

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 730 962**

51 Int. Cl.:

**C12Q 1/6869** (2008.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.02.2015 PCT/EP2015/052436**

87 Fecha y número de publicación internacional: **13.08.2015 WO15118077**

96 Fecha de presentación y número de la solicitud europea: **05.02.2015 E 15706700 (0)**

97 Fecha y número de publicación de la concesión europea: **03.04.2019 EP 3102702**

54 Título: **Secuenciación libre de errores de ADN**

30 Prioridad:

**05.02.2014 EP 14154044**

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

**13.11.2019**

73 Titular/es:

**FRAUNHOFER-GESELLSCHAFT ZUR  
FÖRDERUNG DER ANGEWANDTEN  
FORSCHUNG E.V. (100.0%)  
Hansastraße 27c  
80686 München, DE**

72 Inventor/es:

**KLEIN, CHRISTOPH;  
KIRSCH, STEFAN;  
CZYZ, ZBIGNIEW TADEUSZ y  
LAHRMANN, URS**

74 Agente/Representante:

**ARIAS SANZ, Juan**

ES 2 730 962 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

## Secuenciación libre de errores de ADN

5 La invención se refiere a un método novedoso de secuenciación libre de errores de ADN. Además, la presente invención proporciona un oligonucleótido de cuatro partes, que comprende una secuencia fija, una secuencia aleatorizada, un sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción, y un sitio de unión a cebador. La invención también se refiere al uso de los fragmentos de ADN secuenciado obtenidos mediante los métodos de la invención en métodos para el análisis de secuencias de ADN, generación de árboles de linaje celular o evaluación de números de copias.

10 Aplicaciones clínicas tales como diagnóstico previo a la implementación molecular, detección temprana de cáncer y monitorización longitudinal basada en biomarcadores de ácido nucleico de respuesta a la terapia requieren una preparación de muestras precisa y métodos de amplificación de genoma completo para proporcionar cantidades suficientes de ADN de alta calidad para análisis moleculares. Un método para la amplificación de ADN particularmente útil para la amplificación del ADN o del genoma completo de una célula individual se describe en el documento WO 00/17390. Sin embargo, la amplificación de genoma completo así como la preparación de muestras son propensas a la introducción de errores en la secuencia de ADN. Por tanto, la evaluación de las secuencias correctas a partir de células individuales se ve dificultada por un alto nivel de fondo de errores de secuenciación. Por tanto, tiene que establecerse un procedimiento para monitorizar y evaluar de manera apropiada errores de secuencia introducidos de manera metodológica. El requisito más importante para lograr este objetivo es recuperar correctamente las secuencias de ADN a partir de células individuales tal como eran antes de la manipulación experimental. Los enfoques actuales usan la información de la cadena de ADN complementaria de cada secuencia de ADN para corregir todos los cambios que no están presentes en ambas cadenas de una molécula de ADN bicatenario original.

15 Se han usado diversas categorías fundamentalmente distintas de enfoques con sensibilidades significativamente diferentes para separar ruido derivado de error de secuencia de variación genética real en enfoques de secuenciación de nueva generación. La primera categoría de enfoques consiste exclusivamente en análisis bioinformáticos y representa un análisis posterior clave para conjuntos de datos de secuenciación de nueva generación convencionales; véase por ejemplo DePristo *et al.*, (2011) Nature Genetics 43(5):491-8 y referencias citadas en el mismo. La mayoría de los lectores de secuencias variantes se construyen basándose en algoritmos bayesianos e incorporan la probabilidad de detección para una variante específica en una posición particular dada la tasa de polimorfismos y errores de secuenciación conocidos. Para la secuenciación de exoma y genoma completo, se desarrollaron procedimientos posteriores a la alineación más sofisticados para aumentar adicionalmente la precisión de la lectura de variantes. Nuevos métodos para realineación local alrededor de indels, recalibración de puntuaciones de calidad de bases y modelado de error adaptativo antes de la lectura de variantes permiten identificar variantes falsas positivas. Adaptar este flujo de análisis de secuencias y evaluar su rendimiento sobre datos de secuenciación derivados de un genoma de célula individual supone actualmente el mayor desafío en la genómica de células individuales. La segunda categoría comprende enfoques no computacionales basados en el uso de secuencias de etiquetas aleatorias acopladas directamente a los fragmentos genómicos que van a secuenciarse. Estos desarrollos tecnológicos se centraron principalmente en la detección de variantes poco frecuentes en poblaciones celulares heterogéneas tal como se encuentran con frecuencia, por ejemplo, en tejidos tumorales mediante secuenciación profunda. La detección de mutaciones poco frecuentes mediante secuenciación de nueva generación se ha descrito por Schmitt *et al.* (2012) PNAS 109(36):14508-13 y en el documento WO2013/142389. La precisión de secuenciación se logra mediante la adición de adaptadores de ADN bicatenario complementario aleatorio a ambas cadenas de una molécula genómica bicatenaria antes de la amplificación. Todas las lecturas de secuenciación que comparten la misma etiqueta de secuenciación pueden fusionarse para dar una secuencia consenso monocatenaria. Además, todas las lecturas de secuenciación derivadas de la cadena complementaria se identifican mediante la secuencia de etiqueta complementaria permitiendo la creación de una secuencia consenso bicatenaria denominada secuencia dúplex. Una variación genética real se caracteriza por una coincidencia perfecta en el mismo nucleótido en el consenso monocatenario opuesto.

20 Un enfoque adicional denominado Safe-SeqS se describió por Kinde *et al.* (2011) PNAS 108(23):9530-5. El enfoque consiste en dos etapas básicas. En primer lugar, la asignación de un identificador único a cada molde de ADN y, en segundo lugar, la amplificación de cada molde marcado con etiqueta de manera única. A diferencia de Schmitt *et al.*, Safe-SeqS no usa la información de la cadena complementaria.

25 Se han publicado enfoques adicionales que usan adaptadores de código de barras (o bien ligados mediante clonación de T/A o bien con ligación de extremos romos) para identificar el origen de muestra. Sin embargo, ninguno de los enfoques comprende un enfoque de corrección de errores. El primer enfoque usa adaptadores de ADN bicatenario que se ligan a extremos romos de fragmentos de ADN bicatenario tal como se describe en el documento WO 2012/042374. Como enfoque adicional descrito a continuación, estas técnicas que usan moléculas de adaptador bicatenario no pueden amplificar y secuenciar cantidades bajas de moléculas de ácido nucleico, en particular ADN, en particular el ADN de una célula individual o una molécula de ADN individual.

De manera similar, enfoques conocidos adicionales comprenden el uso de adaptadores de ADN parcialmente bicatenario, denominados adaptadores en Y. Tales técnicas requieren aplicación de cola de dA antes de la ligación de adaptador con el fin de minimizar la dimerización de adaptador. Además, puede usarse ADN bicatenario generado por enzimas de restricción para la ligación de adaptador en Y, tal como por ejemplo en Monson-Miller *et al.* (2012) BMC Genomics 13:72. Este método implica el uso de secuencias de código de barras cortas (4 bases) para la identificación de muestras. Sin embargo, estas técnicas no pueden basarse en la información de secuencia redundante contenida en las cadenas de ADN complementarias para identificación de errores de secuenciación y/o análisis de mutaciones.

Un método iterativo y regenerativo para secuenciación de ADN se proporciona por Jones (1997) BioTechniques 22:938-946 y en el documento US 08/742.755. El método implica la digestión con enzimas de restricción usando una enzima de restricción de tipo II s, en particular FokI o BseRI. El sitio de reconocimiento está comprendido en la secuencia de adaptador. Tras cada ciclo de PCR, se escinde el adaptador del ADN diana, dando como resultado una pérdida del sitio de unión a cebador para PCR. Se realiza un ciclo de PCR para secuenciar una base, es decir, una amplificación de muestra significativa a partir de una baja cantidad de ADN sólo es posible con ciclos de ligación-amplificación iterativos. Por consiguiente, el método no es adecuado para bajas cantidades de ADN.

Ninguno de los enfoques de la técnica anterior permite la identificación de secuencias de ácido nucleico en muestras que comprenden bajas cantidades de ADN o en muestras que comprenden bajas cantidades de moléculas de ácido nucleico, en particular la(s) molécula(s) de ácido nucleico/ADN de una célula individual. De hecho, todos los enfoques conocidos tal como se mencionó anteriormente en el presente documento se limitan a grandes cantidades de ADN. Esto se debe, entre otras cosas, al hecho de que la secuenciación de dúplex tal como se describe, por ejemplo, en el documento WO2013/142389, usa adaptadores para la ligación, que se añaden en grandes cantidades a ADN celular. El exceso de moléculas de adaptador bicatenario genera inhibidores de PCR. Por tanto, no pueden amplificarse satisfactoriamente bajas cantidades de ADN usando el enfoque descrito. Otras técnicas requieren información de múltiples muestras y/o un enfoque de bioinformática de base estocástica para la identificación de mutaciones y/o errores de secuenciación. Por tanto, existe una necesidad de un método de secuenciación que proporcione información de secuencia libre de errores también a partir de bajas cantidades de moléculas de ácido nucleico/ADN de entrada, en particular el ADN de una célula individual.

Por tanto, el problema técnico subyacente a la presente invención es proporcionar un método de secuenciación libre de errores mejorado de moléculas de ácido nucleico.

La solución a este problema técnico se proporciona mediante las realizaciones tal como se definen a continuación en el presente documento y tal como se caracterizan en las reivindicaciones.

En particular, la presente invención se refiere a lo siguiente:

1. Un método de secuenciación libre de errores de ADN, que comprende las etapas de:

(a) en una muestra que comprende ADN;

(b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar,

en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5', en el que el nucleótido terminal de la proyección está fosforilado o,

en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 3', en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN;

(c) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada;

(d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN, mediante lo cual el primer oligonucleótido no se liga al ADN;

(e) rellenar las proyecciones generadas;

(f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido; y

- (g) secuenciar dichos fragmentos de ADN amplificados.
2. El método según el punto 1, en el que dicho segundo oligonucleótido comprende además una cuarta secuencia que comprende un sitio de restricción de una endonucleasa específica de sitio.
- 5 3. El método según uno cualquiera de los puntos 1 ó 2, en el que dicho segundo oligonucleótido es un oligonucleótido de ADN, un oligonucleótido de ARN o un oligonucleótido de ADN/ARN.
- 10 4. El método según uno cualquiera de los puntos 1 a 3, en el que dicho método comprende además la etapa (e'), en el que se añade una exonucleasa en dicha etapa (e').
- 5 15 5. El método según el punto 4, en el que dicha exonucleasa es una enzima que degrada moléculas de ADN, ARN y/o ADN/ARN monocatenarios.
- 15 6. El método según uno cualquiera de los puntos 1 a 5, en el que dicho ADN comprende (i) el genoma o transcriptoma de una célula individual, (ii) cromosoma(s) de una célula individual, (iii) ácidos nucleicos de exosomas u otras microvesículas de una célula individual o (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii).
- 20 7. El método según uno cualquiera de los puntos 1 a 5, en el que dicho ADN comprende (i) el ADN de más de una célula individual, (ii) ADN fetal libre de células de más de una célula individual, (iii) ADN libre de células de más de una célula individual, en suero y/o plasma de pacientes con cáncer o (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii).
- 25 8. El método según uno cualquiera de los puntos 1 a 7, en el que dicha endonucleasa de restricción usada en la etapa (b) es MseI o un isoesquizómero de la misma.
- 30 9. El método según uno cualquiera de los puntos 1 a 8, en el que dicha secuencia aleatorizada comprende de 3 a 24 nucleótidos.
- 35 10. El método según uno cualquiera de los puntos 1 a 9, en el que dicho primer oligonucleótido comprende la secuencia 5'-TAACTGACdd-3' y/o en el que dicho segundo oligonucleótido comprende la secuencia tal como se muestra en SEQ ID NO: 1 y/o en el que dicho tercer oligonucleótido comprende la secuencia tal como se muestra en SEQ ID NO: 2.
- 40 11. El método según uno cualquiera de los puntos 1 a 10, en el que el último nucleótido en 3' del primer oligonucleótido es un dd-nucleótido.
- 45 12. El método según uno cualquiera de los puntos 2 a 11, en el que dicho método comprende además la etapa (f') en el que se añade una endonucleasa de guiado en dicha etapa (f').
- 50 13. Uso de los fragmentos de ADN secuenciado obtenidos mediante el método según uno cualquiera de los puntos 1 a 12 en métodos para el análisis de secuencias de ADN, generación de árboles de linaje celular o evaluación de números de copias.
- 55 14. El uso según el punto 13, en el que el método para el análisis de secuencias de ADN es secuenciación de genoma completo, secuenciación de exoma completo, secuenciación de reguloma completo, análisis de metilación basado en secuenciación, detección de punto de rotura basado en secuenciación, secuenciación de CHIP o secuenciación dirigida y variaciones de los mismos.
- 60 15. Un oligonucleótido de cuatro partes que comprende una secuencia fija, secuencia aleatorizada, sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción, y sitio de unión a cebador, en el que dicha secuencia aleatorizada comprende de 3 a 24 nucleótidos y dicho sitio de reconocimiento de nucleasa de restricción es un sitio de reconocimiento de una endonucleasa de guiado, en el que dicha secuencia fija comprende GTCAGT, en el que dicho sitio de reconocimiento de nucleasa de restricción comprende SEQ ID NO: 3 y en el que dicho sitio de unión a cebador comprende SEQ ID NO: 4.
- 65 16. El oligonucleótido de cuatro partes según el punto 15 que comprende SEQ ID NO: 14, 5 ó 12.
- Por consiguiente, la presente invención proporciona un método de secuenciación libre de errores de ADN. Tal como resulta evidente a partir de los ejemplos adjuntos, el método de secuenciación libre de errores tal como se proporciona en el presente documento comprende en particular análisis de ADN libre de errores que comprende antes del análisis de ADN libre de errores una etapa de secuenciación de ADN. Se prefieren bajas cantidades de ADN como material de partida que va a emplearse en los métodos de la presente invención. En particular, la presente invención proporciona métodos novedosos e inventivos de secuenciación libre de errores de ADN de una célula individual o una molécula de ADN individual. El método de secuenciación libre de errores de ADN, en

particular bajas cantidades de ADN, proporcionado en el presente documento comprende las etapas de (a) proporcionar una muestra que comprende ADN; (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5', en el que el nucleótido terminal de la proyección está fosforilado, o, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 3', en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN; (c) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada; (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN, mediante lo cual el primer oligonucleótido no se liga al ADN; (e) rellenar las proyecciones generadas; (f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido; y (g) secuenciar dichos fragmentos de ADN amplificados.

Según la invención, el segundo oligonucleótido puede comprender además una cuarta secuencia, en el que la cuarta secuencia comprende un sitio de restricción de una endonucleasa, preferiblemente una endonucleasa de guiado.

Los inventores han encontrado sorprendentemente que el uso de oligonucleótidos monocatenarios en lugar de adaptadores bicatenarios evita la creación de inhibidores de PCR y/o adaptadores que se ligan a sí mismos, es decir moléculas de adaptador que se unen a otras moléculas de adaptador y se someten a ligación. Para lograr una alta eficiencia de ligación de adaptadores a bajas cantidades de ADN objetivo se requiere la adición de moléculas de adaptador en exceso de la cantidad de ADN objetivo. Los adaptadores bicatenarios impedirán entonces la unión del cebador de PCR al ADN diana, porque la concentración de la secuencia complementaria es excesivamente alta. Este exceso de moléculas de adaptador hace imposible usar pequeñas cantidades de ADN objetivo, que no puede purificarse de la mezcla de adaptador-ADN objetivo sin pérdida significativa de ADN objetivo. En cambio, la presente invención proporciona métodos que son independientes de la cantidad de ADN usado as material de partida. En particular, los métodos de la presente invención permiten la secuenciación libre de errores de cantidades particularmente bajas de ADN objetivo, que pueden ser de tan sólo el ADN de una célula individual y/o una molécula de ADN individual.

Además, los inventores han encontrado sorprendentemente que el uso de tres oligonucleótidos permite la secuenciación libre de errores independiente de la cantidad de ADN usado como material de partida. Más específicamente, métodos de la técnica anterior no pueden proporcionar secuenciación libre de errores, por ejemplo, de bajas cantidades de ADN. En cambio, los métodos proporcionados en el presente documento están diseñados para amplificar y obtener la secuencia de cualquier cantidad de ADN, en particular ADN de una célula individual y/o ADN de una molécula de ADN individual. Los métodos proporcionados en el presente documento también pueden usarse para eliminar errores de secuenciación en el ADN amplificado y secuenciado con el fin de identificar mutaciones reales y secuencias correctas. Esto se logra mediante el uso de tres oligonucleótidos. En particular, un primer oligonucleótido es parcialmente complementario a una proyección en 3' y/o en 5' generada de una muestra de ADN que se ha fragmentado mediante una endonucleasa de restricción. Cuando se une a la(s) proyección/proyecciones de los fragmentos de ADN, el primer oligonucleótido también genera una proyección, es decir el primer oligonucleótido es más largo que la proyección en 3' y/o en 5' del fragmento de ADN. Un segundo oligonucleótido que comprende de tres a cuatro partes/secuencias funcionales comprende una primera secuencia, también denominada en el presente documento secuencia fija, que es parcialmente complementaria al primer oligonucleótido, que permite al segundo oligonucleótido unirse a la proyección generada por el primer oligonucleótido, formando así un complejo de ADN-oligonucleótido-oligonucleótido. Por consiguiente, el primer oligonucleótido dirige el segundo oligonucleótido al ADN diana. El uso del primer oligonucleótido aumenta sorprendentemente la eficiencia de ligación del segundo oligonucleótido a cada fragmento de ADN diana. La secuencia fija puede variarse, pero comprende necesariamente una secuencia complementaria al primer oligonucleótido. Tras la unión del segundo oligonucleótido, se realiza una ligación monocatenaria que une de manera covalente el segundo oligonucleótido al fragmento de ADN. El primer oligonucleótido no se liga al ADN. Esto puede lograrse sintetizando el oligonucleótido sin un extremo 5'-fosfato-terminal.

Además, una segunda secuencia del segundo oligonucleótido comprende una secuencia aleatorizada usada como código de barras/identificador para marcar de manera única cada complejo de oligonucleótido-ADN. La longitud del código de barras puede variar. En particular, la longitud del código de barras puede variar dependiendo del número de fragmentos de ADN generados. En una etapa adicional, la(s) proyección/proyecciones generada(s) por el segundo oligonucleótido ligado se rellena(n) mediante una reacción de polimerasa. El segundo oligonucleótido comprende además una tercera secuencia que está diseñada para permitir que se una un tercer oligonucleótido. Por tanto, el tercer oligonucleótido es complementario a la tercera secuencia del segundo oligonucleótido. El tercer oligonucleótido está diseñado para permitir una amplificación eficiente basada en PCR de toda la representación de muestra, cuya secuencia se obtiene entonces. Por consiguiente, el tercer oligonucleótido sirve como cebador para amplificación mediante PCR.

El término “secuencia” se refiere a información de secuencia sobre una molécula de ácido nucleico o cualquier porción de la molécula de ácido nucleico que tiene dos o más unidades (nucleótidos) de longitud. El término también puede usarse como referencia a la propia molécula de ácido nucleico o una porción relevante de la misma.

5 Información de secuencia de molécula de ácido nucleico se refiere a la sucesión de bases de nucleótido en la molécula de ácido nucleico. Por ejemplo, si la molécula de ácido nucleico contiene bases adenina, guanina, citosina, timina o uracilo, o análogos químicos de las mismas, la secuencia de molécula de ácido nucleico puede representarse mediante una sucesión correspondiente de letras A, G, C, T o U, por ejemplo, una molécula de ADN o ARN.

10 Métodos a modo de ejemplo, no limitativos, que van a usarse con el fin de determinar la secuencia de una molécula de ácido nucleico son, por ejemplo, métodos para la secuenciación de ácidos nucleicos (por ejemplo secuenciación didesoxi de Sanger), métodos de secuenciación en paralelo masivos tales como pirosecuenciación, terminador de colorante inverso, detección de protones, nucleótidos fluorescentes unidos a fósforo.

