmentos de ADN ligados obtenidos de la etapa d) o e), cuya reticulación ha sido revertida, se circularizan a continuación en la etapa f). Puede ser ventajoso revertir la reticulación antes de la circularización, porque puede ser desfavorable circularizar el ADN reticulado mientras está reticulado. Sin embargo, la circularización también puede realizarse mientras los fragmentos de ADN ligados están reticulados. Puede ser posible incluso que no se requiera una etapa de circularización adicional, ya que durante la etapa de ligación se forman ya fragmentos de ADN ligados circularizados y, por tanto, la etapa de circularización f) se produciría simultáneamente con la etapa c). Sin embargo, se prefiere realizar una etapa de circularización adicional. La circularización implica la ligación de los extremos de los fragmentos de ADN ligados de manera que se forma un círculo cerrado. El ADN circularizado que comprende fragmentos de ADN ligados que comprenden la secuencia de nucleótidos diana, puede amplificarse posteriormente usando al menos un cebador que se hibrida con la secuencia de nucleótidos diana. Para la etapa de amplificación, se requiere revertir la reticulación, ya que el ADN reticulado puede obstaculizar o impedir la amplificación. Preferiblemente, se usan dos cebadores que se hibridan con la secuencia de nucleótidos diana en una reacción de PCR inversa. De esta manera pueden amplificarse los fragmentos de ADN del ADN circularizado, que se ligan con el fragmento de ADN que comprende la secuencia de nucleótidos diana.

A continuación se determina la secuencia de los fragmentos de ADN ligados (amplificados) obtenidos en las etapas d), e), f) o g) que comprenden la secuencia de nucleótidos diana. La determinación de la secuencia se realiza preferiblemente usando tecnología de secuenciación de alto rendimiento, ya que esto es más cómodo y permite determinar un gran número de secuencias para cubrir la región genómica completa de interés. A partir de estas secuencias determinadas, se puede construir un cóntigo de la región genómica de interés. Cuando se determinan las secuencias de los fragmentos de ADN, se pueden obtener lecturas solapantes a partir de las cuales se puede construir la región genómica de interés. En caso de que los fragmentos de ADN se obtuvieran por fragmentación al azar, la naturaleza aleatoria de la etapa de fragmentación puede dar ya como resultado fragmentos de ADN que, cuando se secuencian, dan lecturas solapantes. Aumentando el tamaño de la muestra, p. ej. aumentando el número de células analizadas, puede aumentarse la fiabilidad de la región genómica de interés que se construye. Alternativamente, cuando en la etapa b) se analiza una pluralidad de submuestras, usando diferentes enzimas de restricción, también se obtendrán lecturas solapantes. Al aumentar la pluralidad de submuestras, aumentará el número de fragmentos solapantes, lo que puede aumentar la fiabilidad del cóntigo de la región genómica de interés que se construye. A partir de estas secuencias determinadas que pueden solaparse, se puede construir un cóntigo. Alternativamente, si las secuencias no se solapan, p. ej. cuando puede haberse usado una única enzima de restricción en la etapa b), la alineación de los fragmentos de ADN (ligados) con una secuencia de referencia puede permitir construir un cóntigo de la región genómica de interés.

Secuencias diana múltiples.

También se describe un método para determinar la secuencia de una región genómica de interés que comprende dos secuencias de nucleótidos diana. Este método puede implicar las mismas etapas esbozadas anteriormente hasta la etapa de amplificación. La etapa de amplificación no usa ahora una secuencia de nucleótidos diana, sino dos. Para las dos secuencias de nucleótidos diana, se usan dos cebadores diferentes en una reacción de PCR, un cebador para cada secuencia de nucleótidos diana. Cuando los dos sitios de unión del cebador de las dos secuencias de nucleótidos diana están presentes en un fragmento de ADN ligado, los dos cebadores amplificarán la secuencia entre los dos sitios de unión del cebador, siempre que los sitios de unión del cebador tengan la orientación correcta. Tener un fragmento de ADN ligado circularizado puede ser ventajoso ya que la probabilidad de que los dos sitios de unión del cebador tengan la orientación correcta es mayor si se compara con un fragmento de ADN ligado lineal (dos de las cuatro orientaciones se amplificarán, en comparación con uno de cuatro para un fragmento de ADN ligado lineal). En una realización adicional, además de las dos secuencias de nucleótidos diana, la región genómica de interés comprende nucleótidos diana adicionales, para cada nucleótido diana se usa un cebador en la reacción de amplificación por PCR. Combinando nucleótidos diana múltiples y los cebadores correspondientes en una sola amplificación, aumentará la posibilidad de que las combinaciones de cebadores produzcan un amplicón.

Por ejemplo, como se describe en la sección de ejemplo 5 diferentes nucleótidos diana se usaron para el gen BRCA1 (véase, p. ej., la figura 2). Se puede realizar una PCR seleccionando un cebador de una secuencia de nucleótidos diana (también denominado punto de vista), p. ej. A con otra B. También se puede realizar una PCR usando un cebador de cada secuencia de nucleótidos diana, A, B, C, D y E. Como estos nucleótidos diana se encuentran en proximidad física entre sí, la realización de tal amplificación enriquecerá la región genómica de interés, siempre y cuando los sitios de unión del cebador terminen en fragmentos de ADN ligados de manera que se pueda generar un amplicón.

Por tanto, se proporcionan métodos para determinar la secuencia de una región genómica de interés de acuerdo con la invención, en donde la región genómica de interés comprende además una o más secuencias de nucleótidos diana, y en donde en la etapa de amplificación se proporciona un cebador que se hibrida con la secuencia de nucleótidos diana y se proporcionan uno o más cebadores para los correspondientes uno o más nucleótidos diana adicionales, en donde los fragmentos de ADN ligados se amplifican, o se amplifica el ADN circularizado, usando los

cebadores.

Determinación de la secuencia de fragmentos de ADN ligados.

La etapa de determinación de la secuencia de fragmentos de ADN ligados comprende preferiblemente una secuenciación de alto rendimiento. Los métodos de secuenciación de alto rendimiento son bien conocidos en la técnica, y en principio se puede considerar cualquier método para que sea usado en la invención. Se pueden realizar tecnologías de secuenciación de alto rendimiento de acuerdo con las instrucciones del fabricante (como por ejemplo, las proporcionadas por Roche, Illumina o Applied Biosystems). En general, los adaptadores de secuenciación se pueden ligar a los fragmentos de ADN ligados (amplificados). En el caso de que el fragmento lineal o circularizado sea amplificado, usando por ejemplo PCR como se describe en este documento, el producto amplificado es lineal, permitiendo la ligación de los adaptadores. Se pueden proporcionar extremos adecuados para ligar secuencias adaptadoras (por ejemplo, extremos escalonados, complementarios romos). Alternativamente, el cebador o cebadores usados para la PCR u otro método de amplificación pueden incluir secuencias adaptadoras, de manera que se forman productos amplificados con secuencias adaptadoras en la etapa de amplificación g). En caso de que el fragmento circularizado no se amplifique, el fragmento circularizado puede fragmentarse, preferiblemente utilizando, por ejemplo, una enzima de restricción entre los sitios de unión del cebador para la reacción de PCR inversa, de manera que los fragmentos de ADN ligados con el fragmento de ADN que comprende la secuencia de nucleótidos diana permanezcan intactos. Los adaptadores de secuenciación también se pueden incluir en los pasos c) y f) de los métodos de la invención.