15 En particular, los productos de PCR resultantes pueden someterse o bien a métodos de secuenciación de dideoxi-nucleótidos basados en Sanger convencionales o bien empleando métodos de secuenciación en paralelo masivo novedosos (“secuenciación de nueva generación”) tales como los comercializados por Roche (tecnología 454), Illumina (tecnología Solexa), ABI (tecnología Solid) o Pacific Biosciences (tecnología SMRT). Pueden identificarse mutaciones a partir de lecturas de secuencia mediante comparación con bases de datos de secuencias públicamente disponibles o mediante algoritmos de predicción de pérdida de función y/o ganancia de función implementados en herramientas bioinformáticas *in silico* tales como SIFT y PolyPhen. En particular, las mutaciones pueden ser mutaciones que se conoce en la técnica que son relevantes/decisivas para indicaciones médicas. Alternativamente, pueden identificarse mutaciones mediante incorporación específica de alelo de etiquetas moleculares que pueden detectarse usando reacciones de detección enzimática, fluorescencia, espectrometría de masas u otros; véase Vogeser (2007) Dtsch Arztebl 104 (31-32), A2194-200.

20 El término secuenciación “libre de errores” se refiere a un enfoque que permite eliminar hasta un alto grado los errores técnicos introducidos durante el procesamiento de muestras, por ejemplo aislamiento, amplificación y/o secuenciación de ADN. Mediante el uso de códigos de barras/identificadores aleatorizados cada alelo diferenciado se marca en ambos extremos con una etiqueta de secuencia única. Las secuencias flanqueadas por dos códigos de barras ligados pueden rastrearse fácilmente ya que puede determinarse la secuencia consenso de dos cadenas de ADN complementarias. La falta de complementariedad en cualquier nucleótido entre las secuencias consenso monocatenarias del mismo fragmento de ADN bicatenario debe reconocerse como error técnico. Sin embargo, el experto en la técnica entenderá fácilmente que el término “libre de errores” no se refiere a una eliminación completa de errores técnicos, sino más bien a una reducción de su frecuencia hasta un nivel despreciable. En particular, el experto apreciará que cuando se usa el ADN de una célula individual o una molécula de ADN individual como material de partida que va a emplearse en los métodos de la presente invención, la amplificación del ADN introducirá errores que posteriormente pueden retirarse mediante corrección de errores usando los métodos de la presente invención. Cantidades más grandes de material de partida, por ejemplo ADN extraído de tejido, tal como también se proporciona en el presente documento, pueden no requerir amplificación de ADN antes de la secuenciación de ADN, dando posiblemente como resultado una tasa de errores reducida en la secuenciación de ADN, que también pueden corregirse usando los métodos de la presente invención.

35 El término “secuencia aleatorizada” según la invención debe entenderse como una secuencia de nucleótidos en la que cada posición tiene una probabilidad independiente e igual de ser cualquier nucleótido. Los nucleótidos aleatorios pueden ser cualquiera de los nucleótidos, por ejemplo G, A, C, T, U, o análogos químicos de los mismos, en cualquier orden, en los que: se entiende que G representa nucleótidos guanílicos, A nucleótidos adenílicos, T nucleótidos timidílicos, C nucleótidos citidílicos y U nucleótidos uracílicos. El experto apreciará que los métodos de síntesis de oligonucleótidos conocidos pueden conducir de manera inherente a una representación distinta de los nucleótidos G, A, C, T o U. Por ejemplo, la síntesis puede conducir a una sobrerrepresentación de nucleótidos, tales como G en secuencias de ADN aleatorizadas. Esto puede conducir a un número reducido de secuencias aleatorias únicas tal como se espera basándose en una representación igual de nucleótidos. Sin embargo, el experto es consciente de que el número global de secuencias aleatorias únicas comprendido en el segundo oligonucleótido usado en los métodos de la invención será generalmente suficiente para identificar claramente cada fragmento de ADN diana. Esto se debe a que el experto también es consciente del hecho de que la longitud de la secuencia aleatorizada puede variarse dependiendo del número de fragmentos resultantes de la fragmentación de ADN. El número previsto de fragmentos de ADN puede derivarse del número de sitios de escisión de una endonucleasa de restricción y la longitud del ADN diana. Por consiguiente, el experto puede tener fácilmente en cuenta la posible representación diferente de nucleótidos en la secuencia aleatorizada del segundo oligonucleótido usado en los métodos de la invención, que se debe a eficiencias de acoplamiento diferentes de nucleótidos en métodos de síntesis de oligonucleótidos convencionales conocidos, basándose en el conocimiento general en la técnica.

60 En particular, el experto es consciente de que la longitud de la secuencia aleatorizada puede aumentarse con el fin de obtener un número aumentado de secuencias aleatorizadas únicas.

65

Los términos “complementario” o “complementariedad” se refieren a la unión natural de polinucleótidos en condiciones de sal y temperatura permisivas mediante apareamiento de bases. Por ejemplo, la secuencia “A-G-T” se une a la secuencia complementaria “T-C-A”. La complementariedad entre dos moléculas monocatenarias puede ser “parcial”, en las que sólo algunos nucleótidos de los ácidos nucleicos se unen entre sí, o puede ser completa cuando existe complementariedad total entre moléculas monocatenarias. El grado de complementariedad entre cadenas de ácido nucleico tiene efectos significativos sobre la eficiencia e intensidad de la hibridación entre cadenas de ácido nucleico. Esto tiene particular importancia en reacciones de amplificación, que dependen de la unión entre cadenas de ácidos nucleicos. Tal como se usa según la presente invención, el término “fragmentos de ADN de longitud similar” designa fragmentos que, a un nivel estadístico, tienen un tamaño que es de longitud comparable. Fragmentos de ADN de longitud comparable son, por ejemplo, fragmentos de desde aproximadamente 50 +/- 5 pb hasta aproximadamente 4 kpb +/- 0,4 kpb. El intervalo de longitud de fragmentos de ADN que se generan preferiblemente es ventajosamente de entre aproximadamente 50 pb y aproximadamente 4 kpb. También pueden usarse fragmentos de ADN de mayor o menor longitud, aunque pueden amplificarse o representarse en menor medida que los fragmentos del intervalo de tamaño definido anteriormente. Los fragmentos de ADN son preferiblemente adecuados para amplificación lineal y/o exponencial. Preferiblemente, los fragmentos de ADN tienen un tamaño de < 3 kpb, más preferiblemente dichos fragmentos de ADN tienen una longitud promedio de aproximadamente 1000 pb y se prefieren particularmente fragmentos de aproximadamente 100-400 pb.

Los términos “proyecciones en 5'” y “proyecciones en 3'” tal como se usan en el presente documento significan el grupo fosfato en 5' o grupo hidroxilo en 3', proporcionados por ejemplo mediante una escisión escalonada de ADN mediante endonucleasas de restricción, y designa un extremo 5' con proyección monocatenario o un extremo 3' con proyección monocatenario en ADN, respectivamente.

El término “amplificar” se refiere al copiado repetido de una secuencia de nucleótidos especificada dando como resultado un aumento de la cantidad de dicha secuencia de nucleótidos especificada y permite la generación de una multitud de moléculas de ácido nucleico idénticas o esencialmente idénticas (es decir idénticas en al menos el 95%, más preferiblemente al menos el 98%, incluso más preferiblemente al menos el 99% y lo más preferiblemente al menos el 99,5%, tal como el 99,9%) o partes de las mismas. Tales métodos están bien establecidos en la técnica; véase Sambrook *et al.* “Molecular Cloning, A Laboratory Manual”, 2ª edición 1989, CSH Press, Cold Spring Harbor. Pueden aplicarse diversos métodos de amplificación, estos son por ejemplo, amplificación por círculo rodante (tal como en Liu, *et al.*, “Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases”, J. Am. Chem. Soc. 118:1587-1594 (1996)), amplificación isotérmica (tal como en Walker, *et al.*, “Strand displacement amplification— an isothermal, *in vitro* DNA amplification technique”, Nucleic Acids Res. 20(7): 1691-6 (1992)), reacción en cadena de la ligasa (tal como en Landegren, *et al.*, “A Ligase-Mediated Gene Detection Technique”, Science 241: 1077-1080, 1988, o, en Wiedmann, *et al.*, “Ligase Chain Reaction (LCR)—Overview and Applications”, PCR Methods and Applications (Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, NY, 1994) págs. S51-S64). Sin embargo, se prefiere la amplificación mediante reacción en cadena de la polimerasa. Incluyen reacción en cadena de la polimerasa (PCR) y modificaciones de la misma, reacción en cadena de la ligasa (LCR) por nombrar algunos métodos de amplificación preferidos.

Los términos “aparear” o “hibridar” y “apareamiento” o “hibridación” se refieren a la formación de complejos entre secuencias de nucleótidos que son suficientemente complementarias como para formar complejos mediante apareamiento de bases de Watson-Crick. Con respecto a la presente invención, las secuencias de ácido nucleico que son “complementarias a” o “complementarias con” o que se “hibridan” o “se aparean” entre sí deben poder formar o forman “híbridos” o “complejos” que son suficientemente estables como para servir para el propósito previsto. La hibridación o el apareamiento y la intensidad de hibridación (es decir, la intensidad de la asociación entre cadenas de ácido nucleico) se ve influida por muchos factores bien conocidos en la técnica incluyendo por ejemplo el grado de complementariedad entre los ácidos nucleicos, la rigurosidad de las condiciones implicadas afectadas por condiciones tales como la concentración de sales, la  $T_m$  (temperatura de fusión) del híbrido formado, la presencia de otros componentes (por ejemplo, la presencia o ausencia de polietilenglicol o betaína), la molaridad de las cadenas hibridantes y el contenido en G:C de las cadenas de ácido nucleico.

Los métodos de la presente invención permiten la secuenciación libre de errores de ADN. Se conoce bien en la técnica que la amplificación y/o secuenciación introducen errores en el resultado de secuenciación final. Esto se debe, entre otras cosas, a la tasa de errores natural de la polimerasa usada y/o ambigüedades de la identidad de la base durante la secuenciación. La mejor manera de identificar errores es usar la información de secuencia redundante de una molécula de ADN bicatenario. Esto se debe a que la probabilidad de introducción de errores durante la amplificación y/o secuenciación en la misma posición de una muestra de ADN es mínima. Los métodos proporcionados en el presente documento permiten la identificación de tales errores haciendo uso de la información de secuencia redundante de una molécula de ADN bicatenario añadiendo un identificador/código de barras único a cada molécula/fragmento de ADN monocatenario. Por consiguiente, cada molécula/fragmento de ADN bicatenario se marca con etiqueta con un identificador/código de barras único o bien en ambos extremos 5' o bien en ambos extremos 3', respectivamente. La posterior reacción de relleno conduce a la generación de moléculas de ADN monocatenario que están marcadas con etiqueta en cada sitio del ADN diana con dos códigos de barras distintos. Más específicamente, cada molécula de ADN objetivo monocatenario se marca con etiqueta con el código de barras del segundo oligonucleótido que se ligó al fragmento de ADN diana y la secuencia complementaria del código de

- barras del segundo oligonucleótido que se ligó a la cadena complementaria del fragmento de ADN diana. Por consiguiente, ambas moléculas de ADN monocatenario de un fragmento de ADN bicatenario pueden identificarse de manera no ambigua basándose en ambas secuencias de código de barras que están unidas al extremo 5' y 3', respectivamente, de cada ADN monocatenario. En una primera etapa, se clasifican las secuencias según los códigos de barras en los extremos 3' y 5'. En una segunda etapa, se identifica la cadena complementaria de la molécula de ADN bicatenario original mediante los códigos de barras complementarios. De ese modo es posible identificar, entre otras cosas, errores que se han introducido durante la preparación de muestras, tal como amplificación y/o secuenciación, y por tanto identificar mutaciones auténticas, por ejemplo, en alelos de células tumorales. La identificación fiable de mutaciones auténticas en células tumorales de pacientes individuales puede permitir desarrollar y/o mejorar terapias contra el cáncer dirigidas personalizadas eficientes. Además de la identificación de mutaciones en una molécula de ADN, los métodos de la presente invención pueden usarse para la identificación libre de errores de modificaciones de la molécula de ADN objetivo, por ejemplo metilación, en particular metilación de citosinas (Laird 2010 Nature Reviews Genetics 11, 191-203).
- Los oligonucleótidos de la invención pueden ser oligonucleótidos de ADN, oligonucleótidos de ARN u oligonucleótidos de ADN/ARN. En particular, los oligonucleótidos pueden estar parcial o completamente compuestos por nucleótidos de ácido ribonucleico y/o nucleótidos de ácido desoxirribonucleico, respectivamente. En una realización de la invención, la segunda secuencia del segundo oligonucleótido consiste en nucleótidos de ácido ribonucleico. Por consiguiente, tras ligar el segundo oligonucleótido de composición ADN-ARN-ADN a una molécula monocatenaria de un fragmento de ADN bicatenario, la adición de enzima(s) con propiedades de síntesis de ADN dependientes de ARN y/o dependientes de ADN generará fragmentos de ADN bicatenario de la composición ADN:ADN-ARN:ADN-ADN:ADN en el sitio del segundo oligonucleótido ligado. La adición de ribonucleasa(s) que hidrolizan específicamente enlaces fosfodiéster en híbridos de ARN-ADN y que digieren la parte de ARN monocatenario de segundos oligonucleótidos libres eliminará todas las secuencias de ARN en la reacción creando así un fragmento bicatenario con un hueco interno en el sitio del segundo oligonucleótido y el tercer oligonucleótido. El hueco resultante en el fragmento de ADN bicatenario se rellenará mediante la ADN polimerasa ya presente en la reacción. La adición de una ligasa unirá de manera covalente la parte extendida del segundo oligonucleótido a la secuencia fija del segundo oligonucleótido.
- El segundo oligonucleótido de la invención puede comprender además una cuarta secuencia que puede contener un motivo específico de sitio para una enzima de restricción, por ejemplo de una endonucleasa de guiado. Por tanto, se prefiere que la cuarta secuencia esté ubicada o bien entre la segunda y la tercera secuencia o bien dentro de la tercera secuencia. Por consiguiente, la presencia de la cuarta secuencia del segundo oligonucleótido permite reducir la longitud del producto de amplificación eliminando el sitio de unión a oligonucleótido, es decir la tercera secuencia, que no se requiere para la secuenciación de ADN. Esto permite la secuenciación de fragmentos de ADN derivados de muestra o de célula más largos debido al acortamiento del producto de amplificación. El sitio de enzima de restricción puede usarse para ligar adaptadores de secuenciación con alta eficiencia o para concatenar los fragmentos de ADN tras la amplificación para tecnologías de secuenciación de nueva generación tales como enfoques de secuenciación de genoma completo.
- Tal como se usa en el presente documento, el término "endonucleasa de restricción" se refiere a enzimas que pueden cortar ADN bicatenario en o cerca de una secuencia de nucleótidos específica. Las endonucleasas de restricción (enzimas de restricción) están presentes en muchas especies y pueden unirse de manera específica de secuencia a ADN (en un sitio de reconocimiento), y escindir ADN en o cerca del sitio de unión. Determinadas enzimas de restricción (por ejemplo, de tipo IIS) escinden ADN en sitios remotos del sitio de reconocimiento y tienen unidades de unión y escisión diferenciadas. Por ejemplo, la enzima FokI de tipo IIS cataliza la escisión bicatenaria de ADN, a 9 nucleótidos desde su sitio de reconocimiento en una cadena y a 13 nucleótidos desde su sitio de reconocimiento en la otra. Véanse, por ejemplo, las patentes estadounidenses n.ºs 5.356.802; 5.436.150 y 5.487.994; así como Li *et al.*, Proc. Natl. Acad. Sci. USA, 89:4275-4279, 1992; Li *et al.*, Proc. Natl. Acad. Sci. USA, 90:2764-2768, 1993; Kim *et al.*, J. Biol. Chem., 269:31,978-31,982, 1994b; Kim *et al.*, Proc. Natl. Acad. Sci. USA, 91:883-887, 1994a.
- Se describen enzimas de restricción de tipo IIS a modo de ejemplo en la publicación internacional WO 07/014.275. Otras enzimas de restricción también contienen dominios de unión y escisión separables, y la presente divulgación contempla las mismas. Véase por ejemplo, Roberts *et al.*, Nucleic Acids Res., 31:418-420, 2003.
- Cualquier nucleasa que tenga un sitio diana en el ADN diana puede usarse en los métodos dados a conocer en el presente documento. Por ejemplo, las endonucleasas de guiado y meganucleasas tienen secuencias de reconocimiento muy largas, algunas de las cuales es probable que estén presentes, de manera estadística, una vez en un genoma de tamaño humano.
- Las endonucleasas de guiado a modo de ejemplo adecuadas para su uso en la etapa (f') de los métodos de la invención incluyen I-SceI, ICeul, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-Scell, I-Ppol, I-ScellI, I-CreI, I-TevI, I-TevII y I-TevIII. Sus secuencias de reconocimiento se conocen. Véanse también la patente estadounidense n.º 5.420.032; patente estadounidense n.º 6.833.252; Belfort *et al.*, Nucleic Acids Res., 25:3379-3388, 1997; Dijon *et al.*, Gene, 82:115-118, 1989; Perler *et al.*, Nucleic Acids Res., 22:1125-1127, 1994; Jasin, Trends Genet., 12:224-228, 1996;

Gimble *et al.*, J. Mol. Biol., 263:163-180, 1996; Argast *et al.*, J. Mol. Biol., 280:345-353, 1998 y el catálogo de New England Biolabs.

Aunque la especificidad de escisión de la mayoría de las endonucleasas de guiado no es absoluta con respecto a sus sitios de reconocimiento, los sitios tienen una longitud suficiente como para que pueda obtenerse un único acontecimiento de escisión por genoma de tamaño de mamífero expresando una endonucleasa de guiado en una célula que contiene una única copia de su sitio de reconocimiento. También se ha notificado que la especificidad de endonucleasas de guiado y meganucleasas puede modificarse por ingeniería para unirse a sitios diana no naturales. Véanse, por ejemplo, Chevalier *et al.*, Molec. Cell, 10:895-905, 2002; Epinat *et al.*, Nucleic Acids Res., 31:2952-2962, 2003; Ashworth *et al.*, Nature, 441:656-659, 2006; Paques *et al.*, Current Gene Therapy, 7:49-66, 2007.

Además, la presente invención proporciona un oligonucleótido de cuatro partes, que comprende una primera, segunda, tercera y cuarta secuencias, en el que la primera secuencia comprende una secuencia fija, la segunda secuencia comprende una secuencia aleatorizada, la tercera secuencia comprende un sitio de unión a cebador y la cuarta secuencia comprende un sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción. Según la presente invención, preferiblemente la secuencia fija comprende aproximadamente de 4 a 15 nucleótidos, preferiblemente la secuencia aleatorizada comprende aproximadamente de 3 a 24 nucleótidos y el sitio de reconocimiento de nucleasa de restricción comprende preferiblemente un sitio de reconocimiento de una endonucleasa de guiado. El oligonucleótido de cuatro partes tiene el siguiente orden preferido de 5' a 3' de las cuatro secuencias/partes: extremo 5' seguido por sitio de unión a cebador (tercera secuencia) o sitio de restricción (cuarta secuencia) seguido por sitio de unión a cebador o sitio de restricción (dependiendo de la selección de la secuencia más en 5') seguido por secuencia aleatoria (segunda secuencia) seguida por secuencia fija (primera secuencia) seguida por extremo 3'.

El término "oligonucleótido", según la presente invención, incluye cualquier molécula de ácido nucleico, tal como ADN, por ejemplo ADNc o ADN genómico, y ARN. Se incluyen además moléculas que imitan ácido nucleico conocidas en la técnica tales como derivados sintéticos o semisintéticos de ADN o ARN y polímeros mixtos. Tales moléculas que imitan ácido nucleico o derivados de ácido nucleico según la invención incluyen ácido nucleico de fosforotioato, ácido nucleico de fosforamidato, ácido 2'-O-metoxietil-ribonucleico, ácido morfolinonucleico, ácido hexitolnucleico (AHN) y ácido nucleico bloqueado (ANB) (véase Braasch y Corey, Chem Biol 2001, 8: 1), etc. ANB es un derivado de ARN en el que el anillo de ribosa está restringido mediante un enlace metileno entre el oxígeno en 2' y el carbono en 4'. Pueden contener bases de nucleótido derivadas o no naturales adicionales, tal como apreciarán fácilmente los expertos en la técnica.

Además, la presente invención se refiere a un oligonucleótido que puede amplificar específicamente las moléculas de ácido nucleico de la presente invención. Por consiguiente, los oligonucleótidos dentro del significado de la invención pueden ser capaces de servir como punto de partida para la amplificación, es decir pueden ser capaces de servir como cebadores. En particular, el tercer oligonucleótido de la invención sirve preferiblemente como cebador para la amplificación mediante PCR. Dicho oligonucleótido también puede comprender oligorribos o desoxirribonucleótidos que son complementarios a una región de una de las cadenas de una molécula de ácido nucleico. Según la presente invención, un experto en la técnica entenderá fácilmente que el término "cebador" también puede referirse a un par de cebadores que, con respecto a una región complementaria de una molécula de ácido nucleico, se dirigen en sentido opuesto uno hacia el otro para permitir, por ejemplo, la amplificación mediante reacción en cadena de la polimerasa (PCR). Generalmente se prevé la purificación del/de los cebador(es), antes de su uso en el método de la presente invención. Tales etapas de purificación pueden comprender HPLC (cromatografía de líquidos de alta resolución) o PAGE (electroforesis en gel de poliacrilamida) y las conoce el experto en la técnica.

Cuando se usa en el contexto de cebadores, en particular el tercer oligonucleótido de la invención, el término "específicamente" significa que sólo se amplifican las moléculas de ácido nucleico deseadas tal como se describe en el presente documento. Por tanto, un cebador según la invención es preferiblemente un cebador que se une a una región de una molécula de ácido nucleico que es única para esta molécula. En relación con un par de cebadores, según la invención, es posible que uno de los cebadores del par sea específico en el sentido anteriormente descrito o que ambos de los cebadores del par sean específicos.

El extremo 3'-OH de un cebador se usa por una polimerasa para extenderse mediante incorporación sucesiva de nucleótidos. El cebador o par de cebadores de la presente invención puede usarse, por ejemplo, en experimentos de extensión por cebador con ADN de molde según métodos conocidos por el experto en la técnica. Preferiblemente, el cebador o par de cebadores de la presente invención se usan para reacciones de amplificación con ADN de molde, preferiblemente ADN genómico. El término "ADN de molde" se refiere a moléculas de ADN o fragmentos de las mismas de cualquier fuente o composición de nucleótidos, que comprenden una secuencia de nucleótidos diana tal como se definió anteriormente. El cebador o par de cebadores también pueden usarse para experimentos de hibridación tal como se conoce en la técnica. Preferiblemente, el cebador o par de cebadores se usan en reacciones en cadena de la polimerasa para amplificar secuencias correspondientes a una secuencia de la molécula de ácido nucleico de la presente invención. Se sabe que la longitud de un cebador resulta de diferentes parámetros (Gillam, Gene 8 (1979), 81-97; Innis, PCR Protocols: A guide to methods and applications, Academic Press, San Diego,

EE.UU. (1990)). Preferiblemente, el cebador sólo debe hibridarse o unirse a una región específica de una secuencia de nucleótidos diana. La longitud de un cebador que estadísticamente sólo se hibrida a una región de una secuencia de nucleótidos diana puede calcularse mediante la siguiente fórmula:  $(\frac{1}{4})^x$  (donde x es la longitud del cebador). Sin embargo, se sabe que un cebador que coincide exactamente con una cadena de molde complementaria debe tener al menos 9 pares de bases de longitud, de lo contrario no pueden generarse cadenas dobles estables (Goulian, Biochemistry 12 (1973), 2893-2901). También se prevé que pueden usarse algoritmos basados en ordenador para diseñar cebadores que pueden amplificar ADN. También se prevé que se marcan el cebador o par de cebadores. El marcador puede ser, por ejemplo, un marcador radiactivo, tal como  $^{32}\text{P}$ ,  $^{33}\text{P}$  o  $^{35}\text{S}$ . En una realización preferida de la invención, el marcador es un marcador no radiactivo, por ejemplo, digoxigenina, biotina y tinte o tintes de fluorescencia.

El término "rellenar" o "relleno" tal como se usa en el presente documento significa una reacción de síntesis de ADN, iniciada en extremos 3'-hidroxilo, que conduce a un relleno de la cadena complementaria. Esta reacción de síntesis de ADN se lleva a cabo preferiblemente en presencia de dNTP (dATP, dGTP, dCTP y dTTP, y/o análogos químicos de los mismos). Con frecuencia se usan ADN polimerasas termoestables tales como polimerasas Taq y las conoce bien el experto en la técnica.

En una realización preferida de la presente invención, la cuarta secuencia del segundo oligonucleótido de la invención comprende un sitio de restricción de una endonucleasa específica de sitio. Con respecto a esto, se prefieren sitios de restricción de endonucleasas de corte poco frecuente. Más específicamente, se prefieren sitios de restricción de endonucleasas nativas o sintéticas modificadas por ingeniería o enzimas nativas o modificadas por ingeniería que cortan normalmente ADN genómico humano de manera muy infrecuente o que permiten la escisión dirigida al sitio de secuencias de ácido nucleico presentes de manera infrecuentemente en el genoma humano. Las enzimas de restricción modificadas por ingeniería pueden ser capaces de reconocer de manera específica de sitio secuencias que no están presentes en el genoma humano, pero que se han introducido de manera artificial. La endonucleasa puede ser una endonucleasa de guiado, endonucleasa de dedos de cinc, TALEN, nucleasa TFO o Targetron. Se prefiere que la endonucleasa sea I-SceI o I-CeuI. Por tanto se prefiere que el sitio de restricción reconocido por la endonucleasa sea 5'-TAGGGATAACAGGGTAAT-3' o 5'-TAAGTATAACGGTCTAAGGTAGCGAA-3' cuando se usa I-SceI o I-CeuI, respectivamente.

La endonucleasa específica de sitio puede añadirse antes de la etapa (g) de los métodos de la invención, es decir antes de la secuenciación de las moléculas de ácido nucleico amplificadas, preferiblemente el ADN. Se prefiere que la endonucleasa específica de sitio se añada en una etapa (f'), tras la etapa (f). Por consiguiente, se prefiere que los métodos de la invención comprendan además la etapa (f'), en los que se añade una endonucleasa específica de sitio, en particular una endonucleasa de guiado, en la etapa (f'). Se prefiere que la endonucleasa específica de sitio sea una endonucleasa de guiado, más preferiblemente o bien I-SceI o bien I-CeuI.

Los métodos de la presente invención pueden comprender además la adición de una exonucleasa. Dado que las exonucleasas no tienen sitios de reconocimiento específicos, tras la reacción de relleno, el ADN diana con las secuencias artificiales añadidas consistirá en ADN bicatenario que no puede digerirse por exonucleasas específicas de cadena sencilla. Por tanto, la exonucleasa específica de cadena sencilla sólo afectará a secuencias de oligonucleótido sin ligar. En particular, la exonucleasa puede añadirse con el fin de degradar moléculas de ácido nucleico monocatenario en exceso, por ejemplo ADN y/o ARN, preferiblemente ADN. Específicamente, la exonucleasa puede añadirse con el fin de degradar el primer y el exceso del segundo oligonucleótidos usados en los métodos de la invención antes de la amplificación. Por tanto, se prefiere que la exonucleasa se añada antes de la etapa (f) de los métodos de la presente invención. Más específicamente, se prefiere que los métodos de la invención comprendan además una etapa (e'), en los que se añade una exonucleasa en la etapa (e'). La adición de una exonucleasa evita productos secundarios no deseados durante la amplificación. Preferiblemente, la exonucleasa degrada moléculas de ácido nucleico monocatenario en el sentido de 5' a 3' mientras que no actúa en ninguna otra molécula de ácido nucleico o degrada moléculas de ácido nucleico monocatenario en el sentido de 3' a 5' mientras que no actúa sobre ninguna otra molécula de ácido nucleico. Se prefiere que el sitio de restricción reconocido por la exonucleasa sea el de exonucleasa I, nucleasa de frijol mungo, exonucleasa T o RecJ<sub>f</sub>. Lo más preferido es que la eliminación de nucleótidos a partir de moléculas de ADN monocatenario se catalice mediante exonucleasa I o RecJ<sub>f</sub>.

Aunque puede usarse cualquier muestra de ADN para los métodos de la presente invención, se prefiere que la muestra de ADN usada comprenda (i) el genoma o transcriptoma de una célula individual, (ii) cromosoma(s) de una célula individual, (iii) ácidos nucleicos de exosomas u otras microvesículas de una célula individual o (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii).

La célula individual usada en los métodos de la presente invención puede obtenerse a partir de material biológico usado medicina forense, medicina reproductiva o medicina regenerativa. Por consiguiente, la célula individual puede ser una célula tumoral, una célula sanguínea, una célula procedente de aspirados de médula ósea, una célula procedente de un ganglio linfático y/o una célula obtenida a partir de un tejido microdisecado, un blastómero o blastocisto de un embrión, un espermatozoide, una célula obtenida a partir de líquido amniótico, o una célula obtenida a partir de hisopos bucales. Se prefiere que la célula tumoral sea una célula tumoral diseminada, célula tumoral circulante o una célula procedente de biopsias de tumor. Además se prefiere que la célula sanguínea sea

una célula de sangre periférica o una célula obtenida a partir de sangre de cordón umbilical. Se prefiere particularmente que la muestra de ADN consista en (i) el genoma o transcriptoma de una célula individual, (ii) cromosoma(s) de una célula individual, (iii) ácidos nucleicos de exosomas u otras microvesículas de una célula individual o (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii).

En otro aspecto de la invención, la muestra de ADN también puede comprender (i) el ADN de más de una célula individual, (ii) ADN fetal libre de células de más de una célula individual, (iii) ADN libre de células de más de una célula individual en suero y/o plasma (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii). La muestra de ADN también puede consistir en el ADN de más de una célula individual, ADN fetal libre de células de más de una célula individual, o ADN libre de células de más de una célula individual en suero y/o plasma de pacientes con cáncer. La muestra de ADN puede obtenerse a partir de más de una célula individual, en particular dos o más. Se prefiere que la muestra de ADN se obtenga a partir de desde 2 hasta 5000 células individuales.

Además, el ADN usado en los métodos de la presente invención puede modificarse. En particular, el ADN usado para los métodos de la presente invención puede modificarse mediante introducción de sitios de restricción o etiquetas artificiales. Se prefiere que la modificación del ADN usado en los métodos de la presente invención tenga lugar antes de amplificar el ADN. En particular, se prefiere que los métodos de la invención comprendan además una etapa (a'), en los que el ADN se modifica mediante introducción de sitios de restricción y/o etiquetas artificiales en la etapa (a').

La célula individual y/o las más de una célula individual usadas en los métodos de la presente invención pueden ser de cualquier origen. Puede(n) obtenerse a partir de diversas fuentes de material biológico. El material biológico usado como origen de la(s) célula(s) individual(es) puede usarse, por ejemplo, en medicina forense, medicina reproductiva o medicina regenerativa. Se prefiere que la célula individual o las células individuales usadas en los métodos de la invención sean célula(s) tumoral(es) diseminada(s), célula(s) tumoral(es) circulante(s), célula(s) de sangre periférica, célula(s) procedente(s) de aspirados de médula ósea, célula(s) procedente(s) de biopsias de tumor, célula(s) obtenida(s) a partir de sangre de cordón umbilical, célula(s) obtenida(s) a partir de un ganglio linfático y/o célula(s) obtenida(s) a partir de tejido microdisecado, blastómero (s) o blastocisto(s) de un embrión, espermatozoide(s), célula(s) obtenida(s) a partir de líquido amniótico, o célula(s) obtenida(s) a partir de hisopos bucales, o cuerpos polares.

Los métodos de la presente invención pueden comprender además, antes de la etapa (b), en particular tras la etapa (a), la etapa (a''), en los que dicha muestra que comprende ADN se digiere con una proteinasa. La proteinasa puede ser termolábil o inactivarse mediante otros medios tales como inactivación química. Preferiblemente, dicha proteinasa es termolábil. Por consiguiente, dicha proteinasa puede inactivarse térmicamente en la etapa (a'''). Se prefiere particularmente que dicha proteinasa sea proteinasa K.

Los métodos de la presente invención también pueden comprender una etapa de análisis de metilación de ADN. Se sabe que mecanismos epigenéticos desempeñan papeles importantes durante el desarrollo normal, el envejecimiento y una variedad de estados patológicos. Tales enfermedades pueden ser enfermedades humanas, incluyendo cáncer, esclerosis múltiple, diabetes y/o esquizofrenia. La hipermetilación de islas de CpG ubicadas en las regiones de promotor de genes supresores de tumores está firmemente establecida como mecanismo frecuente para la inactivación génica en cánceres (Hansen *et al.* 2011. *Nat. Genet.* 43, 768-775). La metilación del carbono en 5' de citosina es una forma de modificación epigenética que no afecta a la secuencia de ADN primaria, pero afecta a interacciones secundarias que desempeñan un papel crítico en la regulación de la expresión génica. La metilación de ADN aberrante puede suprimir la transcripción y posteriormente la expresión génica. El análisis de metilación como en los métodos de la presente invención puede comprender la modificación selectiva del ADN diana. Tal modificación puede comprender la adición de enzimas de restricción dependientes de la metilación (MDRE) o enzimas de restricción sensibles a la metilación (MSRE), preferiblemente MDRE. La modificación selectiva del ADN diana también puede comprender la adición de un agente químico que puede diferenciar de manera selectiva entre nucleótidos metilados o sin metilar. En particular, el análisis de metilación como en la presente invención puede identificar de manera selectiva citosinas metiladas que posteriormente pueden leerse usando el método de secuenciación libre de errores de la presente invención. Por ejemplo, se sabe que el tratamiento con bisulfito convierte citosinas sin metilar (C) en uracilo (U) mientras que las citosinas metiladas no se convierten (Frommer *et al.* 1992. *Proc. Natl. Acad. Sci. USA* 89, 1827-1831). La secuenciación de ADN tras el tratamiento con bisulfito puede usarse para identificar nucleótidos metilados, en particular citosinas. El tratamiento con MDRE conduce a restricción dependiente de la metilación de fragmentos de ADN, mientras que el tratamiento con MSRE conduce a inhibición dependiente de la metilación de la restricción. La secuenciación de ADN tras la restricción con MDRE/MSRE además de la restricción con MseI puede usarse para identificar nucleótidos metilados, en particular citosinas. Por consiguiente, la presente invención proporciona un método de análisis de metilación de ADN libre de errores que comprende como etapa adicional a los métodos de la presente invención una etapa de modificar selectivamente el ADN diana, en particular una etapa de diferenciación entre nucleótidos metilados y sin metilar comprendidos en el ADN diana. En una realización, la presente invención proporciona un método libre de errores de análisis de metilación de ADN que comprende antes de la etapa (g) de los métodos de la invención una etapa de tratar el ADN con bisulfito.

Por consiguiente, la presente invención proporciona un método que comprende las etapas de:

(a) proporcionar una muestra que comprende ADN;

5 (b) añadir un agente a dicho ADN que modifica selectivamente residuos de ácido nucleico metilados, en particular bisulfito;

(c) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar,

10 en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5', en el que el nucleótido terminal de la proyección está fosforilado o, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 3', en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN;

15 (d) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada;

(e) ligar dicho segundo oligonucleótido a dicho fragmento de ADN;

25 (f) rellenar las proyecciones generadas;

(g) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido;

30 (h) secuenciar dichos fragmentos de ADN amplificados; y

(i) identificar residuos de ácido nucleico metilados, en el que cuando se usa bisulfito como agente en la etapa (b), una citosina (C) corresponde a un residuo metilado en dicha muestra de ADN y un uracilo (U) corresponde a un residuo sin metilar en dicha muestra de ADN.

35 Se prefiere que el método libre de errores de preparación de ADN con posterior análisis de secuencia de ADN también comprenda análisis de metilación que comprende la adición de una enzima de restricción dependiente de la metilación (MDRE) o enzima de restricción sensible a la metilación (MSRE), preferiblemente una MDRE, para diferenciar selectivamente entre nucleótidos metilados y sin metilar comprendidos en el ADN diana. Además se prefiere que la MDRE o MSRE, preferiblemente MDRE, se añada antes de amplificar el fragmento de ADN diana, es decir antes de la etapa (f) de los métodos de la invención. Preferiblemente, la MDRE o MSRE se añade tras ligar el segundo oligonucleótido de la invención a los fragmentos de ADN, es decir tras la etapa (d) de los métodos de la invención. Sin embargo, también se contempla que la MDRE o MSRE se añada junto con, o antes de, la digestión del ADN con una endonucleasa de restricción, es decir la etapa (b) de los métodos de la invención. Tras la adición de la MDRE o MSRE, preferiblemente MDRE, los fragmentos de ADN generados se ligan con el segundo oligonucleótido de la presente invención con el fin de identificar de manera única cada fragmento de ADN y permitir el análisis de ADN libre de errores tal como se proporciona en el presente documento. Ejemplos de MDRE y MSRE preferidas para su uso en los métodos de la presente invención son, entre otros, FspEI, MspJI, LpnPI y AcclI, HpaII, DpnI, respectivamente.

50 Por consiguiente, la presente invención proporciona un método que comprende las siguientes etapas:

(a) proporcionar una muestra que comprende ADN;

55 (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar,

en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5', en el que el nucleótido terminal de la proyección está fosforilado o, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 3', en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN;

60 (c) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada;

65

(d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN;

(e) digerir dichos fragmentos de ADN ligados con una MDRE o MSRE, preferiblemente una MDRE;

(f) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada;

(g) ligar dicho segundo oligonucleótido a dicho fragmento de ADN;

(h) rellenar las proyecciones generadas;

(i) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido; y

(j) secuenciar dichos fragmentos de ADN amplificados.

Se prefiere particularmente que la endonucleasa de restricción usada según la presente invención, en particular usada en los métodos de la invención en la etapa (b), reconozca un motivo con de cuatro a seis bases definidas. Tales endonucleasas comprenden enzimas que tienen cuatro nucleótidos diferenciados, por ejemplo MseI, en su sitio de reconocimiento así como enzimas en las que una(s) base(s) de balanceo adicional(es) se encuentra(n) dentro del sitio de restricción, tales como por ejemplo Apol. Preferiblemente, la endonucleasa de restricción usada según la presente invención, en particular usada en los métodos de la invención en la etapa (b), reconoce la secuencia consenso TTAA.

Lo que más se prefiere para los métodos de la presente invención, en particular para la etapa (b) de los métodos de la invención, es que la endonucleasa de restricción sea MseI o un isoesquizómero de la misma.

Además se prefiere que la endonucleasa de restricción usada en la etapa (b) de los métodos de la invención no sea una endonucleasa de restricción que pueda escindir ADN en sitios remotos del sitio de reconocimiento. Por ejemplo, la enzima FokI de tipo IIS cataliza la escisión bicatenaria de ADN, a 9 nucleótidos desde su sitio de reconocimiento en una cadena y a 13 nucleótidos desde su sitio de reconocimiento en la otra. Por consiguiente, se prefiere que la endonucleasa de restricción usada en la etapa (b) de los métodos de la invención no sea una endonucleasa de restricción de tipo IIS.

En un aspecto adicional de la presente invención, el segundo oligonucleótido usado en los métodos de la presente invención es más largo que el primer oligonucleótido. La longitud en exceso del segundo oligonucleótido con respecto al primer oligonucleótido regula la unión/hibridación del segundo oligonucleótido al primer oligonucleótido, en particular de la primera secuencia del segundo oligonucleótido, es decir la secuencia fija, a la segunda secuencia del primer oligonucleótido. Además, se prefiere que el primer oligonucleótido se disocie del complejo de oligonucleótido-ADN tras ligarse el segundo oligonucleótido al fragmento de ADN. Por consiguiente, se prefiere que la unión/hibridación se optimice para permitir la unión específica de la segunda secuencia del primer oligonucleótido a la primera secuencia, es decir secuencia fija, del segundo oligonucleótido y que la unión/hibridación se optimice para permitir la disociación del primer oligonucleótido a partir del complejo de oligonucleótido-ADN. Preferiblemente, el primer oligonucleótido comprende aproximadamente de 4 a 15 nucleótidos y el segundo oligonucleótido comprende aproximadamente de 30 a 60 nucleótidos.

La longitud de la secuencia aleatorizada, es decir la segunda secuencia del segundo oligonucleótido, depende del número deseado de códigos de barras/identificadores únicos y puede variarse en consecuencia. Además, se prefiere que la segunda secuencia del segundo oligonucleótido usado en los métodos de la presente invención, es decir la secuencia aleatorizada del segundo oligonucleótido, comprenda aproximadamente de 3 a 24 nucleótidos. Se prefiere más que la segunda secuencia del segundo oligonucleótido usado en los métodos de la presente invención, es decir la secuencia aleatorizada del segundo oligonucleótido, comprenda al menos 3, más preferiblemente al menos 4 y lo más preferiblemente al menos 5 nucleótidos.