Estos adaptadores de secuenciación se pueden incluir como parte de las secuencias adaptadoras de los adaptadores que ya se pueden usar opcionalmente en estos pasos y/o se pueden proporcionar además adaptadores de secuencia distintos en estos pasos.

Preferiblemente, se pueden generar lecturas largas en el método de secuenciación de alto rendimiento utilizado. Las lecturas largas pueden permitir leer a través de múltiples fragmentos de ADN de fragmentos de ADN ligados. De esta manera pueden identificarse los fragmentos de ADN de la etapa b). Las secuencias de fragmentos de ADN se pueden comparar con una secuencia de referencia y/o comparar entre sí. Por ejemplo, como también se expone a continuación, tales secuencias de fragmentos de ADN pueden usarse para determinar la proporción de fragmentos de células portadoras de una mutación genética. Secuenciando también secuencias de fragmentos de ADN de fragmentos de ADN adyacentes a tales secuencias, se pueden identificar fragmentos de ADN ligados únicos. Este es particularmente el caso cuando se obtuvieron fragmentos de ADN en la etapa b) mediante fragmentación aleatoria. La probabilidad de que dos células proporcionen exactamente el mismo fragmento de ADN es muy baja, y mucho menos de que el fragmento de ADN al que se liga dicho fragmento sea el mismo. De este modo, identificando los fragmentos de ADN de esta manera, se puede determinar la proporción de células y/o regiones genómicas de interés que comprenden una mutación particular.

Por tanto, no se requiere proporcionar una secuencia completa de los fragmentos de ADN ligados. Se prefiere secuenciar al menos a través de (múltiples) fragmentos de ADN, de forma que se determinen las secuencias de fragmentos de ADN.

También se puede tener en cuenta leer secuencias incluso más cortas, por ejemplo, lecturas cortas de 50 - 100 nucleótidos. En tal escenario, se prefiere fragmentar el ADN ligado (amplificado) en fragmentos más pequeños, que se pueden ligar posteriormente con un adaptador apropiado adecuado para el método de secuenciación de alto rendimiento. En caso de que se use un protocolo de secuenciación estándar, esto puede significar que puede perderse la información relativa a los fragmentos de ADN ligados. Con lecturas cortas, puede que no sea posible identificar una secuencia completa de fragmentos de ADN. En el caso de que se contemplen tales lecturas cortas, se puede prever proporcionar etapas de procesamiento adicionales de forma que fragmentos de ADN ligados distintos cuando están fragmentados, sean ligados o equipados con identificadores, de modo que a partir de las lecturas cortas, se construyan cónfigos para los fragmentos de ADN ligados. Tales tecnologías de secuenciación de alto rendimiento que implican lecturas cortas de secuencia pueden implicar una secuenciación final apareada. Mediante el uso de secuencias finales apareadas y lecturas de secuencia cortas, las lecturas cortas de ambos extremos de una molécula de ADN utilizada para la secuenciación, cuya molécula de ADN puede comprender diferentes fragmentos de ADN, pueden permitir el acoplamiento de fragmentos de ADN que se ligaron. Esto es porque se pueden acoplar dos lecturas de secuencia que abarcan una secuencia de ADN relativamente grande con respecto a la secuencia que se determinó a partir de ambos extremos. De esta forma, se pueden construir cónfigos para los fragmentos de ADN ligados (amplificados).

Sin embargo, puede contemplarse el uso de lecturas cortas sin identificar fragmentos de ADN, ya que a partir de la secuencia corta se lee una región genómica de interés que puede construirse, especialmente cuando la región genómica de interés ha sido amplificada. La información con respecto a los fragmentos de ADN y/o la región genómica de interés separada (por ejemplo, de una célula diploide) puede perderse, pero las mutaciones de ADN aún pueden identificarse.

Por tanto, la etapa de determinación de al menos parte de la secuencia de ADN ligada (amplificada) puede comprender lecturas de secuencia cortas, pero preferiblemente se determinan lecturas de secuencia más largas, de

manera que las secuencias de fragmentos de ADN puedan ser identificadas. Además, también se puede considerar el uso de diferentes estrategias de secuenciación de alto rendimiento para los fragmentos de ADN ligados (amplificados), p. ej. combinando lecturas de secuencia cortas a partir de secuenciación de extremos emparejados con los extremos relativamente alejados con lecturas de secuencia más largas, de esta manera, se pueden construir 5 contiguos para los fragmentos de ADN ligados (amplificados).

En una realización, el método se puede usar para proporcionar control de calidad de la información de secuencia generada. En el análisis de las secuencias proporcionadas por un método de secuenciación de alto rendimiento, pueden presentarse errores de secuenciación. Puede producirse un error de secuenciación, por ejemplo, durante el alargamiento de la cadena de ADN, en el que se incorpora la base equivocada (es decir, no complementaria a la plantilla) en la cadena de ADN. Un error de secuenciación es diferente de una mutación, ya que el ADN original que es amplificado y/o secuenciado no comprendería esa mutación. De acuerdo con la invención, las secuencias de fragmentos de ADN se pueden determinar, con secuencias de fragmentos de ADN (o al menos parte de ellas) ligados a las mismas, las cuales pueden ser únicas. La singularidad de los fragmentos de ADN ligados a medida que se forman en la etapa c) puede proporcionar un control de calidad de la secuencia determinada en la etapa h). 10 Cuando los fragmentos de ADN ligados se amplifican y se secuencian a una profundidad suficiente, se secuenciarán múltiples copias del mismo fragmento de ADN único (ligado). Pueden compararse secuencias de copias que se originan a partir del mismo fragmento de ADN ligado original y pueden identificarse errores de amplificación y/o secuenciación. 15

Otras formas de realización.