Según la presente invención, el primer oligonucleótido comprende una primera secuencia complementaria a la proyección generada del/de los fragmento(s) de ADN, que comprende preferiblemente los nucleótidos T y A, una segunda secuencia complementaria a la secuencia fija del segundo oligonucleótido, que se varía a partir de cada muestra que va a analizarse y que comprende preferiblemente de 4 a 15 nucleótidos, y un nucleótido sin un extremo 5'-fosfato-terminal, preferiblemente el nucleótido C.

La variación de la secuencia fija del segundo oligonucleótido y por consiguiente la variación de la segunda secuencia

- del primer oligonucleótido, permite la identificación diferenciada de muestras basándose en la secuencia fija que puede asociarse con cada muestra como identificador de muestra (SID). Esto permite la identificación inequívoca de lecturas de secuencia de una muestra particular dentro de una multitud de lecturas que se originan a partir de otros ejemplares, permitiendo el análisis en paralelo de múltiples muestras dentro de una serie de secuenciación. Por consiguiente, esto permite un mayor rendimiento de muestra ya que pueden procesarse simultáneamente para su secuenciación muestras amplificadas de manera más independiente. Además, el SID puede usarse para analizar sintácticamente archivos de secuenciación asignando así claramente secuencias a una muestra. Las secuencias previamente clasificadas permitirán una evaluación más rápida de las muestras de pacientes.
- 5
- 10 En una realización particular de la invención, se prefiere que el primer oligonucleótido usado en los métodos de la presente invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia 5'-TAACTGACdd-3' y/o que el segundo oligonucleótido usado en los métodos de la presente invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 1 y/o que el tercer oligonucleótido usado en los métodos de la presente invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 2.
- 15
- 20 En una realización adicional de la invención, se prefiere que el primer oligonucleótido usado en los métodos de la presente invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia 5'-TAACGACdd-3' y/o que el segundo oligonucleótido usado en los métodos de la presente invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 6 y/o que el tercer oligonucleótido usado en los métodos de la presente invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 2.
- 25
- 30 Lo más preferido es que el primer oligonucleótido usado en los métodos de la invención tenga la secuencia 5'-TAACTGACdd-3' y/o que el segundo oligonucleótido usado en los métodos de la invención tenga una secuencia tal como se muestra en SEQ ID NO: 1 y/o que el tercer oligonucleótido usado en los métodos de la invención tenga una secuencia tal como se muestra en SEQ ID NO: 2.
- 35
- 40 En un aspecto adicional de la invención, el primer y el segundo oligonucleótidos usados en los métodos de la presente invención pueden hibridarse entre sí de manera separada del ADN. En particular, los oligonucleótidos hibridados pueden añadirse al ADN o fragmentos de ADN antes de la etapa (d) de los métodos de la invención, es decir antes de la ligación, en particular tras la etapa (c) de los métodos de la invención. Por consiguiente, se prefiere que los métodos de la invención comprendan además una etapa (c'), en los que el primer oligonucleótido y el segundo oligonucleótido se hibridan entre sí de manera separada de los fragmentos de ADN y se añaden a los fragmentos de ADN en la etapa (c').
- 45
- El primer oligonucleótido usado en los métodos de la presente invención puede modificarse adicionalmente ya que el último nucleótido en 3' de dicho oligonucleótido es un didesoxi (dd)-nucleótido.
- 50
- Los métodos de la invención permiten la amplificación y/o el análisis de ADN libre de errores esencialmente de todo el genoma nuclear de una célula, preferiblemente la amplificación esencialmente del genoma completo de una célula individual. Tal como entiende un experto en la técnica, la amplificación de genoma completo se refiere a métodos en los que se amplifica esencialmente el genoma completo, no refiriéndose necesariamente a un método de amplificación en el que se amplifica cada nucleótido presente en el genoma. Sin embargo, se prefiere que los métodos de la presente invención amplifiquen el genoma completo de una célula, preferiblemente el genoma completo de una célula individual.
- 55
- La célula individual usada en los métodos de la presente invención puede haberse sometido a fijación química. La fijación química puede comprender la fijación usando formalina y/o acetona.
- 60
- Según la invención, puede añadirse una endonucleasa de guiado antes de la etapa (g) de los métodos de la invención, en particular tras la etapa (f) de los métodos de la invención. Por consiguiente, se prefiere que los métodos de la invención comprendan además una etapa (f'), en los que se añade una endonucleasa de guiado en la etapa (f'). Con respecto a esto, se prefiere que dicha endonucleasa de guiado sea I-SceI o I-CeuI.
- 65
- Los métodos de la presente invención pueden llevarse a cabo en un recipiente de reacción. En particular, las etapas (a) a (f) de los métodos de la presente invención pueden llevarse a cabo en un recipiente de reacción. Sin embargo, se prefiere que la endonucleasa específica de sitio reacción (etapa f'), preferiblemente endonucleasa de guiado reacción, se lleve a cabo en un recipiente de reacción independiente.

En una realización preferida de la presente invención, el método de secuenciación libre de errores de ADN comprende las etapas de (a) proporcionar una muestra que comprende ADN de una célula individual; (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5' en el que el nucleótido terminal de la proyección está fosforilado o proyecciones en 3' en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN y en el que dicha endonucleasa de restricción reconoce la secuencia consenso TTAA; (c) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada; (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN; (e) rellenar proyecciones generadas; (f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una primera secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido; y (g) secuenciar dichos fragmentos de ADN amplificados.

En una realización de la presente invención, el método de secuenciación libre de errores de ADN comprende las etapas de (a) proporcionar una muestra que comprende ADN; (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5' en el que el nucleótido terminal de la proyección está fosforilado o proyecciones en 3' en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN y en el que dicha endonucleasa de restricción reconoce la secuencia consenso TTAA; (c) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada, en el que dicho primer oligonucleótido comprende la secuencia de ácido nucleico 5'-TAACTGACdd-3' y en el que dicho segundo oligonucleótido comprende una secuencia de ácido nucleico tal como se expone en SEQ ID NO: 1; (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN; (e) rellenar proyecciones generadas; (f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una primera secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido, en el que dicho tercer oligonucleótido comprende una secuencia de ácido nucleico tal como se expone en SEQ ID NO: 2; y (g) secuenciar dichos fragmentos de ADN amplificados.

En otra realización de la presente invención el método de secuenciación libre de errores de ADN comprende las etapas de (a) proporcionar una muestra que comprende ADN; (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5' en el que el nucleótido terminal de la proyección está fosforilado o proyecciones en 3' en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN y en el que dicha endonucleasa de restricción es Msel o un isoesquizómero de la misma; (c) aparear el primer oligonucleótido a dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada, en el que dicho primer oligonucleótido tiene la secuencia de ácido nucleico 5'-TAACTGACdd-3' y en el que dicho segundo oligonucleótido tiene una secuencia de ácido nucleico tal como se expone en SEQ ID NO: 1; (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN; (e) rellenar las proyecciones generadas; (f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una primera secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido, en el que dicho tercer oligonucleótido tiene una secuencia de ácido nucleico tal como se expone en SEQ ID NO: 2; y (g) secuenciar dichos fragmentos de ADN amplificados.

En una realización adicional de la presente invención el método de secuenciación libre de errores de ADN comprende las etapas de (a) proporcionar una muestra que comprende ADN; (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5' en el que el nucleótido terminal de la proyección está fosforilado o proyecciones en 3' en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN y en el que dicha endonucleasa de restricción es Msel o un isoesquizómero de la misma; (c) aparear el primer oligonucleótido a dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda, una tercera y una cuarta secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada y en el que dicha cuarta secuencia de dicho segundo oligonucleótido comprende un sitio de restricción de una endonucleasa específica de sitio; (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN; (e)

rellenar las proyecciones generadas; (e') añadir una exonucleasa; (f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una primera secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido; (f') añadir una endonucleasa específica de sitio; y (g) secuenciar dichos fragmentos de ADN amplificados.

5 La invención se refiere a un método de secuenciación libre de errores de ADN comprende las etapas de (a) proporcionar una muestra que comprende ADN; (a'') digerir el ADN con una proteinasa tal como, por ejemplo, proteinasa K; (a''') inactivar térmicamente la proteinasa; (b) digerir el ADN con una endonucleasa de restricción en  
10 condiciones adecuadas para obtener fragmentos de ADN de longitud similar, en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5' en el que el nucleótido terminal de la proyección está fosforilado o proyecciones en 3' en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN y en el que dicha endonucleasa de restricción puede ser MseI o un isoesquímico de la misma; (c) aparear el primer oligonucleótido a dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda  
15 secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia de ARN aleatorizada; (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN; (e) rellenar las proyecciones generadas mediante adición de transcriptasa inversa y una ADN polimerasa termoestable; (e') añadir enzimas de digestión de ARN, tales como, por ejemplo, ARNasa H y ARNasa If; (e'') añadir ligasa; (f) amplificar dichos fragmentos de ADN; y (g)  
20 secuenciar dichos fragmentos de ADN amplificados.

La invención se refiere además al uso de los fragmentos de ADN secuenciado obtenidos mediante los métodos de la invención. En particular, la invención se refiere al uso de la información de secuencia obtenida mediante los métodos  
25 de la invención. La información de secuencia puede usarse, por ejemplo, en métodos para el análisis de secuencias de ADN, generación de árboles de linaje celular o evaluación de números de copias. En particular, la información de secuencia obtenida mediante los métodos de la invención puede usarse en métodos para el análisis de secuencias de ADN tales como secuenciación de genoma completo, secuenciación de exoma completo, secuenciación de reguloma completo, análisis de metilación basado en secuenciación, detección de punto de rotura basado en  
30 secuenciación, secuenciación de ChIP o secuenciación dirigida. El método de presente invención es particularmente útil para todos los enfoques anteriormente mencionados, cuando la cantidad de ácido nucleico de entrada, preferiblemente ADN, está fuertemente limitada, es decir ADN de célula individual o fracciones del mismo. Adicionalmente, el método de presente invención puede ser particularmente útil para unos enfoques de secuenciación de alto rendimiento, incluyendo enfoques de secuenciación profunda, que buscan variantes de  
35 secuencia poco frecuentes (es decir transcritos, isoformas/variantes de transcripción, productos intermedios de corte y empalme, sitios aberrantes de cambios epigenéticos, mutaciones puntuales, indels y otras variaciones de secuencia y/o mutaciones) ocultas en el trasfondo de secuencias que muestran perfil de expresión/perfil epigenético/genotipo de tipo natural (sin cambiar). Además, la información de secuencia generada mediante los métodos de la presente invención puede usarse para identificar sitios de metilación dentro del ADN diana.

40 La presente invención también proporciona un oligonucleótido de cuatro partes, que comprende una primera, segundo, tercera y cuarta secuencias, en el que la primera secuencia comprende una secuencia fija, la segunda secuencia comprende una secuencia aleatorizada, la tercera secuencia comprende un sitio de unión a cebador y la cuarta secuencia comprende un sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción. Según la  
45 presente invención, preferiblemente la secuencia fija comprende aproximadamente de 4 a 15 nucleótidos, preferiblemente la secuencia aleatorizada comprende aproximadamente de 3 a 24 nucleótidos y el sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción es preferiblemente un sitio de reconocimiento y/o sitio de restricción de una endonucleasa de guiado.

50 El sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción del oligonucleótido de la invención está ubicado preferiblemente en el lado de 5' de la secuencia aleatorizada. Además se prefiere que el sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción y el sitio de unión a cebador, es decir la tercera y la cuarta secuencias del oligonucleótido de la invención, sean idénticos y/o solapantes. Lo más preferido es que el sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción y el sitio de unión a cebador, es decir la tercera  
55 y la cuarta secuencias del oligonucleótido de la invención, sean idénticos y solapantes al 100 por ciento.

Se prefiere que el oligonucleótido de cuatro partes de la invención comprenda una secuencia fija que comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia GTCAGT y/o una secuencia  
60 aleatorizada y/o un sitio de reconocimiento de nucleasa de restricción que comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 3 y/o un sitio de unión a cebador que comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%,  
65 preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 4. Por consiguiente, se prefiere que el oligonucleótido de cuatro partes de la invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el

90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 5. También se prefiere que el oligonucleótido de cuatro partes de la invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 12.

5 Lo más preferido es que el oligonucleótido de cuatro partes de la invención comprenda una secuencia fija que tenga la secuencia GTCAGT y/o una secuencia aleatorizada y/o un sitio de reconocimiento de nucleasa de restricción y/o sitio de restricción que tenga una secuencia tal como se muestra en SEQ ID NO: 3 y/o un sitio de unión a cebador que tenga una secuencia tal como se muestra en SEQ ID NO: 4. Por consiguiente, se prefiere que el oligonucleótido de cuatro partes la invención, tenga una secuencia tal como se muestra en SEQ ID NO: 5. También se prefiere que el oligonucleótido de cuatro partes tenga una secuencia tal como se muestra en SEQ ID NO: 12.

10 Lo más preferido es que el oligonucleótido de cuatro partes de la invención comprenda una secuencia fija y/o una secuencia aleatorizada y/o una endonucleasa de restricción sitio de reconocimiento y/o un sitio de unión a cebador, en el que el sitio de unión a cebador tiene una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 13. Por consiguiente, lo más preferido es que el oligonucleótido de cuatro partes de la invención comprenda una secuencia idéntica en al menos el 50%, el 55%, el 60%, el 65%, el 70%, el 75%, el 80%, el 85%, preferiblemente el 90%, el 95% o, lo más preferiblemente, el 100% a la secuencia tal como se muestra en SEQ ID NO: 14.

15 A menos que se defina lo contrario, todos los términos técnicos y científicos usados en el presente documento tienen los mismos significados que entiende habitualmente un experto habitual en la técnica a la que se refiere esta invención. Aunque en la práctica o las pruebas de la presente invención pueden usarse métodos y materiales similares o equivalentes a los descritos en el presente documento, a continuación se describen métodos y materiales adecuados. En caso de conflicto, regirá la presente memoria descriptiva, incluyendo las definiciones. Además, los materiales, métodos y ejemplos son únicamente ilustrativos y no se pretende que sean limitativos.

20 Los métodos y las técnicas de la presente invención se realizan generalmente según métodos convencionales bien conocidos en la técnica y tal como se describe en diversas referencias generales y más específicas que se mencionan y se comentan a lo largo de la totalidad de la presente memoria descriptiva a menos que se indique lo contrario. Véase, por ejemplo, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2<sup>a</sup> ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) y Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992), y Harlow y Lane Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990).

25 La invención también cubre todas las características adicionales mostradas en las figuras de manera individual, aunque pueden no haberse descrito en la descripción anterior o siguiente. Además, puede renunciarse a alternativas individuales de las realizaciones descritas en las figuras y la descripción y alternativas individuales de características de las mismas con respecto al objeto del otro aspecto de la invención.

30 Además, en las reivindicaciones el término "que comprende" no excluye otros elementos o etapas, y el artículo indefinido "un" o "una" no excluye una pluralidad. Una única unidad puede cumplir las funciones de varias características mencionadas en las reivindicaciones. Los términos "esencialmente", "alrededor de", "aproximadamente" y similares en relación con un atributo o un valor particularmente también definen exactamente el atributo o exactamente el valor, respectivamente. Ningún signo de referencia en las reivindicaciones debe interpretarse como que limita el alcance.

35 La presente invención también se ilustra mediante las siguientes figuras.

50 Figura 1. Secuencias de oligonucleótido.

La figura muestra secuencias de oligonucleótido del primer, segundo y tercer oligonucleótidos tal como pueden usarse en los métodos de la presente invención. Además, se muestra la secuencia del oligonucleótido de cuatro partes.

Figura 2. Precisión de ensayo de tres marcadores de PCR seleccionados para predecir un experimento de CGH de metafase satisfactorio.

60 Figura 3. Detección de secuencias específicas tras WGA.

Figura 4. Ensayo de control de calidad y predicción de calidad de WGA en muestras clínicas.

65 Figura 5. Selección del diseño de oligonucleótido óptimo para una preparación de muestras eficiente.

(A, B) Resultados de PCR de 3 plex realizada con muestras procesadas con dúplex de oligonucleótidos o bien

DSL<sub>12</sub>-DSS<sub>6</sub> (A) o bien DSL<sub>9</sub>-DSS<sub>5</sub> (B). Las muestras 1-10 indican muestras de ADN que se originan a partir de una célula individual de un donante sano. Las muestras P1 y P2 se generaron usando combinaciones de células. + y - indican control positivo y negativo, respectivamente. (C-D) Perfiles de aCGH de célula individual de un donante sano (C) y de célula individual de líneas celulares OE-19 (D). Se muestran todos los autosomas. Ambas muestras se generaron con DSL<sub>12</sub>-DSS<sub>6</sub>.

Figura 6. Análisis de la tasa de abandono alélico (ADO). Resultados de RFLP-PCR específica para cuatro marcadores de SNP (SNP1-SNP4). En el análisis se incluyeron tres muestras seleccionadas aleatoriamente, originándose cada una a partir de células individuales. Se sometió cada muestra a una variante diferente de la PCR primaria variando la temperatura de la etapa de apareamiento. L indica la posición del marcador de tamaño de ADN, Ø indica un control de PCR negativo y + un control positivo de la digestión de restricción.

Figura 7. Impacto de la temperatura de apareamiento de oligonucleótidos sobre el apareamiento del oligonucleótido DSL<sub>12</sub> durante la PCR primaria.

A, B, C) Resultados de la PCR de control de calidad de 4 plex de muestras procesadas con tres variantes de la PCR primaria, que difieren en cuanto a la temperatura del apareamiento de oligonucleótidos: 57°C (A), 60°C (B), 63°C (C). Cada una de las muestras NZ1-NZ10 se origina a partir de células individuales procesadas de manera individual de un donante varón sano, mientras que las muestras Combinación 1 y Combinación 2 se generaron con una combinación de células.

Figura 8. Funcionamiento de la exonucleasa en el tampón de reacción de PCR usado para la PCR primaria.

Se trató con exonucleasa I producto de PCR bicatenario (ADNbc) con o sin adición de adaptadores de PCR. Obsérvese la ausencia de efecto del tratamiento con exonucleasa I sobre el rendimiento de ADNbc y reducción del contenido de adaptador de PCR en la reacción.

Figura 9. Detección de secuencia de oligonucleótido DSL tras la amplificación de aplicones de células individuales.

(A, B, C) Secuencias de secuencias seleccionadas aleatoriamente que representan tres fragmentos de restricción con MseI específicos flanqueados por secuencias de DSL<sub>12</sub>. El código de color indica la posición de las secuencias de oligonucleótido: amarillo - secuencia de unión del cebador de DSPCR que incluye el sitio de I-SceI; verde - secuencia de código de barras aleatorizado, violeta - extremo 3' fijo. Obsérvese que la secuencia del código de barras aleatorizado es cada vez diferente demostrando que se introdujeron de hecho códigos de barras únicos.

Figura 10. Condiciones para digestión con proteinasa K.

Figura 11. Condiciones para digestión con MseI.

Figura 12. Condiciones para apareamiento de los adaptadores de PCR.

Figura 13. Condiciones para ligación.

Figura 14. Condiciones para PCR primaria (variante sin tratamiento con exonucleasa I).

Figura 15. Condiciones para PCR primaria (variante que incluye tratamiento con exonucleasa I).

Figura 16. Ilustración de la unión y ligación de oligonucleótidos de la invención a ADN objetivo.

Resumen del mecanismo subyacente a la ligación de los adaptadores de PCR y la posterior reacción de relleno. Debido a la falta de fosfato en el extremo 5' del primer oligonucleótido (DSS), no puede ligarse de manera covalente al extremo 3' de la secuencia de ADN objetivo. Los acontecimientos de ligación se producen preferiblemente entre el extremo 3' del segundo oligonucleótido (DSL) y el extremo 5' de la secuencia de ADN objetivo. Adicionalmente, la modificación del extremo 3' del primer oligonucleótido (introducción del 2',3'-didesoxinucleótido) impide cualquier acontecimiento de cebado iniciado por este oligonucleótido. Durante la etapa de relleno (68°C), el oligonucleótido DSS se disocia de sus parejas de unión, permitiendo que la ADN polimerasa sintetice la secuencia complementaria al oligonucleótido DSL. La posterior amplificación basada en PCR se realiza con el uso del tercer oligonucleótido (DSPCR), que es complementario a la tercera secuencia del oligonucleótido DSL.

Figura 17. Ilustración de uso de exonucleasa con el fin de degradar el segundo oligonucleótido.

El uso de una exonucleasa de cadena sencilla tras la reacción de relleno de la secuencia de adaptador de PCR facilita la eliminación del segundo oligonucleótido residual (no unido a la secuencia diana). La eliminación del segundo oligonucleótido sin ligar impide su interferencia con la posterior amplificación mediante el tercer oligonucleótido no degenerado como cebador.

Figura 18. Aplicación del método a la amplificación de genoma completo (WGA) y comparación con un método del estado de la técnica.

5 Fotografías presentan resultados de PCR de control realizada con productos de WGA generados o bien con el método descrito en Jones (1997) *BioTechniques* 22:938-946 y el documento US 08/742.755 (A) o bien con el método de la presente invención (B). En ambos casos se generaron productos de WGA con cantidad variable de ADN de molde (es decir célula individual, combinación de 10 células y combinación de ~100 células). Carriles M: marcador de peso molecular (marcador de peso molecular de ADN de 2 log; New England Biolabs); carriles 1-2: controles negativos (sin molde) de WGA; carriles 3-7: productos de WGA, cada uno generado con ADN de una célula individual diferente; carriles 8-10: productos de WGA generados con combinaciones de 10 células individuales; carriles 11-13, productos de WGA generados con combinaciones de ~100 células; carril (-), control negativo de PCR; carril (+), control positivo de PCR. Se realizaron experimentos tal como se describe en el ejemplo 10.

15 Figura 19a-d. Ilustración de corrección de errores mediante agrupamiento de lecturas de secuenciación que comparten la misma secuencia aleatorizada.

Se alinearon lecturas de secuenciación contra el conjunto de genoma humano 19 y se agruparon basándose en la secuencia aleatorizada identificada dentro de su secuencia de adaptador respectiva. Se muestran pequeñas secciones de lecturas que se mapean (a) al cromosoma 6 desde la posición 124.934.321 hasta la 124.934.396, (b) al cromosoma 22 desde la posición 39.277.551 hasta la 37.277.608, (c) al cromosoma 22 desde la posición 18.455.108 hasta la 18.455.137 y de posición 18.455.218 a la 18.455.247 y (d) al cromosoma 22 desde la posición 42.820.701 hasta la 42.820.754. En cada ejemplo, la secuencia de referencia respectiva en la posición respectiva se muestra encima de las lecturas. Los asteriscos "\*" debajo de las secuencias de grupos de lectura indican posiciones dentro de un grupo de secuencias aleatorizadas en las que el 100% de las bases concuerdan con la referencia, mientras que los huecos " " indican posiciones dentro de un grupo de secuencias aleatorizadas que difieren en una minoría con respecto a la referencia. Estos errores se consideran errores de secuenciación o de amplificación y por tanto necesitarán corrección de errores. Una "X" debajo de las secuencias de grupos de lectura indica además posiciones dentro de todas las lecturas en la alineación en la que la mayoría de las lecturas en al menos un grupo de secuencias aleatorizadas difieren de la referencia. Estas se consideran o bien SNP (si la diferencia se produce únicamente en un grupo de secuencias aleatorizadas) o bien mutaciones (si la diferencia se produce sistemáticamente en todos los grupos de secuencias aleatorizadas). La presente invención se describe adicionalmente mediante los siguientes ejemplos ilustrativos, no limitativos, que proporcionan una mejor comprensión de la presente invención y de sus muchas ventajas. Los siguientes ejemplos se incluyen para demostrar realizaciones preferidas de la invención. Los expertos en la técnica deben apreciar que las técnicas dadas a conocer en los siguientes ejemplos representan técnicas usadas en la presente invención para funcionar bien en la práctica de la invención, y por tanto puede considerarse que constituyen modos preferidos para su puesta en práctica. Sin embargo, los expertos en la técnica deben apreciar, a la luz de la presente divulgación, que pueden realizarse muchos cambios en las realizaciones específicas que se dan a conocer y todavía obtenerse un resultado igual o similar sin alejarse del espíritu y alcance de la invención.

Figura 20. Secuenciación tasas de errores.

45 Se mapearon secuencias de lectura de una serie de secuenciación libre de errores al conjunto de genoma humano 19. Se muestra la abundancia relativa de deleciones (Del), inserciones (Ins), coincidencias erróneas/sustituciones (Sus) y coincidencias (Coincidencia) de posiciones de mapeo en relación con la puntuación de Phred del mapeo de bases de lectura para esta posición. Esto indica una correlación entre los números crecientes de variaciones a partir de la referencia (Del/Ins/Sus) con la calidad de secuenciación y soporta adicionalmente la necesidad de corrección de errores.

50 A menos que se indique lo contrario, se usaron métodos establecidos de tecnología génica recombinante tal como se describe, por ejemplo, en Sambrook, Russell "Molecular Cloning, A Laboratory Manual", Cold Spring Harbor Laboratory, N.Y. (2001).

55 Ejemplo 1 - Diseño de un método de secuenciación libre de errores de ADN.

El componente crucial del método es el diseño de oligonucleótidos que permiten la amplificación de moléculas de ADN individuales y la identificación y eliminación de mutaciones por artefacto. Estos oligonucleótidos se unen a moléculas de ADN individuales mediante ligación. La eliminación de mutaciones por artefacto se basa en la identificación de cadenas de ADN complementarias que formaban una molécula de ADN bicatenario antes de la manipulación de la muestra (por ejemplo, en una célula individual). Por tanto, los oligonucleótidos que se añaden a la molécula de ADN de interés forman una molécula bicatenaria en el sitio de ligación. Por simplicidad, los oligonucleótidos que forman estos adaptadores para secuenciación libre de errores se denominaron DSL o DSS, para oligonucleótido de secuenciación en dúplex largo y oligonucleótido de secuenciación en dúplex corto, respectivamente. Las secuencias de ambos oligonucleótidos se exponen en la figura 1. Ambos oligonucleótidos son parcialmente complementarios entre sí permitiendo la formación de dúplex de oligonucleótido-oligonucleótido, que

se usan como adaptador en un enfoque de PCR mediada por ligación (figura 1). En la estructura de dúplex, el oligonucleótido DSS forma una proyección en 5', que es compatible con los sitios de restricción introducidos en la representación genómica mediante endonucleasa de restricción (para la enzima de restricción MseI, estas bases son TA), permitiendo una ligación más eficiente de los adaptadores de PCR. Las bases restantes son complementarias al oligonucleótido DSL. Se usa un subíndice m ( $DSS_m$ ) para indicar la longitud de la secuencia complementaria. El oligonucleótido DSS puede contener un didesoxinucleótido para impedir su elongación durante las etapas de polimerización (figura 1). El oligonucleótido DSL está compuesto por de tres a cuatro partes. La secuencia en el extremo 3' es una secuencia fija que es responsable de la formación de dúplex de oligonucleótido-oligonucleótido con el oligonucleótido DSS. Puede variarse y también usarse para generar oligonucleótidos de identidad variable, por ejemplo, para marcar con etiqueta células de un individuo específico (obsérvese que en este caso el oligonucleótido DSS necesita tener la secuencia complementaria). La sección central del oligonucleótido DSL contiene una secuencia aleatorizada, que se usa como código de barras para marcar de manera única cada oligonucleótido en la reacción que se liga al ADN diana. La longitud del código de barras puede variar. Para los ejemplos en este caso se añade el subíndice ( $DSL_n$ ) para indicar el número de bases aleatorias. La tercera secuencia, ubicada más en 5', del oligonucleótido puede contener un motivo específico de sitio para una enzima de restricción, por ejemplo de una endonucleasa de guiado, tal como I-SceI. Este extremo 5' del oligonucleótido  $DSL_n$  se diseña para tener una amplificación basada en PCR eficiente de toda la representación de muestra.

Para permitir la amplificación, se usan las siguientes etapas:

1) Formación del adaptador de DSL y DSS.

2) Ligación del adaptador; en el ejemplo en este caso el DSL está ligándose de manera covalente mientras que el DSS no. Después de eso, se liberará el oligonucleótido DSS más corto a partir del apareamiento de bases en una etapa de desnaturalización leve durante la polimerización (etapa 3).

3) Polimerización de la secuencia complementaria (reacción de relleno) del oligonucleótido DSL. Durante esta etapa el código de barras está generándose como código de barras bicatenario.

4) Se usa un tercer oligonucleótido para amplificación que se une a la región de unión a cebador de PCR en la región 5' del oligonucleótido DSL.

Ejemplo 2 - Resumen general de un método de secuenciación libre de errores de ADN.

El procedimiento usado para la preparación de representación/representaciones genética(s) que permite(n) la posterior secuenciación libre de errores consistió en las siguientes etapas:

(a) Acceso a ADN mediante eliminación de las estructuras celular y proteínas que encapsulan el material de ADN (normalmente realizado usando enzimas proteolíticas, es decir, proteinasa K y/o detergentes).

(b) Digestión del material de ADN usando una enzima de restricción de corte frecuente; en este caso MseI.

(c) Apareamiento de oligonucleótidos  $DSL_n$  y  $DSL_m$ .

(d) Ligación de los dúplex de oligonucleótidos  $DSL_n$ - $DSL_m$  al material de ADN con inspección.

(e) Opcional: digestión de oligonucleótidos (DSL/DSS) sin unir usando exonucleasa u otras enzimas específicas de ADN monocatenario (ADN<sub>mc</sub>) o ARNasa H en caso de oligonucleótidos de ADN/ARN.

(f) Amplificación de la(s) representación/representaciones genómica(s) seleccionada(s) como diana usando un cebador universal (en este caso denominado DSPCR), cuya secuencia es idéntica a la región de unión a cebador en 5' del oligonucleótido  $DSL_n$ .

Puede usarse el protocolo completo o partes seleccionadas del mismo para diversos tipos de muestras, por ejemplo, células individuales, una multitud de células individuales, ADN libre de células, ADN exosómico, ejemplares tisulares químicamente fijados (es decir, muestras tisulares fijadas en formalina, incrustadas en parafina), etc.

Ejemplo 3 - Muestras procesadas sin la adición de exonucleasa.

Se han aislado células individuales a partir de sangre periférica de un individuo sano o se han recogido a partir de cultivo de células adherentes de células de cáncer de esófago OE-19. Se escogió una célula individual en 1,0  $\mu$ l de PBS y se transfirió al interior de un tubo de reacción que contenía 2,0  $\mu$ l de tampón de digestión con proteinasa K (Tris-acetato 10 mM, pH 7,5, acetato de Mg 10 mM, acetato de K 50 mM (0,2  $\mu$ l de 10 x tampón One-Phor-All-Buffer-Plus); Tween 20 al 0,67%; Igepal al 0,67%; proteinasa K 0,67 mg/ml). Todas las etapas posteriores del protocolo se realizaron en una máquina de PCR con una tapa calentada. Se realizó la digestión con proteinasa K durante 10 h a 42°C, seguido por una etapa de inactivación a 80°C durante 10 min. A continuación, se sometió ADN de célula

individual a digestión con endonucleasa de restricción MseI (Fermentas) añadiendo 0,2 µl de 10 x tampón One-Phor-All-Buffer-Plus, 10 U de MseI (New England Biolabs) y H<sub>2</sub>O hasta un volumen total de 5,0 µl. La digestión de restricción se realizó durante 3 horas a 37°C y se inactivó térmicamente a 65°C durante 5 min. La preparación de los adaptadores para PCR mediada por ligación se logró mediante apareamiento de oligonucleótidos DSL<sub>n</sub> y DSS<sub>m</sub>. Con este fin, se mezclaron 0,5 µl de disolución madre 100 QM de cada uno de los oligonucleótidos con 1,0 µl de H<sub>2</sub>O. Se inició el apareamiento a 65°C y se continuó a temperatura decreciente de manera continua, con una rampa de 1°C/min, hasta 15°C. Se complementaron los oligonucleótidos apareados con 0,5 µl de oligonucleótido de DSPCR [disolución madre 100 µM], 1 µl de ATP (10 mM) y 1 µl de ADN ligasa de T4 (5 U; Roche). Posteriormente se añadió una mezcla de oligonucleótido/ATP/ligasa previamente mezclada a representación de ADN fragmentada y se ligó durante la noche a 15°C. La reacción de PCR posterior se inició tras añadir 3 µl de tampón de PCR (Expand Long Range Buffer 1, Roche), 2 µl de dNTP (10 mM), 5 U de mezcla de ADN polimerasa Pwo/Taq (PolMix, Expand Long Range Buffer 1, Roche) y H<sub>2</sub>O hasta un volumen total de 50 µl y se ejecutó durante 47 ciclos en una máquina de PCR. En la figura 14 se exponen detalles del procedimiento de ciclación usado en la PCR primaria.

Ejemplo 4 - Preparación de muestras incluyendo tratamiento con exonucleasa.

La inclusión de un código de barras con de 3 a 18 bases aleatorias marca básicamente todas las moléculas de ADN ligadas de una única o unas pocas células de tal manera que cada secuencia resultante es única. Por tanto, el oligonucleótido DSL no puede usarse como oligonucleótido de PCR, porque o bien se pierde el código de barras único o bien la eficiencia de amplificación es mala. Además, el oligonucleótido DSL con código de barras intacto puede afectar negativamente a la reacción de PCR mediante acontecimientos de cebado aleatorios no deseados.

Para prevenir la unión no deseada del oligonucleótido DSL puede añadirse una etapa de exonucleasa tras la reacción de relleno y antes de la amplificación exponencial de la muestra. En esta realización del procedimiento las etapas de digestión con proteinasa K y digestión con MseI permanecieron inalteradas. El apareamiento de las secuencias de oligonucleótido estuvo compuesto por 0,5 µl de 10 x tampón One-Phor-All-Buffer-Plus, 0,5 µl de oligonucleótido DSL<sub>n</sub> y DSS<sub>m</sub> (100 µM cada uno) y 1,5 µl de H<sub>2</sub>O. De manera idéntica a la variación anterior del procedimiento, se inició el apareamiento a 65°C y se continuó a temperatura decreciente de manera continua, con una rampa de 1°C/min, hasta 15°C. Posteriormente, se mezclaron los oligonucleótidos apareados con el producto de la digestión con MseI y se ligaron durante la noche a 15°C. La posterior reacción de PCR primaria se ensambló añadiendo 3,0 µl de tampón de PCR (Expand Long Range Buffer 1, Roche), 2 µl de dNTP (10 mM), 5 U de combinación de ADN polimerasas Pwo/Taq (PolMix, Expand Long Range Buffer 1, Roche) y 34,0 µl de H<sub>2</sub>O. Tras una etapa de relleno (3 min a 68°C), se añadieron 0,5 µl de exonucleasa I (20 U/µl) para eliminar los oligonucleótidos sin unir. Se realizó la digestión con exonucleasa a 37°C durante 30 minutos y se inactivó térmicamente a 85°C durante 15 minutos. A continuación, se añadieron 0,5 µl de oligonucleótido DSPCR (100 µM) y se inició la PCR. En la figura 15 se expone la especificación para el procedimiento de ciclación usado.

Ejemplo 5 - Control de calidad de muestras para secuenciación libre de errores.

Anteriormente se desarrolló un ensayo sustituto que predice la calidad de la amplificación de genoma completo (WGA). Para establecer este ensayo, se usaron una CGH de metafase, CGH en matriz y tasas de abandono alélico para evaluar para cada célula individual la calidad de la WGA. En resumen, con el fin de diseñar una prueba adecuada para evaluar una amplificación de genoma completo fiable y homogénea de ADN de célula individual con *Ampli1*<sup>TM</sup>, se realizó el siguiente experimento: a partir del biobanco de WGA de células individuales existente, se seleccionaron 72 productos de WGA de células cancerosas diseminadas (DCC) individuales aisladas a partir de médula ósea de pacientes con cáncer de mama y con cáncer de próstata, así como a partir de ganglios linfáticos de pacientes con melanoma. A partir de cada uno de los tres tipos de tumor, se seleccionaron 12 DCC que se habían hibridado satisfactoriamente en cromosomas humanos en experimentos de CGH anteriores (n=36), y 3 x 12 DCC que no superaron experimentos de CGH. Se diseñaron ocho pares de oligonucleótidos diferentes para fragmentos de restricción con MseI ubicados en diferentes regiones cromosómicas y con diferente longitud de fragmento, que oscilaba entre 239 pb y 1936 pb. Se realizó PCR específica con ocho pares de oligonucleótidos en todas las DCC seleccionadas y se correlacionaron los resultados con resultado de CGH conocido. Se encontró que tres pares de oligonucleótidos podían predecir un experimento de CGH de metafase satisfactorio con una especificidad del 94% y una sensibilidad del 97%, si un producto de WGA de célula individual era positivo para al menos 2/3 de los marcadores (figura 2).

Se validó el ensayo en una cohorte de 100 células no cancerosas diploides que se habían aislado y se había amplificado su ADN entre 1999 y 2008. Se seleccionaron veintidós productos de WGA de células individuales que se predijo mediante los tres pares de oligonucleótidos seleccionados que permitían el análisis de CGH y 10 productos de WGA de células individuales que se predijo que no lo superaban. El rendimiento de CGH se predijo correctamente en los 32 casos. Posteriormente, se incluyó un cuarto par de oligonucleótidos de PCR ubicado en un fragmento de restricción con MseI de 192 pb de longitud que abarcaba el codón 12/13 mutado con frecuencia del gen *KRAS* y se diseñó un ensayo de PCR de multiplex de 4 marcadores (kit de QC *Ampli1*<sup>TM</sup>) para predecir la integridad del genoma de células aisladas.

Después, se aislaron 88 células mononucleares individuales a partir de sangre periférica de un donante varón usando un micromanipulador controlado manualmente, se amplificó ADN genómico y se evaluó la calidad de amplificación con el kit de QC *Ampli1*<sup>TM</sup> en células recientemente aisladas y sin fijar. Los resultados mostraron que 83/88 (94,3%) de las células presentaban dos o tres de las bandas de QC (figura 3).

El ensayo de QC final asigna un índice de integridad genómica (GII) de 0 si no se amplifica ninguna banda; de 1 si se amplifica una banda; de 2 si se amplifican dos bandas; de 3 si se amplifican tres bandas; de 4, si se amplifican cuatro bandas (figura 3). Se sometieron a prueba los GII en células tumorales circulantes de pacientes aisladas mediante el sistema CellSearch. Se investigaron todas las células con diversos análisis aguas abajo, es decir qPCR, secuenciación de Sanger dirigida y aCGH. De nuevo, el ensayo era perfectamente adecuado para evaluar la calidad de las muestras.

El ensayo de QC basado en PCR multiplexada se usó para determinar la calidad de muestras para la secuenciación libre de errores. Esta reacción evalúa la presencia de tres loci diferentes en el genoma humano. La tasa positiva de la reacción de multiplex se correlaciona con una tasa satisfactoria de la aplicación aguas abajo múltiple, por tanto puede usarse como marcador sustituto para la amplificación de genoma completo satisfactoria. Para evaluar adicionalmente la calidad y el rendimiento de la preparación de muestras, se analizó la tasa de abandono alélico (ADO) resultante del sesgo introducido durante la preparación de muestras. Con este fin, se eligieron cuatro marcadores de SNP diferentes, se sometieron a prueba como heterocigóticos en todos los ejemplares incluidos en la serie de pruebas y se sometieron a prueba para determinar su presencia en la muestra en un ensayo de RFLP-PCR.

Ejemplo 6 - Optimización del procedimiento de preparación de muestras para secuenciación libre de errores.

*Determinación del diseño de oligonucleótido óptimo.*

El cebado no deseado del oligonucleótido DSL depende de la razón de la longitud de secuencia fija en 3' del oligonucleótido DSL y la longitud del código de barras. Cuanto más corta es la secuencia fija en 3' más débil es la unión en 3' de bases crucial para la extensión mediante polimerasa Taq. Cuanto más largo es el código de barras menos probable es la posibilidad de que oligonucleótidos DSL totalmente complementarios se unan a productos de adaptador de ADN durante PCR. Por tanto, códigos de barras cortos pueden funcionar mejor con secuencias fijas cortas, y secuencias fijas largas pueden requerir códigos de barras más largos.

Para someter a prueba cómo influye la longitud del oligonucleótido DSS<sub>m</sub> en el rendimiento de la PCR primaria, se sometieron a prueba dos variantes del oligonucleótido, DSS<sub>5</sub> y DSS<sub>6</sub>, con la longitud de 5 ó 6 bases además de las dos bases que reconstituyen el motivo TTAA, respectivamente.

Se sometieron a prueba dos combinaciones de oligonucleótidos, DSL<sub>12</sub>-DSS<sub>6</sub> y DSL<sub>9</sub>-DSS<sub>5</sub>, en diez células individuales y dos combinaciones de células de un donante varón sano. Se evaluó el rendimiento de ambos dúplex de oligonucleótidos usando la PCR de multiplex (figura 5).