Además, de acuerdo con los métodos de la invención, a partir de una muestra de ADN reticulado se determinan las secuencias de regiones genómicas múltiples de interés. Para cada región genómica de interés se proporciona una secuencia de nucleótidos diana, para la que se pueden diseñar los cebadores correspondientes. Las regiones genómicas múltiples de interés pueden ser regiones genómicas de interés que también pueden solaparse, aumentando así el tamaño de la secuencia que puede determinarse. Por ejemplo, en el caso de una secuencia de una región genómica de interés que comprende una secuencia de nucleótidos diana típicamente comprendería 1 MB, combinar regiones genómicas de solapamiento parcial de interés, p. ej. con un solapamiento de 0,1 MB, cada una con una secuencia de nucleótidos diana en correspondencia, la combinación de 5 regiones genómicas de interés daría como resultado una secuencia de 4,6 MB ( $0,9 + 3 * (0,1 + 0,8) + 0,1 + 0,9 = 4,6$  MB), con ello extendiendo en gran medida el tamaño de la región genómica de interés cuya secuencia puede determinarse o analizarse de otra manera. También se pueden usar múltiples secuencias de nucleótidos diana a distancias definidas dentro de una región genómica de interés para aumentar la cobertura promedio y/o la uniformidad de cobertura a través de la región genómica. 20 25 30

Además, se puede incluir un identificador en al menos uno de los cebadores de oligonucleótido de la etapa g). Los identificadores también se pueden incluir en secuencias adaptadoras, tales como las que se usan para la ligación entre fragmentos durante la etapa c) de ligación. Incluyendo un identificador en el cebador oligonucleotídico, cuando se analiza una pluralidad de muestras o una pluralidad de submuestras de ADN reticulado simultáneamente, puede determinarse fácilmente el origen de cada muestra. Las (sub)muestras de ADN reticulado pueden haber sido procesadas de forma diferente mientras que la muestra original de ADN reticulado es la misma, y/o se pueden haber obtenido muestras de ADN, por ejemplo, de diferentes organismos o pacientes. Los identificadores permiten 35 40 combinar muestras procesadas de manera diferente cuando el procesamiento de las muestras puede converger, p. ej. se realizan idénticos pasos de procedimiento. Tal convergencia del procesamiento puede ser especialmente ventajosa cuando la etapa de secuenciación h) implica una secuenciación de alto rendimiento.

Antes o después de la etapa de amplificación g), de acuerdo con los métodos de la invención, se puede realizar una etapa de selección de tamaño. Tal etapa de selección de tamaño puede realizarse usando cromatografía de extracción en gel, electroforesis en gel o centrifugación en gradiente de densidad, que son métodos conocidos generalmente en la técnica. Preferiblemente, se selecciona el ADN de un tamaño entre 20 - 20.000 pares de bases, preferiblemente 50 - 10.000 pares de bases, lo más preferiblemente entre 100 - 3.000 pares de bases. Una etapa de separación por tamaño permite seleccionar fragmentos de ADN ligados (amplificados) en un margen de tamaño que puede ser óptimo para la amplificación por PCR y/o óptimo para la secuenciación de lecturas largas por secuenciación de próxima generación. La secuenciación de lecturas de 500 nucleótidos está disponible comercialmente en la actualidad, los avances recientes de compañías como la tecnología de Secuenciación de ADN de Molécula Única en Tiempo Real (SMRT™: Single Molecule Real Time) desarrollada por Pacific Biosciences <http://www.pacificbiosciences.com/>) indican que están al alcance lecturas de 1.000 a 10.000 nucleótidos. 45 50

En el caso en el que la ploidía en una célula de una región genómica de interés es mayor que 1, para cada ploidía se construye un contiguo en la etapa h) de los métodos de acuerdo con la invención. Como el entorno genómico de cualquier sitio diana dado en el genoma consiste principalmente en secuencias del genoma de ADN que están físicamente cerca de la secuencia diana en la plantilla del cromosoma lineal, permite la reconstrucción de cada plantilla de cromosoma particular. En el caso en que la ploidía de una región genómica de interés es mayor que 1, están presentes múltiples regiones genómicas de interés en una célula (o un equivalente de la misma). Estas regiones genómicas múltiples de interés no ocupan generalmente el mismo espacio, es decir, están separadas en el espacio. Cuando una muestra de ADN reticulado de dicha célula se fragmenta, de cada región genómica de interés 55 60

- en una célula se formará el correspondiente fragmento de ADN que comprende la secuencia de nucleótidos diana. Estos fragmentos de ADN se ligarán con fragmentos de ADN en su proximidad. Los fragmentos de ADN ligados serán así representativos de las diferentes regiones genómicas de interés. Por ejemplo, en el caso en que la ploidía es dos, cuando dos fragmentos, cada uno de los cuales tiene una mutación única, y separados por 1 MB se encuentren juntos en fragmentos de ADN ligados, se puede concluir que estos dos fragmentos pertenecen a la misma región genómica de interés. Por tanto, en este escenario se identificaron dos fragmentos, y ambos son asignados a la misma región genómica. Así pues, al construir un cóntigo a partir de las secuencias de fragmentos identificados, estos dos fragmentos que llevan una mutación se usarían para construir un cóntigo para una región genómica particular, mientras que el cóntigo construido para la otra región genómica no portaría las mutaciones.
- 5
- 10 Así pues, de acuerdo con los métodos de la invención, la etapa h) de construcción de un cóntigo comprende los pasos de:
- 1) identificar los fragmentos de la etapa b);
  - 2) asignar los fragmentos a una región genómica;
  - 3) construir un cóntigo para la región genómica a partir de las secuencias de los fragmentos.
- 15 También, cuando se presentan tres fragmentos que comprenden una mutación única (A\*, B\* y C\*) y la ploidia de interés genómico es dos. Esta vez, se identifican los productos de ligación que comprenden dos de los fragmentos mutados, un producto de ligación que comprende A\*B\* y uno con A\*C\*. También los productos de ligación que comprenden fragmentos no mutados se identifican BC y AC. En este escenario, los fragmentos de ADN ligados A\*B y A\*C\* están acoplados por el fragmento A\*, y los fragmentos de ADN ligados BC y AC están acoplados por el
- 20 fragmento C. En este escenario, los fragmentos de ADN A\*, B\* y C\* son asignado a la misma región genómica, mientras que A, B y C son asignados a la otra región genómica. Así pues, en consecuencia, la etapa 2) de asignar los fragmentos a una región genómica comprende identificar los diferentes productos de ligación y acoplar los diferentes productos de ligación que comprenden los fragmentos de ADN.
- 25 De modo semejante, lo mismo sería válido para las poblaciones de células heterogéneas. Por ejemplo, en caso de proporcionarse una muestra de ADN reticulado que comprende una población de células heterogéneas (por ejemplo, células con diferente origen o células de un organismo que comprende células normales y células mutadas genéticamente (por ejemplo, células cancerosas)), pueden construirse cóntigos para cada región genómica de interés que corresponde a entornos genómicos diferentes (que pueden ser, por ejemplo, diferentes entornos genómicos en una célula o diferentes entornos genómicos de diferentes células).
- 30 Identificación de mutaciones.
- En formas de realización alternativas, se proporcionan métodos para identificar la presencia o ausencia de una mutación genética.
- En una primera realización, se proporciona un método para identificar la presencia o ausencia de una mutación genética, que comprende las etapas a) - h) de cualquiera de los métodos de la invención como se describió
- 35 anteriormente, en donde los cóntigos se construyen para una pluralidad de muestras, que comprende las etapas adicionales de:
- i) alinear los cóntigos de una pluralidad de muestras;
  - j) identificar la presencia o ausencia de una mutación genética en las regiones genómicas de interés de la pluralidad de muestras.
- 40 Alternativamente, se proporciona un método para identificar la presencia o ausencia de una mutación genética, que comprende las etapas a) - g) de cualquiera de los métodos de la descripción como se ha descrito anteriormente, que comprende las etapas adicionales de:
- i) alinear el cóntigo con una secuencia de referencia;
  - j) identificar la presencia o ausencia de una mutación genética en la región genómica de interés.
- 45 Se pueden identificar mutaciones genéticas, por ejemplo, comparando los cóntigos de muestras múltiples, en el caso de que una (o más) de las muestras comprenda una mutación genética, esto puede observarse ya que la secuencia del cóntigo es diferente cuando se compara con la secuencia de las otras muestras, es decir, la presencia de una mutación genética es identificada. En caso de que no se observen diferencias de secuencia entre cóntigos de las muestras, se identifica la ausencia de mutación genética. Alternativamente, también se puede usar una secuencia
- 50 de referencia con la que se puede alinear la secuencia de un cóntigo. Cuando la secuencia del cóntigo de la muestra es diferente de la secuencia de referencia, se observa una mutación genética, es decir, se identifica la presencia de una mutación genética. En caso de que no se observen diferencias de secuencia entre el cóntigo de la muestra o muestras y la secuencia de referencia, se identifica la ausencia de mutación genética.

No es necesario construir un cóntigo para identificar la presencia o la ausencia de una mutación genética. Siempre y cuando las secuencias de fragmentos de ADN puedan alinearse, entre sí o con una secuencia de referencia, se puede identificar la presencia o la ausencia de una mutación genética. Por tanto, en realizaciones alternativas de la invención se proporciona un método para identificar la presencia o la ausencia de una mutación genética, de acuerdo con cualquiera de los métodos descritos anteriormente, sin la etapa h) de construcción de un cóntigo. Tal método comprende las etapas a) - g) de cualquiera de los métodos como se ha descrito anteriormente y los pasos adicionales de:

- 5 h) alinear las secuencias determinadas de los fragmentos de ADN ligados (amplificados) con una secuencia de referencia;
- 10 i) identificar la presencia o ausencia de una mutación genética en las secuencias determinadas.

Alternativamente, se proporciona un método para identificar la presencia o ausencia de una mutación genética, en donde de una pluralidad de muestras se determinan secuencias de fragmentos de ADN ligados (amplificados), que comprenden los pasos a) - g) de cualquiera de los métodos como se describió anteriormente, que comprende las etapas adicionales de:

- 15 h) alinear las secuencias determinadas de los fragmentos de ADN ligados (amplificados) de una pluralidad de muestras;
- i) identificar la presencia o ausencia de una mutación genética en las secuencias determinadas.

Relación de alelos o células que llevan una mutación genética.

Como ya se ha mencionado anteriormente, cuando de poblaciones de células heterogéneas se proporciona una muestra de ADN reticulado (por ejemplo, células con diferente origen o células de un organismo que comprende células normales y células mutadas genéticamente (por ejemplo, células cancerosas)), se pueden construir cóntigos para cada región genómica de interés correspondiente a un entorno genómico diferente (que pueden ser, por ejemplo, entornos genómicos diferentes de diferentes alelos en una célula o diferentes entornos genómicos de diferentes células). Además, se puede determinar la relación de fragmentos o fragmentos de ADN ligados que llevan una mutación genética, que puede correlacionarse con la relación de alelos o células portadoras de la mutación genética. La ligación de fragmentos de ADN es un proceso aleatorio, la colección y el orden de los fragmentos de ADN que son parte de los fragmentos de ADN ligados pueden ser únicos y representan una única célula y/o una única región genómica de interés de una célula. Además, en caso de que la etapa de fragmentación b) comprenda un proceso de fragmentación aleatoria, tal como p. ej. sonicación, los puntos en los que se ha roto el ADN pueden proporcionar una característica única adicional, especialmente dentro del contexto de los otros fragmentos de ADN a los que se liga (que también puede tener extremos de fragmentos únicos).

Así pues, la identificación de fragmentos de ADN ligados que comprenden el fragmento con la mutación genética también puede comprender identificar fragmentos de ADN ligados con un orden único y una colección de fragmentos de ADN. La relación de alelos o células que llevan una mutación genética puede ser de importancia en la evaluación de terapias, p. ej. en el caso en que los pacientes están siendo sometidos a una terapia para el cáncer. Las células cancerosas pueden tener una mutación genética particular. El porcentaje de células que portan dicha mutación puede ser una medida del éxito o el fracaso de una terapia. En realizaciones alternativas, se proporcionan métodos para determinar la relación de fragmentos que llevan una mutación genética, y/o la proporción de fragmentos de ADN ligados que llevan una mutación genética. En esta realización, una mutación genética se define como una mutación genética particular o una selección de mutaciones genéticas particulares.

En una primera realización, se proporciona un método para determinar la proporción de fragmentos que llevan una mutación genética de una población celular sospechosa de ser heteróloga que comprende las etapas a) - h) de cualquiera de los métodos descritos anteriormente, que comprende las siguientes etapas de:

- i) identificar los fragmentos de la etapa b);
- 45 j) identificar la presencia o ausencia de una mutación genética en los fragmentos;
- k) determinar el número de fragmentos que llevan la mutación genética;
- l) determinar el número de fragmentos que no llevan la mutación genética;
- m) calcular la relación de fragmentos que llevan la mutación genética.

En una realización alternativa, se proporciona un método para determinar la relación de productos de ligación que llevan un fragmento con una mutación genética de una población de células sospechosa de ser heteróloga, que comprende los pasos a) - h) de cualquiera de los métodos descritos anteriormente, que comprende las etapas adicionales de:

- i) identificar los fragmentos de la etapa b);

- j) identificar la presencia o ausencia de una mutación genética en los fragmentos;
- k) identificar los productos de ligación de la etapa f) que llevan los fragmentos con o sin la mutación genética;
- l) determinar el número de productos de ligación que llevan los fragmentos con la mutación genética;
- m) determinar el número de productos de ligación que llevan los fragmentos sin la mutación genética;
- 5 n) calcular la relación de productos de ligación que llevan la mutación genética.