Los resultados del ensayo de QC indican que el adaptador compuesto por DSL<sub>12</sub> y DSS<sub>6</sub> proporciona de manera reproducible productos de PCR de alta calidad, mientras que el uso de DSL<sub>9</sub>-DSS<sub>5</sub> parece ser menos adecuado (figura 5A-B). La amplificación exhaustiva de un genoma de célula individual usando DSL<sub>12</sub>/DSS<sub>6</sub> se confirmó en experimentos de aCGH (figura 5C-D). Por este motivo, se realizaron experimentos adicionales únicamente con la combinación de oligonucleótidos DSL<sub>12</sub>-DSS<sub>6</sub>.

*Determinación de la temperatura óptima para el apareamiento de oligonucleótidos durante la PCR primaria.*

Para optimizar adicionalmente el rendimiento de la PCR mediada por adaptador, se sometieron a prueba las condiciones de ciclación para la PCR primaria. La incorporación de la etiqueta aleatorizada del oligonucleótido DSL<sub>n</sub> dio como resultado cinética de apareamiento variable de las diferentes variantes de oligonucleótidos. Por tanto, una selección apropiada de la condición de apareamiento puede ser crucial para el éxito de la PCR. Para encontrar los ajustes más óptimos, se sometieron a prueba tres temperaturas de apareamiento diferentes: 57°C, 60°C y 63°C. Se procesaron muestras de células individuales con la PCR primaria modificada. El análisis de SNP posterior mostró tasas de abandono alélico comparables independientes de la temperatura de apareamiento usada (figura 6). Esto sugiere que el protocolo permite amplificar completamente un genoma de célula individual con pérdidas alélicas poco frecuentes. Por tanto, la cobertura del genoma parece ser excelente.

Sin embargo, la evaluación de calidad de productos de PCR usando la PCR de multiplex reveló que las temperaturas de apareamiento aumentadas pueden tener un impacto ligeramente negativo sobre la PCR primaria (figura 7). Las muestras mostraron la mejor calidad cuando se usó la temperatura de apareamiento de 57°C durante la PCR primaria (figura 7). Por consiguiente, se usó este ajuste para experimentos adicionales.

Ejemplo 7 - Tratamiento con exonucleasa adicional.

Tal como ya se mencionó, la presencia del oligonucleótido DSL<sub>n</sub> sin unir en la PCR primaria puede dificultar algunas

aplicaciones aguas abajo u ocasionalmente la eficiencia de la reacción de amplificación. Para impedir esto, se sometieron a prueba los efectos de la etapa de digestión con exonucleasa I introducida entre la reacción de relleno y el inicio de la amplificación exponencial de los fragmentos de restricción en la PCR primaria. Usando este enfoque se intentó eliminar el oligonucleótido sin unir antes de proceder con la amplificación basada en PCR de la representación genómica de las muestras. Pruebas iniciales con la exonucleasa I indicaron que la enzima no afecta a la fracción de ADN bicatenario en la reacción de PCR y permite la eliminación de los adaptadores de PCR sin unir (figura 8).

Ejemplo 8 - Prueba directa para la introducción del código de barras en amplicones derivados de células individuales.

Para demostrar que los fragmentos de restricción con MseI se marcaban de hecho con el código de barras, se secuenciaron tres fragmentos de restricción seleccionados aleatoriamente a partir de una muestra de célula individual generada usando el oligonucleótido DSL<sub>12</sub>. Para aislar fragmentos individuales, la representación seleccionada por tamaño (únicamente fragmentos de más de 300 pb) del genoma de célula individual se clonó en un vector pGEM T-Easy. Tras transformar las bacterias *E. coli* con este constructo y la selección basada en X-Gal colorimétrica de las colonias transformadas, se seleccionaron aleatoriamente tres colonias para pruebas adicionales. Tras el aislamiento de ADN de plásmido, la secuenciación posterior reveló secuencias que albergaban secuencias genómicas humanas flanqueadas por el oligonucleótido DSL<sub>12</sub> y sus secuencias complementarias (figura 8). Tal como se esperaba, la secuencia de código de barras aleatorizado difirió para los tres fragmentos, demostrando que el nuevo enfoque permite marcar con etiqueta de manera única los fragmentos de restricción individuales en la representación genómica que se origina a partir de ADN de célula individual (figura 9). Asimismo, pudieron recuperarse la secuencia de etiqueta de paciente fijada y el sitio de I-SceI.

Ejemplo 9 - Uso de oligonucleótidos de ADN/ARN.

También se prevé el uso de un oligonucleótido de ADN/ARN. Un método de este tipo comprende las siguientes etapas:

1) digestión con proteasa de ADN libre de células;

2) restricción, usando por ejemplo MseI;

3) ligación de adaptador, en el que el segundo oligonucleótido es un oligonucleótido de ADN/ARN que comprende una primera secuencia que consiste en ADN, una segunda secuencia que consiste en ARN y una tercera secuencia que consiste en ADN. Por consiguiente, la secuencia aleatoria del segundo oligonucleótido consiste en ARN;

4) adición de transcriptasa inversa + ADN polimerasa (termoestable) y desoxinucleótidos para generar cadenas dobles, mediante lo cual se crea una molécula de ADN:ADN-ARN:ADN-ADN:ADN en el segundo oligonucleótido ligado. Por consiguiente, se crea la cadena de ADN complementaria a la cadena de ARN de secuencia aleatorizada;

5) adición de ARNasa H para digerir híbridos bicatenarios de ADN/ARN;

6) adición de ARNasa If para digerir ARN monocatenario de los adaptadores libres compuestos por el primer y el segundo oligonucleótidos;

7) tras la eliminación de la secuencia de ARN del segundo oligonucleótido ligado, vuelven a unirse las partes de ADN restantes del segundo oligonucleótido mediante la ADN polimerasa;

8) adición de ligasa para unir la parte extendida del segundo oligonucleótido con la región fija del segundo oligonucleótido;

8) adición de reactivos de PCR restantes;

9) adición del tercer oligonucleótido. Si las etapas de ARNasa son eficientes, el cebador de PCR se crea durante la reacción. Por consiguiente, si las etapas de ARNasa son eficientes, no hay necesidad de añadir un tercer oligonucleótido.

Ejemplo 10 - Aplicación a amplificación de genoma completo (WGA).

Se evaluó el rendimiento de WGA generada mediante el método de la presente invención usando un ensayo de PCR de multiplex diseñado específicamente para someter a prueba la calidad de productos de WGA de células individuales; véase Polzer *et al.* (2014) EMBO Mol Med. 30 de octubre; 6(11):1371-86. Esta prueba evalúa la presencia de cuatro loci genéticos (KRAS, D5S 2117, KRT19 y TP53) en los productos de WGA. La presencia de las cuatro secuencias (4 bandas positivas) es indicativa de producto de alta calidad. A la inversa, la falta de cualquier producto en la PCR de control indica una mala calidad de los productos de WGA. Como comparación, se realizó un

protocolo tal como se describe por Jones (1997) BioTechniques 22:938-946 y en el documento US 08/742.755. A continuación, el protocolo descrito en la técnica anterior, concretamente Jones (1997) BioTechniques 22:938-946 y el documento US 08/742.755, se indica como (A) mientras que el experimento que se realizó usando el método de la presente invención se denomina (B).

5

#### *Proporcionar muestra de ADN*

Se comparó el método descrito en Jones (1997) BioTechniques 22:938-946 y el documento US 08/742.755 con el método de la presente invención. Se usaron ambos métodos para amplificar 13 muestras con cantidad variable de material de partida: cinco reacciones con linfocitos de sangre periférica (PBL) individuales procedentes de un donante normal, tres combinaciones de 10 PBL y tres combinaciones de 100 PBL se sometieron a lisis simultáneamente por duplicado en recipientes individuales para liberar el ácido desoxirribonucleico (ADN) genómico bicatenario.

#### *Restricción de ADN*

(A) Según el método descrito en Jones (1997) BioTechniques 22:938-946 y en el documento US 08/742.755, se digirió el ADN aislado con una enzima de restricción cuyo sitio de escisión está separado de su sitio de reconocimiento (en este caso BseRI), creando así moléculas bicatenarias que tenían una secuencia de proyección monocatenaria correspondiente al sitio de restricción de la enzima usada. Tras la restricción, se inactivó la enzima según la información proporcionada por el fabricante. En la configuración experimental, la concentración de BseRI era de ~0,151 U/pg, ~0,015 U/pg y ~0,001 U/pg de ADN para ADN de célula individual, ADN que se originaba a partir de combinación de células de 10 PBL y 100 PBL, respectivamente (1 unidad se define como la cantidad de enzima requerida para digerir 1 µg de ADN λ en 1 hora a 37°C en un volumen de reacción total de 50 µl). Se realizó la digestión de restricción durante 3 horas a 37°C en un volumen de reacción de 5 µl.

(B) En paralelo, se trató el ADN aislado tal como se describe en el presente documento, es decir usando la enzima de restricción MseI a la concentración de ~1,5 U/pg, ~0,15 U/pg y ~0,015 U/pg de ADN para ADN de célula individual, ADN que se originaba a partir de combinación de células de 10 PBL y 100 PBL, respectivamente (1 unidad se define como la cantidad de enzima requerida para digerir 1 µg de ADN λ en 1 hora a 37°C en un volumen de reacción total de 50 µl). Se realizó la digestión de restricción durante 3 horas a 37°C en un volumen de reacción de 5 µl.

#### *Adaptador de PCR*

(A) Se generó un adaptador correspondiente al conjunto de adaptadores 1 (tal como se muestra en la tabla 2 de Jones (1997) BioTechniques 22 (5), 938-946 y se indica en el ejemplo 2 del documento US 08/742.755) tal como se describe en la sección de Materiales y métodos de la publicación de Jones *et al.* con la excepción de que en lugar de usar cuatro adaptadores diferentes, que portaban cada uno un nucleótido fijo en la posición 3'-terminal de la cadena superior, sólo se usó un adaptador presentando la cadena superior dos bases virtuales (N) en la posición 3'-terminal.

#### *Apareamiento de adaptador*

(A) Se generaron adaptadores bicatenarios mediante apareamiento de los oligonucleótidos de cadena superior e inferior según el procedimiento descrito en Jones DH *et al.*, BioTechniques 22 (5), 938-946.

(B) En paralelo, se aparearon el primer y el segundo oligonucleótidos de la invención usando el siguiente protocolo: etapa 1: 30 s a 80°C, seguido por etapa 2: etapa de incubación a 65°C durante 1 min, y etapa 3: enfriamiento hasta 15°C con una temperatura de rampa constante de 1°C/min.

50

#### *Ligación de adaptador*

Para ambos experimentos, se ligaron adaptadores a los productos de digestión con enzima de restricción en presencia de 1 µl de ATP (10 mM) y 5 U de ADN ligasa de T4.

55

#### *Amplificación mediante PCR*

(A) Se amplificaron las moléculas bicatenarias ligadas mediante un cebador específico para el conjunto de adaptadores 1, mediante lo cual la secuencia era homóloga a la secuencia de la cadena superior del adaptador que sirvió como etiqueta de identificación.

60

(B) Se amplificaron las moléculas bicatenarias ligadas mediante un cebador específico para el oligonucleótido 2, mediante lo cual la parte terminal del oligonucleótido 2 sirvió como etiqueta de unión a cebador. Es decir, se amplificó la molécula bicatenaria ligada usando un tercer oligonucleótido tal como se usa en los métodos de la presente invención.

65

*Evaluación*

5 Se evaluó la idoneidad de ambos métodos para WGA mediante el ensayo de PCR de multiplex de QC2 (Polzer *et al* 2014; EMBO Molecular Medicine (2014) 6,1371-1386).

*Resultados*

10 Los resultados negativos del ensayo de PCR de multiplex en todas las muestras amplificadas mediante el método descrito en Jones DH, BioTechniques 22:938-946 y el documento US 08/742.755 indican que este enfoque no es adecuado para la amplificación de genoma completo y por consiguiente no puede dar como resultado la secuenciación libre de errores de un genoma entero. Por tanto, la aplicación del método descrito en los mismos se limita únicamente a la secuenciación de loci genómicos seleccionados y previamente amplificados (por ejemplo mediante PCR) y no para el análisis de genomas enteros.

15 En cambio, los resultados positivos del ensayo de multiplex obtenidos cuando se aplica a productos de WGA generados usando el método de la presente invención muestran que esta tecnología es adecuada para la amplificación de genoma completo de representación de secuencia genómica entera. Además, el método de la presente invención puede usarse posteriormente para la secuenciación libre de errores del genoma entero. Tal como se muestra en la figura 18, el método de la presente invención permite al experto amplificar ADN y posteriormente recuperar la información de secuencia a partir de muestras, mediante lo cual sólo están presentes bajas cantidades de material de partida. En particular, los métodos de la presente invención pueden amplificar el ADN de células individuales cuya información de secuencia puede recuperarse entonces usando los métodos tal como se proporcionan en el presente documento.

25 Ejemplo 11 - Secuenciación libre de errores del ADN de células individuales.

30 Con el fin de demostrar la viabilidad de la corrección de errores usando una secuencia aleatorizada como código de barras/identificador como parte del segundo oligonucleótido usado en los métodos de la presente invención, se realizaron dos experimentos. Para ellos, se ligaron adaptadores que contenían secuencia aleatorizada a ~6 pg de ADN digerido con MseI o bien procedente de una célula individual o bien procedente de cromosoma 22 humano clasificado mediante FACS. Posteriormente se amplificaron fragmentos de MseI usando *Ampli1*<sup>TM</sup> y posteriormente se secuenciaron en la plataforma Roche GS 454 FLX+.

35 *Configuración experimental*

- (a) Se extrajo ADN a partir de una célula individual;
- (b) se digirió la muestra de ADN usando MseI como enzima de restricción;
- (c) se apareó un primer oligonucleótido a la proyección generada. Posteriormente, se apareó un segundo oligonucleótido que comprendía una secuencia aleatorizada al primer oligonucleótido;
- (d) se ligó el segundo oligonucleótido a los extremos de fragmentos de MseI;
- (e) se rellenaron las proyecciones generadas mediante el segundo oligonucleótido ligado;
- (f) se amplificaron los fragmentos usando cebadores de PCR complementarios a una tercera secuencia del segundo oligonucleótido;
- (f) dado que el segundo oligonucleótido comprendía un sitio de escisión de una endonucleasa de guiado, es decir SclI, se cortaron los fragmentos amplificados tras la amplificación;
- (g) se repararon los extremos de los fragmentos cortados y se ligaron mediante T/A al adaptador en Y proporcionado en el kit Rapid Library Prep de Roche Diagnostics. Se realizó la secuenciación en la plataforma GS 454 FLX+ tal como se describe en los manuales de métodos de sistema de secuenciación 454 de la serie XLR70.

*Secuenciación*

60 Aunque el método se realizó usando la plataforma 454FLX, el experto apreciará que pueden usarse métodos de secuenciación alternativos.

*Análisis*

65 El análisis de los fragmentos de ADN secuenciado requirió la identificación de la secuencia aleatorizada y posterior recorte de adaptador. Ambos se realizaron usando un programa de JAVA interno.

En resumen, se identificaron secuencias del segundo oligonucleótido dentro de lecturas basándose en una implementación de BioJava modificada del algoritmo de Smith-Waterman; se escribieron secuencias aleatorizadas en la cabecera de lectura y se cortaron secuencias de adaptador mientras se conservaba el sitio de restricción de Msel ("TTAA"). Se descartaron lecturas en las que no podía identificarse el adaptador y por tanto tampoco la secuencia aleatorizada (1 - 3,5% de todas las lecturas). Además, se descartaron lecturas con una longitud de menos de 20 pb tras el recorte de adaptador (0,7 - 4,3% de todas las lecturas en las que se identificó un adaptador). Un examen para seleccionar secuencias de lectura contaminantes a partir de especies no humanas no reveló ninguna contaminación y por tanto se mapearon directamente las lecturas recortadas contra el conjunto de genoma humano 19 usando el algoritmo de BWA-MEM con ajustes convencionales (Li y Durbin, Bioinformatics, 15 de julio de 2009; 25(14):1754-60). Con el fin de ilustrar la capacidad para distinguir entre errores de secuenciación y mutaciones reales/SNP usando el método de la presente invención, se identificaron las regiones genómicas con una cobertura de lectura de al menos 12. Después se filtraron adicionalmente para detectar regiones en las que se produjeron al menos 8 lecturas con secuencias aleatorizadas idénticas. A partir del conjunto de datos resultante, se seleccionaron aleatoriamente 4 regiones que contenían supuestos SNP/mutaciones/errores de secuenciación. Para las alineaciones mostradas, se usaron todas las lecturas que contenían la misma secuencia aleatorizada.

*Resultados*

Usar las secuencias aleatorizadas de lecturas alineadas a la misma posición del genoma de referencia permitió diferenciar de manera fiable entre errores de secuenciación/amplificación, polimorfismos de un solo nucleótido y mutaciones reales. En cada grupo de lecturas que compartían la misma secuencia aleatorizada, se encontró que las variaciones de un solo nucleótido con respecto a la referencia eran específicas para un grupo de secuencias aleatorizadas (figura 19a-d).

**Lista de secuencias**

<110> Fraunhofer-Gesellschaft zur Förderung der angew. Forschung  
 <120> Secuenciación libre de errores de ADN  
 <130> Documento W3072 PCT S3  
 <150> Documento EP14 15 4044.3  
 <151> 05-02-2014  
 <160> 139  
 <170> BiSSAP 1.2  
 <210> 1  
 <211> 40  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 <222> 1..40  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="Cebador"  
 / organismo="Secuencia artificial"  
 <220>  
 <221> variación  
 <222> 23..34  
 <223> /nota="n = a, t, g, c o u"  
 <400> 1  
**gctagggata acagggtaat gcnnnnnnnn nnnngtcagt**

40

<210> 2  
 <211> 22  
 <212> ADN  
 <213> Secuencia artificial

65

	<220>	
	<221> fuente	
	<222> 1..22	
5	<223> /tipo de mol.=“ADN sin asignar” / nota=“Cebador” / organismo=“Secuencia artificial”	
	<400> 2	
10	<b>gctagggata acagggtaat gc</b>	<b>22</b>
	<210> 3	
	<211> 18	
	<212> ADN	
15	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..18	
20	<223> /tipo de mol.=“ADN sin asignar” / nota=“Cebador” / organismo=“Secuencia artificial”	
	<400> 3	
25	<b>tagggataac agggtaat</b>	<b>18</b>
	<210> 4	
	<211> 22	
	<212> ADN	
30	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..22	
35	<223> /tipo de mol.=“ADN sin asignar” / nota=“Cebador” / organismo=“Secuencia artificial”	
	<400> 4	
40	<b>gctagggata acagggtaat gc</b>	<b>22</b>
	<210> 5	
	<211> 40	
	<212> ADN	
45	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..40	
50	<223> /tipo de mol.=“ADN sin asignar” / nota=“Cebador” / organismo=“Secuencia artificial”	
	<220>	
55	<221> variación	
	<222> 23..34	
	<223> /nota=“n = a, t, g, c o u”	
	<400> 5	
60	<b>gctagggata acagggtaat gcnnnnnnnn nnnngtcagt</b>	<b>40</b>
	<210> 6	
	<211> 36	
	<212> ADN	

<213> Secuencia artificial

<220>  
 <221> fuente  
 5 <222> 1..36  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“Cebador”  
 / organismo=“Secuencia artificial”

10 <220>  
 <221> variación  
 <222> 23..31  
 <223> /nota=“n =a, t, g, c o u”

15 <400> 6  
**gctagggata acagggtaat gcnnnnnnnn ngtcgt** 36

<210> 7  
 <211> 435  
 20 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 25 <222> 1..435  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“Anotación del fragmento: chr10, de 1811253 a 1811607”  
 / organismo=“Secuencia artificial”

30 <400> 7  
**gctagggata acagggtaat gcgtggtcag gggggtcagt taaagtagtt ttggcaaatt** 60  
**ggataaagag tcaagacca tcagtgtgct gtattcagga aaccatctc atgtgcagag** 120  
**acacacatag gctcaaaata aaaggatgga ggaagatcta ccaagcaaat ggaaaacaaa** 180  
**aaaaggcag ggttgcaatc ctagtctctg ataaaacaga ctttatacca acaaagatca** 240  
**aaagagacaa agaaggccat tacataatgg taaagagatc aattcaacaa gaagagctaa** 300  
**ctatcctaaa tatatatgca cccaatacag gagcaccag attcataaag caagtcctga** 360  
**gtgacctaca aagagactta gactcccaga cattaactga cactctcgtc cttgcattac** 420  
**cctgttatcc ctagc** 435