En los métodos de estas realizaciones, la presencia o ausencia de una mutación genética puede identificarse en la etapa j) alineándose con una secuencia de referencia y/o comparando secuencias de fragmentos de ADN de una pluralidad de muestras.

10 En los métodos de acuerdo con la descripción, una mutación genética identificada puede ser un SNP o polimorfismo de un nucleótido único, una inserción, una inversión y/o una translocación. En caso de que se observe una delección y/o inserción, el número de fragmentos y/o productos de la ligación de una muestra que lleva la delección y/o inserción puede compararse con una muestra de referencia para identificar la delección y/o la inserción. También se puede identificar una delección, una inserción, una inversión y/o una translocación en función de la presencia de puntos de corte cromosómicos en los fragmentos analizados.

15 En otra realización, en los métodos descritos anteriormente, la presencia o ausencia de nucleótidos metilados se determina en fragmentos de ADN, fragmentos de ADN ligados y/o regiones genómicas de interés. Por ejemplo, el ADN de las etapas a) - f) puede tratarse con bisulfito. El tratamiento de ADN con bisulfito convierte los restos de citosina en uracilo, pero deja restos de 5-metilcitosina no afectados. Por tanto, el tratamiento con bisulfito introduce cambios específicos en la secuencia de ADN que dependen del estatus de metilación de restos individuales de citosina, produciendo información de resolución de un nucleótido único sobre el estado de metilación de un  
20 segmento de ADN. Dividiendo las muestras en submuestras, en las que una de las muestras es tratada y la otra no, se pueden identificar nucleótidos metilados. Alternativamente, las secuencias de una pluralidad de muestras tratadas con bisulfito también pueden alinearse, o una secuencia de una muestra tratada con bisulfito se puede alinear con una secuencia de referencia.

25 Cuando se analizan lecturas de secuencias (cortas), puede ser de interés evitar la secuenciación de los cebadores utilizados. Por tanto, en un método alternativo, la secuencia del cebador puede eliminarse antes de la etapa de secuenciación de alto rendimiento. Así pues, en una realización alternativa se proporciona el siguiente método para determinar la secuencia de una región genómica de interés que comprende una secuencia de nucleótidos diana, que comprende las etapas de:

- 30 a) proporcionar una muestra de ADN reticulado;
- b) fragmentar el ADN reticulado;
- c) ligar el ADN reticulado fragmentado;
- d) revertir la reticulación;
- e) opcionalmente, fragmentar el ADN de la etapa d), preferiblemente con una enzima de restricción;
- 35 f) opcionalmente, ligar el ADN fragmentado de la etapa d) o e) con al menos un adaptador;
- g) amplificar los fragmentos de ADN ligados de la etapa d) o e) que comprenden la secuencia de nucleótidos diana usando al menos un cebador que preferiblemente contiene un colgajo 5' que lleva un sitio de reconocimiento de enzima de restricción de tipo III y (2) se hibrida a la secuencia de nucleótidos diana, o amplificar los fragmentos de ADN ligados de la etapa f) usando al menos un cebador que (1) contiene preferiblemente un colgajo 5' que lleva  
40 un sitio de reconocimiento de enzima de restricción tipo III y (2) se hibrida con la secuencia de nucleótidos diana y al menos un cebador que se hibrida con al menos un adaptador;
- h) digerir las secuencias de nucleótidos amplificadas de interés con una enzima de restricción de tipo III, seguido de una etapa de selección de tamaño para eliminar las secuencias de cebadores de doble cadena liberadas;
- 45 i) fragmentar el ADN, preferiblemente mediante sonicación;
- j) opcionalmente, ligar secuencias adaptadoras bicatenarias necesarias para la secuenciación de próxima generación
- k) determinar al menos parte de la secuencia de los fragmentos de ADN ligados (amplificados) de las etapas d), e), f) o g), que comprende la secuencia de nucleótidos diana preferiblemente usando secuenciación de alto  
50 rendimiento;



l) identificar la variación genética en, y construir un cóntigo de la región genómica de interés a partir de las secuencias determinadas.

5 En una realización alternativa, en cualquiera de los métodos como se describe en este documento, en la etapa g) se usan cebadores que llevan un resto, p. ej. biotina, para la purificación opcional de fragmentos de ADN ligados (amplificados) mediante la unión a un soporte sólido.

10 En una realización, los fragmentos de ADN ligados que comprenden la secuencia de nucleótidos diana pueden capturarse con una sonda de hibridación (o sonda de captura) que se hibrida con una secuencia de nucleótidos diana. La sonda de hibridación se puede unir directamente a un soporte sólido, o puede comprender un resto, p. ej. biotina, para permitir la unión a un soporte sólido adecuado para capturar restos de biotina (por ejemplo, perlas recubiertas con estreptavidina). En cualquier caso, los fragmentos de ADN ligados que comprenden una secuencia de nucleótidos diana son capturados permitiendo así separar los fragmentos de ADN ligados que comprenden la secuencia de nucleótidos diana de los fragmentos de ADN ligados que no comprenden la secuencia de nucleótidos diana. Por tanto, tales etapas de captura permiten enriquecer fragmentos de ADN ligados que comprenden la secuencia de nucleótidos diana. Por ello, a lo largo de toda la invención, se realiza una etapa de amplificación, que también es una etapa de enriquecimiento, alternativamente se puede realizar una etapa de captura con una sonda dirigida a una secuencia de nucleótidos diana. Para una región genómica de interés, se puede usar al menos una sonda de captura para una secuencia de nucleótidos diana. Para una región genómica de interés, puede usarse más de una sonda para múltiples secuencias de nucleótidos diana. Por ejemplo, de modo similar a como se describe para el gen BRCA1, puede usarse un cebador de una de las 5 secuencias de nucleótidos diana como sonda de captura (A, B, C, D o E). Alternativamente, los 5 cebadores se pueden usar de manera combinada (A, B, C, D y E) capturando la región genómica de interés.

15 En una realización, se puede combinar una etapa de amplificación y una etapa de captura, p. ej. realizando primero una etapa de captura y luego una etapa de amplificación o viceversa.

25 En una realización, se puede usar una sonda de captura que se hibrida con una secuencia adaptadora comprendida en fragmentos de ADN ligados (amplificados).

#### **Ejemplo.**

30 Este es un ejemplo de un planteamiento de secuenciación completa de genes de acuerdo con la invención que se usó para determinar la secuencia del gen Brca1 completa. Las células que se usaron fueron células SUM149PT, una línea de células adherentes de cáncer de mama, con una delección de una T en la posición 2288 en el locus Brca1 (Elstrodt y col., Cancer Res, 2006). La figura 1 muestra un esquema del método.

#### *Cultivo de células.*

35 Se cultivan células SUM149PT en platos de 150 cm<sup>2</sup> hasta placas completas con RPMI/FCS al 10% /penstrep. La escisión previa y el recuento de un plato indicaron que un plato completo de 150 cm<sup>2</sup> contiene ~20 x 10<sup>6</sup> células SUM149PT.

#### 35 *Fijación y lisis celular.*

40 Las células cultivadas se lavan con PBS y se fijan con PBS/FCS al 10%/formaldehído al 2% durante 10 minutos a temperatura ambiente. Posteriormente las células se lavan y se recogen, y se absorben en tampón para lisis (Tris-HCl 50 mM, pH 7,5, NaCl 150 mM, EDTA 5 mM, NP-40 al 0,5%, TX-100 al 1% y 1X inhibidores completos de la proteasa (Roche nº 11245200), y se incuban durante 10 minutos en hielo. A continuación las células se lavan y se recogen en MilliQ.

#### *Fragmentación 1: digestión.*

Las células lisadas fijadas se digieren con NlaIII (New England Biolabs # R0125).

#### *Ligación 1.*

45 La enzima NlaIII se inactiva mediante calor y posteriormente se lleva a cabo una etapa de ligación usando ADN ligasa T4 (Roche, nº 799009).

#### *Reversión de la reticulación.*

50 A la muestra se agrega Prot K (10 mg/ml) y se incuba a 65 °C. Posteriormente se agrega RNasa A (10 mg/ml, Roche Nº 10109169001) y la muestra se incuba a 37 °C. Después se realiza la extracción con fenol-cloroformo, y el sobrenadante que comprende el ADN se precipita y se sedimenta. El sedimento se disuelve en Tris-HCl 10 mM, pH 7,5.

*Fragmentación 2: segunda digestión.*

La muestra digerida y ligada se digiere con Nspl (New England Biolabs # R0602S).

*Ligación 2: Segunda ligación y purificación.*

5 A la muestra se agrega Prot K (10 mg/ml) y se incuba a 65 °C. Posteriormente se añade RNasa A (10 mg/ml, Roche N° 10109169001) y la muestra se incuba a 37 °C. Luego se realiza la extracción con fenol-cloroformo, y el sobrenadante que contiene el ADN se precipita y se sedimenta. El sedimento se disuelve en Tris-HCl 10 mM, pH 7,5. La plantilla de enriquecimiento ahora está terminada y puede almacenarse o continuarse con ella directamente.

*Amplificación de fragmentos de ADN ligados: PCR.*

10 Los cebadores utilizados para el enriquecimiento con PCR del locus Brca1 están diseñados como cebadores únicos invertidos cercanos (< 50 pb) a los sitios de restricción NlaIII con un espaciado de los conjuntos de cebadores, es decir, "puntos de vista", de aproximadamente 20 kb (véanse la figura 2 y la tabla 1).

nombre	vp	ID	secuencia	inicio (5')	final (3')
BRCA1_9.9_fw	A	1	CTGGTGGGATCTGTCATTT	6470734	6470752
BRCA1_9.9_rev	A	2	TGGTAGCAAACACTTCCAC	6470481	6470463
BRCA1_28.9_fw	B	3	TATAAGTTTGCCTGCTGCAC	6489743	6489762
BRCA1_28.9_rev	B	4	TTTCCTTAACAATGCACAAA	6489413	6489394
BRCA_50.1_fw	C	5	CATTACTGTAGAAGTTCCCTAAA	6511331	6511353
BRCA_50.1_rev	C	6	ACCATTGCTGTTCTTCTAA	6510682	6510663
BRCA_65.2_fw	D	7	TCCTCCTGAAGAGAACTTG	6526103	6526122
BRCA_65.2_rev	D	8	AGTCCCACCTTGAAGAATC	6525783	6525764
BRCA_91.5_fw	E	9	AGTGAGCGCCGAATTTGC	6552296	6552313
BRCA_91.5_rev	E	10	GCGAAGACCTTTCATTCC	6552022	6552005

15 Tabla 1. Visión de conjunto de las secuencias de cebadores usados. Los cebadores se nombran (nombre) con referencia al gen BRCA1, la posición en el mapa de secuencia (por ejemplo, 50.1 (kb)) y si es un cebador directo (forward: fw) o inverso (reverse: rev). vp indica los puntos de vista, ID indica la SEQ ID No., es decir, las SEQ ID Nos. 1 - 10. La secuencia del gen BRCA1 a la que corresponde el cebador está también indicada (inicio (5') y final (3')), obsérvese que los cebadores están orientados hacia el exterior, es decir, inversos, no pueden formar un amplicón usando ADN normal como plantilla.

Una reacción de enriquecimiento por PCR típica consiste en 25 µl de:

- 20
- 2,5 µl 10X tampón 3 PCR (suministrado con la Expand Long Template Polymerase)
  - 0,5 µl dNTP (10 mM)
  - 0,5 µl de cebador directo (de una dilución 1/7 de un stock de cebador de 1 µg/µl)
  - 0,5 µl de cebador inverso (de una dilución 1/7 de un stock de cebador de 1 µg/µl)
  - 0,375 µl de Expand Long Template Polymerase (Roche # 11759060001)

25

  - 100 ng de plantilla de enriquecimiento
  - X µl de Milli-Q hasta un volumen total de 25 µl

*Secuenciación de los fragmentos de ADN ligados amplificados.*

Proceder con la preparación de la biblioteca para la secuenciación SOLiD, de acuerdo con los protocolos SOLiD estándar.

30

**Resultados.**

La distribución de lecturas desde los diferentes puntos de vista fue la más alta alrededor del sitio del punto de vista. Se muestran otras estadísticas adicionales en la tabla 2. Las lecturas de secuencias desde librerías de puntos de vista C, D y E identificaron la mutación 2288delT. También se determinó qué secuencias del gen BRCA1 no estaban cubiertas, desde el punto de vista A, 15807 pares de bases no estaban cubiertos, desde B, 50124 b. A partir de C, D y E, todas las secuencias BRCA1 fueron cubiertas.

vp	M	TR	% MtT	media	mediana	% nt 20x
A	898515	13715420	6,55	531	3	11,5
B	17578	7401964	0,24	10	0	4,90
C	2098974	11190246	18,76	1241	29	63
D	3113059	9851741	31,60	1840	45	74,70
E	134324	9108300	1,47	79	42	71,70

Tabla 2. La estadística de lectura de secuencias por punto de vista. Se indican las estadísticas de lectura de secuencia por biblioteca de puntos de vistas (vp, A - E). M (lecturas coincidentes con BRCA1), TR (número total de lecturas), % MtT (% de lecturas totales coincidentes con la diana), media (cobertura media), mediana (cobertura mediana), % nt 20x (el % de los nucleótidos del gen BRCA1 que han sido cobertura más de 20 veces).