<210> 8  
 <211> 543  
 35 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 40 <222> 1..543  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“fragmento: chr5, de 96535437 a 96525889”  
 / organismo=“Secuencia artificial”

45 <400> 8

ES 2 730 962 T3

	gctagggata acagggtaat gcgcgaaggt aggcgtcagt taaccaagaa aagaggagag	60
	aaaatacaaaa taagctcaat aaggaacaaa acaggagaca ttacaactga cactatataa	120
	atacaaaaaga tcattcaagg ctactgtgaa cacctttaca tgcataaact agaaaaccta	180
	aaagacatgg ataaattcct ggaaagtaga gaggcacac attacccaac ttcaaacat	240
	actataaggc tatagtcacc aactcagcat ggcaactggta taaaaatagg catgtagacc	300
	aatgaaacag aaaagagAAC ccagacataa agccaaatac agccaaccaa tctttgacaa	360
	agcaaacaaa aacataaagt ggggaaagga cgctctattc aacaaatggg gctgggataa	420
	ttggcaagcc acatgtagaa gaatgaaact ggatcctcat ctccccttta tacaaaatca	480
	acacaaaata gataaaagac ttaactgacc cacaaggata ggcattacc tgttatccct	540
	agc	543
5	<210> 9 <211> 440 <212> ADN <213> Secuencia artificial	
10	<220> <221> fuente <222> 1..440 <223> /tipo de mol.=“ADN sin asignar” / nota=“restricción fragmento: chrX, de 94553619 a 94553978” / organismo=“Secuencia artificial”	
15	<400> 9 gctagggata acagggtaat gcgtaggaga aatgtcagt taaatctata aattgtgtgc	60
	cacgtgcggt ggctcatgac tgtaatccca gcactttggg agtccgaggt gggcagatca	120
	cgaggtcagg agattgagac catcttggt aacacgggta aaccctgtct ctactaaaaa	180
	tacaaaaaca aaattagccg ggtgaggtgg caggcacctg tagtcccagc tactccagag	240
	gctgaggtga gagaatggcg tgaacctggg aggcagagct tgcagtgagc tgagattgtg	300
	ccactgcact ccagtctggg tgacagagca agactctgtc tcaaaaaaaaa aaaaaaaaaa	360
	aaaaattcta taaattgtgt tgggcagtat ggccttttta actgacacta aagtcacagc	420
	attaccctgt tatccctagc	440
20	<210> 10 <211> 18 <212> ADN <213> Secuencia artificial	
25	<220> <221> fuente <222> 1..18 <223> /tipo de mol.=“ADN sin asignar” / nota=“Cebador” / organismo=“Secuencia artificial”	
30	<400> 10 tagggataac agggtaat	

<210> 11  
 <211> 27  
 <212> ADN  
 <213> Secuencia artificial  
 5  
 <220>  
 <221> fuente  
 <222> 1..27  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“Cebador”  
 / organismo=“Secuencia artificial”  
 10  
 <400> 11  
**taactataac ggtcctaagg tagcgaa** 27  
 15  
 <210> 12  
 <211> 46  
 <212> ADN  
 <213> Secuencia artificial  
 20  
 <220>  
 <221> fuente  
 <222> 1..46  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“cebador”  
 / organismo=“Secuencia artificial”  
 25  
 <220>  
 <221> variación  
 <222> 28..40  
 <223> /nota=“n = a, c, g, t o u”  
 30  
 <400> 12  
**agttacgcta gggataacag ggtaatgcnn nnnnnnnnnn gtcagt** 46  
 35  
 <210> 13  
 <211> 23  
 <212> ADN  
 <213> Secuencia artificial  
 40  
 <220>  
 <221> fuente  
 <222> 1..23  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“cebador”  
 / organismo=“Secuencia artificial”  
 45  
 <400> 13  
**cgcagttacg ctagggataa cag** 23  
 50  
 <210> 14  
 <211> 50  
 <212> ADN  
 <213> Secuencia artificial  
 55  
 <220>  
 <221> fuente  
 <222> 1..50  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“cebador”  
 / organismo=“Secuencia artificial”  
 60  
 <220>  
 <221> variación

ES 2 730 962 T3

<222> 32..43  
 <223> /nota="n = c, t, g, a o u"

<400> 14  
 5 **cgcgattacg ctagggataa cagggtaatg cnnnnnnnnn nnngatcagt** 50

<210> 15  
 <211> 91  
 <212> ADN  
 10 <213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..91  
 15 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01EV3C9"  
 / organismo="Secuencia artificial"

<400> 15  
**gaagagggga gtgggaagta gaggctgagt gatttaggta tttccaataa tcaatagcaa** 60

20 **gaccactgtg agctgaatac atcgggttat c** 91

<210> 16  
 <211> 90  
 <212> ADN  
 25 <213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..90  
 30 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01D9VJ6"  
 / organismo="Secuencia artificial"

<400> 16  
**gaagagggga gtgggaagtg aggctgagtg atttaggtat ttccaataat caatagcaag** 60

35 **accactgtga gctgaataca tcgggttatc** 90

<210> 17  
 <211> 91  
 <212> ADN  
 40 <213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..91  
 45 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01BDRUD"  
 / organismo="Secuencia artificial"

50 <400> 17  
**gaagaggggg agtgggaagt gaggctgagt gatttaggta tttccaataa tcaatagcaa** 60

**gaccactgtg agctgaatac atcgggttat c** 91

<210> 18  
 <211> 89  
 55 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente

<222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01DC2CL"  
 / organismo="Secuencia artificial"

5

<400> 18  
**atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatattc aatagcaaga** 60  
  
**ccactgtgag ctgaatacat caggttattc** 89

10

<210> 19  
 <211> 89  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <221> fuente  
 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01AV5Z7"  
 / organismo="Secuencia artificial"

20

<400> 19  
**atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatattc aatagcaaga** 60  
  
**ccactgtgag ctgaatacat caggttattc** 89

25

<210> 20  
 <211> 89  
 <212> ADN  
 <213> Secuencia artificial

30

<220>  
 <221> fuente  
 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01BN7XB"  
 / organismo="Secuencia artificial"

35

<400> 20  
**atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatattc aatagcaaga** 60  
  
**ccactgtgag ctgaatacat caggttattc** 89

40

<210> 21  
 <211> 89  
 <212> ADN  
 <213> Secuencia artificial

45

<220>  
 <221> fuente  
 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01AH386"  
 / organismo="Secuencia artificial"

50

<400> 21  
**atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatattc aatagcaaga** 60  
  
**ccactgtgag ctgaatacat caggttattc** 89

55

<210> 22  
 <211> 89  
 <212> ADN

<213> Secuencia artificial

<220>  
 <221> fuente  
 5 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01D63DR"  
 / organismo="Secuencia artificial"

10 <400> 22  
 atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatatac aatagcaaga 60  
 ccactgtgag ctgaatacat caggttatac 89

<210> 23  
 <211> 89  
 15 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 20 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01BVTHT"  
 / organismo="Secuencia artificial"

25 <400> 23  
 atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatatac aatagcaaga 60  
 ccactgtgag ctgaatacat caggttatac 89

<210> 24  
 <211> 89  
 30 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 35 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01BI7HX"  
 / organismo="Secuencia artificial"

40 <400> 24  
 atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatatac aatagcaaga 60  
 ccactgtgag ctgaatacat caggttatac 89

<210> 25  
 <211> 89  
 45 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 50 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H02IESZE"  
 / organismo="Secuencia artificial"

55 <400> 25  
 atgagtggga tatgggaagt gaggctgagt gatttaggta tttccatatac aatagcaaga 60  
 ccactgtgag ctgaatacat caggttatac 89

ES 2 730 962 T3

<210> 26  
 <211> 89  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..89  
 10 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“IQ4WJ2H02G69J5”  
     / organismo=“Secuencia artificial”  
  
 <400> 26  
 atgagtgggga tatgggaagt gaggctgagt gatttaggta tttccatata aatagcaaga 60  
  
 15 ccactgtgag ctgaatacat caggttatac 89  
  
 <210> 27  
 <211> 67  
 <212> ADN  
 20 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..67  
 25 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02I8EWU”  
     / organismo=“Secuencia artificial”  
  
 <400> 27  
 gataaatatg aacactgcac cgtgaaattc tgtgcctcga agacaggctc gctatttttag 60  
  
 30 aagaggg 67  
  
 <210> 28  
 <211> 70  
 <212> ADN  
 35 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..70  
 40 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02HST0F”  
     / organismo=“Secuencia artificial”  
  
 <400> 28  
 gataaatatg aacactgcac cgtgaaattc tgtgcctccg aagacaggct cgctattttt 60  
  
 45 agaaagaggg 70  
  
 <210> 29  
 <211> 70  
 <212> ADN  
 50 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..70  
 55 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02IS3DB”  
     / organismo=“Secuencia artificial”  
  
 <400> 29

ES 2 730 962 T3

gataaatatg aacactgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60  
 agaaagaggg 70  
 <210> 30  
 <211> 70  
 5 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 10 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02GVA1A”  
 / organismo=“Secuencia artificial”  
 15 <400> 30  
 gataaatatg aacactgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60  
 agaaagaggg 70  
 <210> 31  
 <211> 70  
 20 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 25 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02GV56X”  
 / organismo=“Secuencia artificial”  
 30 <400> 31  
 gataaatatg aacactgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60  
 agaaagaggg 70  
 <210> 32  
 <211> 70  
 35 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 40 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02HT9UP”  
 / organismo=“Secuencia artificial”  
 45 <400> 32  
 gataaatatg aacactgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60  
 agaaagaggg 70  
 <210> 33  
 <211> 70  
 50 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 55 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”

ES 2 730 962 T3

/ nota="I935FME02IPW2U"  
 / organismo="Secuencia artificial"

<400> 33  
 gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60

5 agaaagaggg 70

<210> 34  
 <211> 70  
 <212> ADN  
 10 <213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..70  
 15 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME02HNKYA"  
 / organismo="Secuencia artificial"

<400> 34  
 gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60

20 agaaagaggg 70

<210> 35  
 <211> 70  
 <212> ADN  
 25 <213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..70  
 30 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME02I0S0E"  
 / organismo="Secuencia artificial"

<400> 35  
 gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60

35 agaaagaggg 70

<210> 36  
 <211> 70  
 <212> ADN  
 40 <213> Secuencia artificial

<220>  
 <221> fuente  
 <222> 1..70  
 45 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME02HXEG8"  
 / organismo="Secuencia artificial"

<400> 36  
 gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT 60

50 agaaagaggg 70

<210> 37  
 <211> 70  
 <212> ADN  
 55 <213> Secuencia artificial

<220>

ES 2 730 962 T3

<221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02JFTGT”  
 5 / organismo=“Secuencia artificial”

<400> 37  
**gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT** 60  
**agaaagaggg** 70

10 <210> 38  
 <211> 70  
 <212> ADN  
 <213> Secuencia artificial

15 <220>  
 <221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02GBSUA”  
 20 / organismo=“Secuencia artificial”

<400> 38  
**gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT** 60  
**agaaagaggg** 70

25 <210> 39  
 <211> 70  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02GR861”  
 35 / organismo=“Secuencia artificial”

<400> 39  
**gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT** 60  
**agaaagaggg** 70

40 <210> 40  
 <211> 70  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02IPFUP”  
 50 / organismo=“Secuencia artificial”

<400> 40  
**gataaatatg aacctgcac cgtgaaattc tgtgcctccg aagacaggct cgctatTTTT** 60  
**agaaagaggg** 70

55 <210> 41  
 <211> 70

ES 2 730 962 T3

<212> ADN  
 <213> Secuencia artificial

5 <220>  
 <221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01C1GTK”  
 / organismo=“Secuencia artificial”

10 <400> 41  
 agacgtttta agcactgcac tgtgaaattc tgtgcctccg aagacaggct cgctatntttt 60  
 agaaagaggg 70

15 <210> 42  
 <211> 69  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <222> 1..69  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01CN901”  
 / organismo=“Secuencia artificial”

25 <400> 42  
 agacgnttaa gcaactgcaact gtgaaattct gtgcctccga agacaggctc gctatntttta 60  
 gaaagaggg 69

30 <210> 43  
 <211> 70  
 <212> ADN  
 <213> Secuencia artificial

35 <220>  
 <221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01EAK2C”  
 / organismo=“Secuencia artificial”

40 <400> 43  
 agacgnttta agcactgcac tgtgaaattc tgtgcctccg aagacaggct cgctatntttt 60  
 agaaagaggg 70

45 <210> 44  
 <211> 70  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <221> fuente  
 <222> 1..70  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01AXYNN”  
 / organismo=“Secuencia artificial”

55 <400> 44  
 taaataggggt ttcactgcac tgtgaaattc tgtgcctccg aagacaggct cgctatntttt 60  
 agaaagaggg 70

<210> 45  
 <211> 43  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..43  
 10 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02G3JAX”  
     / organismo=“Secuencia artificial”  
  
 <400> 45  
 15 **agcggggggtt ggaggtaagg tggatgataa cgggtttccc aaa** 43  
  
 <210> 46  
 <211> 72  
 <212> ADN  
 20 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..72  
 25 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02HPZCG”  
     / organismo=“Secuencia artificial”  
  
 <400> 46  
**agcggggggtt ggaggtaagg tggatgataa cgggtttccca aaacaggtga acggacgtac** 60  
  
 30 **gggagtaggg ac** 72  
  
 <210> 47  
 <211> 45  
 <212> ADN  
 35 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..45  
 40 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02HEGFO”  
     / organismo=“Secuencia artificial”  
  
 <400> 47  
 45 **agcggggggtt ggaggtaagg tggatgataa cgggtttccca aaaca** 45  
  
 <210> 48  
 <211> 64  
 <212> ADN  
 50 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..64  
 55 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02FY2OL”  
     / organismo=“Secuencia artificial”  
  
 <400> 48  
**agcggggggtt ggaggtaagg tggatgataa cgggtttccca aaacaggtga acggacgtac** 60  
  
 60 **ggga** 64

ES 2 730 962 T3

<210> 49  
 <211> 64  
 <212> ADN  
 5 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..64  
 10 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02FUAIM”  
     / organismo=“Secuencia artificial”  
  
 <400> 49  
 agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac 60  
  
 15 gggga 64  
  
 <210> 50  
 <211> 72  
 <212> ADN  
 20 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..72  
 25 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02HNVDR”  
     / organismo=“Secuencia artificial”  
  
 <400> 50  
 agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac 60  
  
 30 gggagtaggg ac 72  
  
 <210> 51  
 <211> 72  
 <212> ADN  
 35 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..72  
 40 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02GP6YE”  
     / organismo=“Secuencia artificial”  
  
 <400> 51  
 agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac 60  
  
 45 gggagtaggg ac 72  
  
 <210> 52  
 <211> 72  
 <212> ADN  
 50 <213> Secuencia artificial  
  
 <220>  
 <221> fuente  
 <222> 1..72  
 55 <223> /tipo de mol.=“ADN sin asignar”  
     / nota=“I935FME02IAAW2”  
     / organismo=“Secuencia artificial”  
  
 <400> 52

	<b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
	<210> 53 <211> 72	
5	<212> ADN <213> Secuencia artificial	
	<220> <221> fuente	
10	<222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02I4UCC” / organismo=“Secuencia artificial”	
15	<400> 53 <b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
	<210> 54 <211> 64	
20	<212> ADN <213> Secuencia artificial	
	<220> <221> fuente	
25	<222> 1..64 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02FVCO0” / organismo=“Secuencia artificial”	
30	<400> 54 <b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac</b>	<b>60</b>
	<b>ggga</b>	<b>64</b>
	<210> 55 <211> 45	
35	<212> ADN <213> Secuencia artificial	
	<220> <221> fuente	
40	<222> 1..45 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02JTEOX” / organismo=“Secuencia artificial”	
45	<400> 55 <b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aaaca</b>	<b>45</b>
	<210> 56 <211> 42	
50	<212> ADN <213> Secuencia artificial	
	<220> <221> fuente	
55	<222> 1..42 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02GYCTS” / organismo=“Secuencia artificial”	

	<400> 56		
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aa		42
5	<210> 57 <211> 67 <212> ADN <213> Secuencia artificial		
10	<220> <221> fuente <222> 1..67 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02HYTON” / organismo=“Secuencia artificial”		
15	<400> 57		
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aacaggtga acggacgtac		60
	gggagta		67
20	<210> 58 <211> 42 <212> ADN <213> Secuencia artificial		
25	<220> <221> fuente <222> 1..42 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02JBAVG” / organismo=“Secuencia artificial”		
30	<400> 58		
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aa		42
35	<210> 59 <211> 72 <212> ADN <213> Secuencia artificial		
40	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02FX9GA” / organismo=“Secuencia artificial”		
45	<400> 59		
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aacaggtga acggacgtac		60
	gggagtaggg ac		72
50	<210> 60 <211> 42 <212> ADN <213> Secuencia artificial		
55	<220> <221> fuente <222> 1..42 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02GYILN” / organismo=“Secuencia artificial”		

	<400> 60	
	<b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aa</b>	<b>42</b>
5	<210> 61 <211> 72 <212> ADN <213> Secuencia artificial	
10	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02ID4XH”	
15	/ organismo=“Secuencia artificial”	
	<400> 61	
	<b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aacaggtga acggacgtac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
20	<210> 62 <211> 71 <212> ADN <213> Secuencia artificial	
25	<220> <221> fuente <222> 1..71 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02HVJJE”	
30	/ organismo=“Secuencia artificial”	
	<400> 62	
	<b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aacaggtga acggacgtac</b>	<b>60</b>
	<b>gggagtaggg a</b>	<b>71</b>
35	<210> 63 <211> 64 <212> ADN <213> Secuencia artificial	
40	<220> <221> fuente <222> 1..64 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02H32G2”	
45	/ organismo=“Secuencia artificial”	
	<400> 63	
	<b>agcggggggtt ggaggtaagg tggtgataac gggtttccca aacaggtga acggacgtac</b>	<b>60</b>
	<b>ggga</b>	<b>64</b>
50	<210> 64 <211> 72 <212> ADN <213> Secuencia artificial	
55	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02F6Q0H”	

ES 2 730 962 T3

	/ organismo="Secuencia artificial"	
	<400> 64	
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac	60
5	gggagtaggg ac	72
	<210> 65	
	<211> 67	
	<212> ADN	
	<213> Secuencia artificial	
10	<220>	
	<221> fuente	
	<222> 1..67	
	<223> /tipo de mol.="ADN sin asignar"	
15	/ nota="I935FME02HNHHY"	
	/ organismo="Secuencia artificial"	
	<400> 65	
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac	60
20	gggagta	67
	<210> 66	
	<211> 72	
	<212> ADN	
	<213> Secuencia artificial	
25	<220>	
	<221> fuente	
	<222> 1..72	
	<223> /tipo de mol.="ADN sin asignar"	
30	/ nota="I935FME02FOIUY"	
	/ organismo="Secuencia artificial"	
	<400> 66	
	agcggggggtt ggaggtaagg tggtgataac gggtttccca aaacaggtga acggacgtac	60
35	gggagtaggg ac	72
	<210> 67	
	<211> 42	
	<212> ADN	
	<213> Secuencia artificial	
40	<220>	
	<221> fuente	
	<222> 1..42	
	<223> /tipo de mol.="ADN sin asignar"	
45	/ nota="I935FME02GSN7O"	
	/ organismo="Secuencia artificial"	
	<400> 67	
50	agcggggggtt ggaggtaagg tggtgataac gggtttccca aa	42
	<210> 68	
	<211> 39	
	<212> ADN	
	<213> Secuencia artificial	
55	<220>	
	<221> fuente	
	<222> 1..39	
	<223> /tipo de mol.="ADN sin asignar"	

ES 2 730 962 T3

/ nota="I935FME02FR6XR"  
/ organismo="Secuencia artificial"