Así pues, desde los puntos de vista únicos C, D y E, se cubrió el BRCA1 completo de 100 kb, desde el punto de vista A fueron cubiertos 85 kb del gen BRCA1, desde el punto de vista B fueron cubiertos 50 kb del gen BRCA1 y desde los puntos de vista C, D y E se confirmó la mutación 2288delT.

**15 Lista de secuencias**

- <110> Laat, Wouter L
- <120> Selección y secuenciación dirigidas
- <130> P30349PC00
- <160> 10
- <170> PatentIn versión 3.5
- <210> 1
- <211> 19
- <212> ADN
- <213> Homo sapiens
- <400> 1  
ctggtgggat ctgtcattt      19
- <210> 2
- <211> 19
- <212> ADN
- <213> Homo sapiens
- <400> 2  
tggtagcaaa cactccac      19
- <210> 3
- <211> 20
- <212> ADN
- <213> Homo sapiens
- <400> 3  
tataagtttg cctgctgcac      20
- <210> 4

ES 2 667 346 T3

<211> 20  
 <212> ADN  
 <213> Homo sapiens

5 <400> 4  
 ttccttaac aatgcacaaa 20

<210> 5  
 <211> 23  
 10 <212> ADN  
 <213> Homo sapiens

<400> 5  
 15 cattactgta gaagttcct aaa 23

<210> 6  
 <211> 20  
 <212> ADN  
 <213> Homo sapiens

20 <400> 6  
 accattgctg tccttctaa 20

<210> 7  
 25 <211> 20  
 <212> ADN  
 <213> Homo sapiens

<400> 7  
 30 tcctcctgaa gagaaactg 20

<210> 8  
 <211> 20  
 <212> ADN  
 35 <213> Homo sapiens

<400> 8  
 agtcccacc ttgaagaatc 20

<210> 9  
 <211> 18  
 <212> ADN  
 <213> Homo sapiens

45 <400> 9  
 atgagcgcc gaattgc 18

<210> 10  
 <211> 18  
 <212> ADN  
 <213> Homo sapiens

50 <400> 10  
 gcgaagacct ttattcc 18

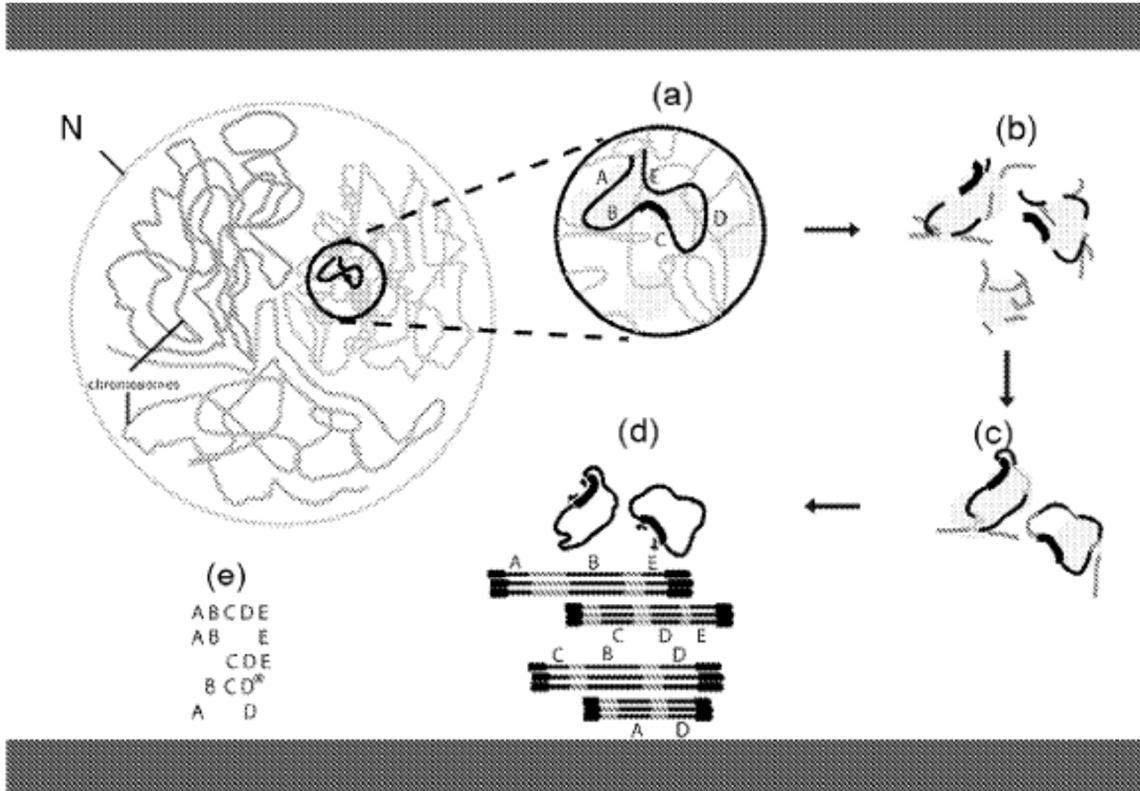
55

**REIVINDICACIONES**

- 1ª. Método para construir un cóntigo de una región genómica de interés que comprende una secuencia de nucleótidos diana, que comprende fragmentar un ADN reticulado,
- 5 ligar el ADN reticulado fragmentado, revertir la reticulación y determinar al menos parte de la secuencias de los fragmentos de ADN ligados, que comprenden un fragmento de ADN con la secuencia de nucleótidos diana, y usar las secuencias determinadas para construir un cóntigo de la región genómica de interés.
- 10 2ª. Método para construir un cóntigo de una región genómica de interés que comprende una secuencia de nucleótidos diana, que comprende las etapas de:
- a) proporcionar una muestra de ADN reticulado;
- b) fragmentar el ADN reticulado;
- c) ligar el ADN reticulado fragmentado;
- 15 d) revertir la reticulación;
- e) opcionalmente, fragmentar el ADN de la etapa d) preferiblemente con una enzima de restricción;
- f) opcionalmente, ligar el ADN fragmentado de la etapa d) o e) con al menos un adaptador;
- g) opcionalmente y preferiblemente amplificar los fragmentos de ADN ligados de la etapa d) o c) que comprenden un fragmento de ADN con la secuencia de nucleótidos diana utilizando al menos un cebador que se hibrida con la secuencia de nucleótidos diana, o amplificar los fragmentos de ADN ligados de la etapa f) usando al menos un cebador que se hibrida con la secuencia de nucleótidos diana y al menos un cebador que se hibrida con al menos un adaptador;
- 20 h) determinar al menos parte de la secuencia de los fragmentos de ADN ligados (amplificados) de las etapas d), e), f) o g), que comprende la secuencia de nucleótidos diana;
- 25 i) construir un cóntigo de la región genómica de interés a partir de las secuencias determinadas.
- 3ª. Método para construir un cóntigo de una región genómica de interés, que comprende una secuencia de nucleótidos diana, comprendiendo el método los pasos de:
- a) proporcionar una muestra de ADN reticulado;
- b) fragmentar el ADN reticulado;
- 30 c) ligar el ADN reticulado fragmentado;
- d) revertir la reticulación;
- e) opcionalmente fragmentar el ADN de la etapa d), preferiblemente con una enzima de restricción;
- f) circularizar el ADN de la etapa d) o e);
- g) opcionalmente y preferiblemente, amplificar el ADN circularizado que comprende un fragmento de ADN con la secuencia de nucleótidos diana usando preferiblemente al menos un cebador que se hibrida con la secuencia de nucleótidos diana;
- 35 h) determinar al menos parte de la secuencia de los fragmentos de ADN ligados (amplificados), que comprende un fragmento de ADN usando el nucleótido diana una secuenciación de alto rendimiento;
- i) construir un cóntigo de la región genómica de interés a partir de las secuencias determinadas.
- 40 4ª. Método para construir un cóntigo de una región genómica de interés según las reivindicaciones 2ª o 3ª, en donde la región genómica de interés comprende una o más secuencias de nucleótidos diana, y en donde en la etapa de amplificación (g) se proporciona un cebador que se hibrida con la secuencia de nucleótidos diana y se proporcionan uno o más cebadores para los correspondientes uno o más nucleótidos diana adicionales, en donde los fragmentos de ADN ligados se amplifican, o se amplifica el ADN circularizado, utilizando los cebadores.

- 5<sup>a</sup>. Método según la reivindicación 2<sup>a</sup> a 4<sup>a</sup>, en el que la etapa de fragmentación b) comprende sonicación, seguida de la reparación enzimática del extremo del ADN.
- 6<sup>a</sup>. Método según una cualquiera de las reivindicaciones 2<sup>a</sup> a 4<sup>a</sup>, en el que la etapa de fragmentación b) comprende la fragmentación con una enzima de restricción.
- 5 7<sup>a</sup>. Método según la reivindicación 5<sup>a</sup> o 6<sup>a</sup>, en el que la etapa de ligación c) se realiza en presencia de un adaptador, ligando secuencias adaptadoras entre fragmentos.
- 8<sup>a</sup>. Método según la reivindicación 6<sup>a</sup> o 7<sup>a</sup>, en el que en la etapa b) se procesa una pluralidad de submuestras, y para cada submuestra se usan enzimas de restricción con sitios de reconocimiento diferentes.
- 10 9<sup>a</sup>. Método según la reivindicación 8<sup>a</sup>, en el que la etapa de fragmentación e) comprende una enzima de restricción con una secuencia de reconocimiento que es más larga que la secuencia de reconocimiento de la enzima de restricción de la etapa b).
- 10<sup>a</sup>. Método según una cualquiera de las reivindicaciones anteriores, en el que se determinan las secuencias de múltiples regiones genómicas de interés.
- 11<sup>a</sup>. Método según una cualquiera de las reivindicaciones 2<sup>a</sup> a 10<sup>a</sup>, en el que antes o después de la etapa de amplificación g) se realiza una etapa de selección por tamaños, en donde preferiblemente
- 15 - la etapa de selección por tamaños se realiza utilizando cromatografía de extracción en gel, electroforesis en gel o centrifugación en gradiente de densidad; y/o
- se elige ADN de un tamaño entre 20 - 20.000 pares de bases, preferiblemente 50 - 10.000 pares de bases, lo más preferiblemente entre 100 y 3.000 pares de bases.
- 20 12<sup>a</sup>. Método según una cualquiera de las reivindicaciones precedentes, en el que fragmentos de ADN ligados (amplificados) que comprenden la secuencia de nucleótidos diana son capturados con una sonda de captura para separar fragmentos de ADN ligados (amplificados) que comprenden la secuencia de nucleótidos diana de fragmentos de ADN ligados (amplificados) que no comprenden la secuencia de nucleótidos diana.
- 25 13<sup>a</sup>. Método según la reivindicación 12<sup>a</sup>, en el que se usan múltiples sondas de captura para secuencias de nucleótidos diana múltiples.
- 14<sup>a</sup>. Método según una cualquiera de las reivindicaciones 2<sup>a</sup> a 13<sup>a</sup>, en el que en el caso en que la ploidía en una célula de una región genómica de interés es mayor que 1, para cada ploidía se construye un cóntigo en la etapa i).
- 15<sup>a</sup>. Método según una cualquiera de las reivindicaciones 2<sup>a</sup> a 13<sup>a</sup>, en el que la etapa i) de construir un cóntigo comprende las etapas de:
- 30 1) identificar los fragmentos de ADN de la etapa b);
- 2) asignar los fragmentos de ADN a una región genómica; y
- 3) construir un cóntigo para la región genómica.
- 16<sup>a</sup>. Método según la reivindicación 15<sup>a</sup>, en el que la etapa 2) de asignar los fragmentos a una región genómica comprende identificar los distintos productos de ligación de la etapa f) y acoplar los diferentes productos de ligación a los fragmentos identificados.
- 35 17<sup>a</sup>. Método según una cualquiera de las reivindicaciones 2<sup>a</sup> a 16<sup>a</sup>, en el que la etapa de construir un cóntigo comprende alinear las secuencias determinadas de los fragmentos de ADN de la etapa b), con una secuencia de referencia.

Figura 1



**Figura 2**

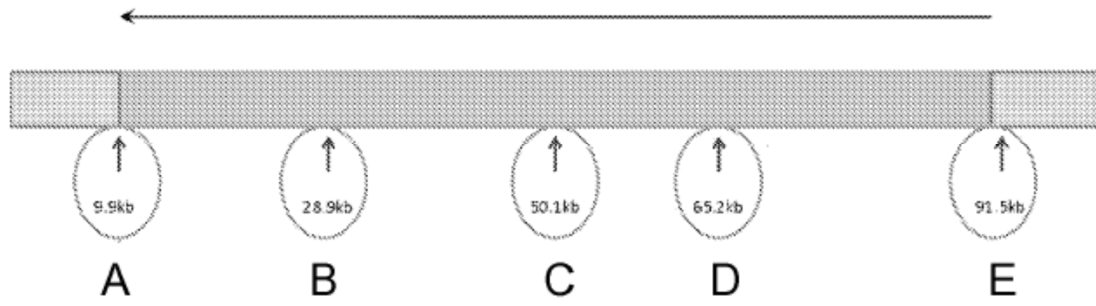




Figura 3