	<400> 68	
5	<b>agcggggggtt ggaggttaagg tggtgataac gggtttccc</b>	<b>39</b>
	<210> 69	
	<211> 71	
	<212> ADN	
10	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..71	
15	<223> /tipo de mol.="ADN sin asignar" / nota="I935FME02IP2KO" / organismo="Secuencia artificial"	
	<400> 69	
	<b>agcgggggttg gaggttaaggt ggtgataacg ggtttcccaa aacaggtgaa cggacgtacg</b>	<b>60</b>
20	<b>ggagtaggga c</b>	<b>71</b>
	<210> 70	
	<211> 71	
	<212> ADN	
25	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..71	
30	<223> /tipo de mol.="ADN sin asignar" / nota="I935FME02JKE75" / organismo="Secuencia artificial"	
	<400> 70	
	<b>agcgggggttg gaggttaaggt ggtgataacg ggtttcccaa aacaggtgaa cggacgtacg</b>	<b>60</b>
35	<b>ggagtaggga c</b>	<b>71</b>
	<210> 71	
	<211> 72	
	<212> ADN	
40	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..72	
45	<223> /tipo de mol.="ADN sin asignar" / nota="I935FME01AGM6X" / organismo="Secuencia artificial"	
	<400> 71	
	<b>atagactagg cgaggttaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>	<b>60</b>
50	<b>gggagtaggg ac</b>	<b>72</b>
	<210> 72	
	<211> 60	
	<212> ADN	
55	<213> Secuencia artificial	
	<220>	
	<221> fuente	
	<222> 1..60	

<223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01AMWUD"  
 / organismo="Secuencia artificial"

5 <400> 72  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

<210> 73  
 <211> 72  
 10 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 15 <222> 1..72  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01EEHFG"  
 / organismo="Secuencia artificial"

20 <400> 73  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

**gggagtaggg ac** 72

<210> 74  
 <211> 71  
 25 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 30 <222> 1..71  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01CP0EN"  
 / organismo="Secuencia artificial"

35 <400> 74  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

**gggatagggga c** 71

<210> 75  
 <211> 72  
 40 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 45 <222> 1..72  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01CAYQ8"  
 / organismo="Secuencia artificial"

50 <400> 75  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

**gggagtaggg ac** 72

<210> 76  
 55 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

ES 2 730 962 T3

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01CT2IH”  
 / organismo=“Secuencia artificial”

5

<400> 76  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

gggagtaggg ac 72

10

<210> 77  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BLNIO”  
 / organismo=“Secuencia artificial”

20

<400> 77  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

gggagtaggg ac 72

25

<210> 78  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

30

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01EW5VF”  
 / organismo=“Secuencia artificial”

35

<400> 78  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

gggagtaggg ac 72

40

<210> 79  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

45

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BR9EW”  
 / organismo=“Secuencia artificial”

50

<400> 79  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60

gggagtaggg ac 72

55

<210> 80  
 <211> 71

<212> ADN  
 <213> Secuencia artificial

<220>  
 5 <221> fuente  
 <222> 1..71  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01D9HRR”  
 / organismo=“Secuencia artificial”

10 <400> 80  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac 60**

**gggatagggga c 71**

<210> 81  
 15 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 20 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DYFB5”  
 / organismo=“Secuencia artificial”

25 <400> 81  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac 60**

**gggagtaggg ac 72**

<210> 82  
 30 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 35 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01D6FUV”  
 / organismo=“Secuencia artificial”

40 <400> 82  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac 60**

**gggagtaggg ac 72**

<210> 83  
 45 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

<220>  
 50 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BB15Y”  
 / organismo=“Secuencia artificial”

55 <400> 83

	<b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
5	<210> 84 <211> 72 <212> ADN <213> Secuencia artificial	
10	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02IGF97” / organismo=“Secuencia artificial”	
15	<400> 84 <b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
20	<210> 85 <211> 72 <212> ADN <213> Secuencia artificial	
25	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02FLWO3” / organismo=“Secuencia artificial”	
30	<400> 85 <b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
35	<210> 86 <211> 72 <212> ADN <213> Secuencia artificial	
40	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02FOXSW” / organismo=“Secuencia artificial”	
45	<400> 86 <b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>	<b>60</b>
	<b>gggagtaggg ac</b>	<b>72</b>
50	<210> 87 <211> 72 <212> ADN <213> Secuencia artificial	
55	<220> <221> fuente <222> 1..72 <223> /tipo de mol.=“ADN sin asignar” / nota=“I935FME02GUX6V”	

/ organismo="Secuencia artificial"

	<400> 87		
	<b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatgc</b>		<b>60</b>
5	<b>gggagtaggg ac</b>		<b>72</b>
	<210> 88		
	<211> 72		
	<212> ADN		
	<213> Secuencia artificial		
10	<220>		
	<221> fuente		
	<222> 1..72		
	<223> /tipo de mol.="ADN sin asignar"		
15	/ nota="I935FME02G14Z4"		
	/ organismo="Secuencia artificial"		
	<400> 88		
	<b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>		<b>60</b>
20	<b>gggagtaggg ac</b>		<b>72</b>
	<210> 89		
	<211> 72		
	<212> ADN		
	<213> Secuencia artificial		
25	<220>		
	<221> fuente		
	<222> 1..72		
	<223> /tipo de mol.="ADN sin asignar"		
30	/ nota="I935FME02H2M7T"		
	/ organismo="Secuencia artificial"		
	<400> 89		
	<b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>		<b>60</b>
35	<b>gggagtaggg ac</b>		<b>72</b>
	<210> 90		
	<211> 72		
	<212> ADN		
40	<213> Secuencia artificial		
	<220>		
	<221> fuente		
	<222> 1..72		
45	<223> /tipo de mol.="ADN sin asignar"		
	/ nota="I935FME02JHRO7"		
	/ organismo="Secuencia artificial"		
	<400> 90		
	<b>atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac</b>		<b>60</b>
50	<b>gggagtaggg ac</b>		<b>72</b>
	<210> 91		
	<211> 72		
	<212> ADN		
55	<213> Secuencia artificial		
	<220>		

ES 2 730 962 T3

<221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DYW8K”  
 / organismo=“Secuencia artificial”

5

<400> 91  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60  
**gggagtaggg ac** 72

10

<210> 92  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

15

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BFXU5”  
 / organismo=“Secuencia artificial”

20

<400> 92  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60  
**gggagtaggg ac** 72

25

<210> 93  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

30

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02IQWMO”  
 / organismo=“Secuencia artificial”

35

<400> 93  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60  
**gggagtaggg ac** 72

40

<210> 94  
 <211> 72  
 <212> ADN  
 <213> Secuencia artificial

45

<220>  
 <221> fuente  
 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02HQ894”  
 / organismo=“Secuencia artificial”

50

<400> 94  
**atagactagg cgaggtaagg tggtgacaac gggtttccca aaacaggtga acggacatac** 60  
**gggagtaggg ac** 72

55

<210> 95  
 <211> 72  
 <212> ADN

<213> Secuencia artificial

<220>  
 <221> fuente  
 5 <222> 1..72  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME02JROVV”  
 / organismo=“Secuencia artificial”

10 <400> 95  
**atagactagg cgaggtaagg tggtagacaac gggtttccca gaacaggtga acggacatac 60**  
**gggagtaggg ac 72**

<210> 96  
 <211> 66  
 15 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 20 <222> 1..66  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DKBPG”  
 / organismo=“Secuencia artificial”

25 <400> 96  
**atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60**  
**gagggg 66**

<210> 97  
 <211> 66  
 30 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 35 <222> 1..66  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01C29WR”  
 / organismo=“Secuencia artificial”

40 <400> 97  
**atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60**  
**gagggg 66**

<210> 98  
 <211> 48  
 45 <212> ADN  
 <213> Secuencia artificial

<220>  
 <221> fuente  
 50 <222> 1..48  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BLAFB”  
 / organismo=“Secuencia artificial”

55 <400> 98  
**atatttaaag ctggtggagt tggaaggaat gatttgtgtc tggagggg 48**  
 <210> 99

<211> 62  
 <212> ADN  
 <213> Secuencia artificial

5 <220>  
 <221> fuente  
 <222> 1..62  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01AN93Y”  
 10 / organismo=“Secuencia artificial”

<400> 99  
 atatttaaag ctcccctgtg tgccttggtg gagttggagg aatgatttgt gtctgggagg 60  
 gg 62

15 <210> 100  
 <211> 64  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <222> 1..64  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01B80B0”  
 25 / organismo=“Secuencia artificial”

<400> 100  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aggaatgatt tgtgtctgga 60  
 gggg 64

30 <210> 101  
 <211> 64  
 <212> ADN  
 <213> Secuencia artificial

35 <220>  
 <221> fuente  
 <222> 1..64  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DFR9W”  
 40 / organismo=“Secuencia artificial”

<400> 101  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aggaatgatt tgtgtctgga 60  
 gggg 64

45 <210> 102  
 <211> 64  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <221> fuente  
 <222> 1..64  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BNPCR”  
 55 / organismo=“Secuencia artificial”

<400> 102

atatttaaag cttatcccct gtgtgccttg gtggagttgg aggaatgatt tgtgtctgga 60  
 gggg 64  
 <210> 103  
 <211> 65  
 5 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 10 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01EO6UA”  
 / organismo=“Secuencia artificial”  
 15 <400> 103  
 atatttaaag cttatcccc tgtgtgcctt ggtggagttg gaggaatgat ttgtgtctgg 60  
 agggg 65  
 <210> 104  
 <211> 65  
 20 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 25 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01D7FI6”  
 / organismo=“Secuencia artificial”  
 30 <400> 104  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60  
 agggg 65  
 <210> 105  
 <211> 65  
 35 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 40 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01B4QJQ”  
 / organismo=“Secuencia artificial”  
 45 <400> 105  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60  
 agggg 65  
 <210> 106  
 <211> 65  
 50 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 55 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01AP8OU”

ES 2 730 962 T3

/ organismo="Secuencia artificial"

<400> 106  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60

5  
 agggg 65

<210> 107  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

10  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01EGAYX"  
 / organismo="Secuencia artificial"

15  
 <400> 107  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60

20  
 agggg 65

<210> 108  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

25  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01COZ0O"  
 / organismo="Secuencia artificial"

30  
 <400> 108  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60

35  
 agggg 65

<210> 109  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

40  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="I935FME01A84M0"  
 / organismo="Secuencia artificial"

45  
 <400> 109  
 atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg 60

50  
 agggg 65

<210> 110  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

55  
 <220>  
 <221> fuente

ES 2 730 962 T3

<222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01CNOA1”  
 / organismo=“Secuencia artificial”

5

<400> 110  
**atatttaaag cttatcccct gtgtgccttg gtggagttgg aaggaatgat ttgtgtctgg** 60  
  
**agggg** 65

<210> 111  
 <211> 64  
 <212> ADN  
 <213> Secuencia artificial

10

<220>  
 <221> fuente  
 <222> 1..64  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01A1C15”  
 / organismo=“Secuencia artificial”

15

<400> 111  
**atatttaaag cttatcccct gtgtgccttg gtggagttga ggaatgattt gtgtctggga** 60  
  
**gggg** 64

<210> 112  
 <211> 60  
 <212> ADN  
 <213> Secuencia artificial

20

<220>  
 <221> fuente  
 <222> 1..60  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01AS8D7”  
 / organismo=“Secuencia artificial”

25

<400> 112  
**atatttaaag ctcccctgtg tgccttggtg gagttgagga atgatttgtg tctggagggg** 60

<210> 113  
 <211> 63  
 <212> ADN  
 <213> Secuencia artificial

30

<220>  
 <221> fuente  
 <222> 1..63  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BJ137”  
 / organismo=“Secuencia artificial”

35

<400> 113  
**atatttaaag cttatcccct gtgtgccttg gtggagttga ggaatgattt gtgtctggag** 60  
  
**ggg** 63

<210> 114  
 <211> 63  
 <212> ADN  
 <213> Secuencia artificial

40

<220>

ES 2 730 962 T3

<221> fuente  
 <222> 1..63  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01A46D0”  
 5 / organismo=“Secuencia artificial”

<400> 114  
**atatttaaag cttatcccct gtgtgccttg gtggagttga ggaatgattt gtgtctggag** 60  
**ggg** 63

10 <210> 115  
 <211> 59  
 <212> ADN  
 <213> Secuencia artificial

15 <220>  
 <221> fuente  
 <222> 1..59  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01ENGKQ”  
 20 / organismo=“Secuencia artificial”

<400> 115  
**atgaatgcag tttgtgtgcc ttggtggagt tggaaggagt gatttgcgtc tgggagggg** 59

25 <210> 116  
 <211> 58  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <221> fuente  
 <222> 1..58  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01A5DNG”  
 35 / organismo=“Secuencia artificial”

<400> 116  
**atgaatgcag tttgtgtgcc ttggtggagt tggaaggagt gatttgcgtc tgggaggg** 58

40 <210> 117  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01C7OXZ”  
 50 / organismo=“Secuencia artificial”

<400> 117  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**gaggg** 65

55 <210> 118  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

60 <220>

ES 2 730 962 T3

<221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DB053”  
 5 / organismo=“Secuencia artificial”

<400> 118  
 atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg 60  
 gaggg 65

10 <210> 119  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

15 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01CU25R”  
 20 / organismo=“Secuencia artificial”

<400> 119  
 atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg 60  
 gaggg 65

25 <210> 120  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

30 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01AJLKP”  
 35 / organismo=“Secuencia artificial”

<400> 120  
 atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg 60  
 gaggg 65

40 <210> 121  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

45 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01B8P5R”  
 50 / organismo=“Secuencia artificial”

<400> 121  
 -  
 atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg 60  
 gaggg 65

55 <210> 122  
 <211> 65

<212> ADN  
 <213> Secuencia artificial

5 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BSJKW”  
 / organismo=“Secuencia artificial”

10 <400> 122  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
  
**gaggg** 65

15 <210> 123  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial

20 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DZ64B”  
 / organismo=“Secuencia artificial”

25 <400> 123  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
  
**gaggg** 65

30 <210> 124  
 <211> 49  
 <212> ADN  
 <213> Secuencia artificial

35 <220>  
 <221> fuente  
 <222> 1..49  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01EPZFR”  
 / organismo=“Secuencia artificial”

40 <400> 124  
**atgaatgcag ttttggtgga gttggaagga gtgatttgcg tctggaggg** 49

45 <210> 125  
 <211> 49  
 <212> ADN  
 <213> Secuencia artificial

50 <220>  
 <221> fuente  
 <222> 1..49  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01ASAI7”  
 / organismo=“Secuencia artificial”

55 <400> 125  
**atgaatgcag ttttggtgga gttggaagga gtgatttgcg tctggaggg** 49

60 <210> 126  
 <211> 50

<212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 5 <221> fuente  
 <222> 1..50  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01C80EK”  
 / organismo=“Secuencia artificial”  
 10  
 <400> 126  
**atgaatgcag ttttgggtgga gttggaagga gtgatttgcg tctggagggg** 50  
  
 <210> 127  
 15 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 20 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01CN455”  
 / organismo=“Secuencia artificial”  
 25  
 <400> 127  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
  
**agggg** 65  
  
 <210> 128  
 30 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 35 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01C304D”  
 / organismo=“Secuencia artificial”  
 40  
 <400> 128  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
  
**agggg** 65  
  
 <210> 129  
 45 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
  
 <220>  
 50 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01CGQON”  
 / organismo=“Secuencia artificial”  
 55  
 <400> 129  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
  
**agggg** 65

<210> 130  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
 5  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01EJQ9M”  
 / organismo=“Secuencia artificial”  
 10  
 <400> 130  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**agggg** 65  
 15  
 <210> 131  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
 20  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01D218C”  
 / organismo=“Secuencia artificial”  
 25  
 <400> 131  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**agggg** 65  
 30  
 <210> 132  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
 35  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01DCUU6”  
 / organismo=“Secuencia artificial”  
 40  
 <400> 132  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**agggg** 65  
 45  
 <210> 133  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
 50  
 <220>  
 <221> fuente  
 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01AI0RD”  
 / organismo=“Secuencia artificial”  
 55  
 <400> 133

**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**agggg** 65  
 5 <210> 134  
 <211> 65  
 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 10 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01B6QGC”  
 / organismo=“Secuencia artificial”  
 15 <400> 134  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**agggg** 65  
 <210> 135  
 <211> 66  
 20 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 25 <222> 1..66  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01B6P77”  
 / organismo=“Secuencia artificial”  
 30 <400> 135  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat tttgcgtctg** 60  
**gagggg** 66  
 <210> 136  
 <211> 65  
 35 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 40 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01BRGZH”  
 / organismo=“Secuencia artificial”  
 45 <400> 136  
**atgaatgcag tttatcccct gtgtgccttg gtggagttgg aaggagtgat ttgcgtctgg** 60  
**agggg** 65  
 <210> 137  
 <211> 65  
 50 <212> ADN  
 <213> Secuencia artificial  
 <220>  
 <221> fuente  
 55 <222> 1..65  
 <223> /tipo de mol.=“ADN sin asignar”  
 / nota=“I935FME01A2VKU”

ES 2 730 962 T3

/ organismo="Secuencia artificial"

<400> 137  
**atgaatgcag tttatcccct gtgtgccttg gtggagtgg aaggagtgat ttgcgtctgg** 60

5 **agggg** 65

<210> 138  
 <211> 92  
 <212> ADN  
 <213> Secuencia artificial

10

<220>  
 <221> fuente  
 <222> 1..92  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01EBTDI"  
 / organismo="Secuencia artificial"

15

<400> 138  
**gaagagggga gtgggaaggt agaggctgag tgatttaggt attccaata atcaatagca** 60

20 **agaccactgt gagctgaata catcgggtta tc** 92

<210> 139  
 <211> 89  
 <212> ADN  
 <213> Secuencia artificial

25

<220>  
 <221> fuente  
 <222> 1..89  
 <223> /tipo de mol.="ADN sin asignar"  
 / nota="IQ4WJ2H01DJ2RX"  
 / organismo="Secuencia artificial"

30

<400> 139  
**gaagagggga gtgggaagtg aggctgagtg atttaggtat ttccataatc aatagcaaga** 60

35 **ccactgtgag ctgaatacat cgggttatc** 89

**REIVINDICACIONES**

1. Método de secuenciación libre de errores de ADN, que comprende las etapas de:
  - 5 (a) en una muestra que comprende ADN;
  - (b) digerir el ADN con una endonucleasa de restricción en condiciones adecuadas para obtener fragmentos de ADN de longitud similar,
  - 10 en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 5', en el que el nucleótido terminal de la proyección está fosforilado o,
  - en el que dicha endonucleasa de restricción puede proporcionar proyecciones en 3', en el que el nucleótido terminal de la proyección está hidroxilado en dichos fragmentos de ADN;
  - 15 (c) aparear un primer oligonucleótido con dichos fragmentos de ADN, en el que una primera secuencia de dicho primer oligonucleótido es complementaria a la proyección en 5' o en 3', respectivamente, de dicho fragmento de ADN, y una segunda secuencia de dicho primer oligonucleótido es complementaria a una primera secuencia de un segundo oligonucleótido, en el que dicho segundo oligonucleótido comprende una segunda y una tercera secuencias, en el que dicha segunda secuencia de dicho segundo oligonucleótido comprende una secuencia aleatorizada;
  - 20 (d) ligar dicho segundo oligonucleótido a dicho fragmento de ADN, mediante lo cual el primer oligonucleótido no se liga al ADN;
  - 25 (e) rellenar las proyecciones generadas;
  - (f) amplificar dichos fragmentos de ADN usando un tercer oligonucleótido que comprende una secuencia que se une a dicha tercera secuencia de dicho segundo oligonucleótido; y
  - 30 (g) secuenciar dichos fragmentos de ADN amplificados.
2. Método según la reivindicación 1, en el que dicho segundo oligonucleótido comprende además una cuarta secuencia que comprende un sitio de restricción de una endonucleasa específica de sitio.
- 35 3. Método según las reivindicaciones 1 ó 2, en el que dicho segundo oligonucleótido es un oligonucleótido de ADN, un oligonucleótido de ARN o un oligonucleótido de ADN/ARN.
- 40 4. Método según una cualquiera de las reivindicaciones 1 a 3, en el que dicho método comprende además la etapa (e'), en el que se añade una exonucleasa en dicha etapa (e').
5. Método según la reivindicación 4, en el que dicha exonucleasa es una enzima que degrada moléculas de ADN, ARN o ADN/ARN monocatenarios.
- 45 6. Método según una cualquiera de las reivindicaciones 1 a 5, en el que dicho ADN comprende (i) el genoma o transcriptoma de una célula individual, (ii) cromosoma(s) de una célula individual, (iii) ácidos nucleicos de exosomas u otras microvesículas de una célula individual o (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii).
- 50 7. Método según una cualquiera de las reivindicaciones 1 a 5, en el que dicho ADN comprende (i) el ADN de más de una célula individual, (ii) ADN fetal libre de células de más de una célula individual, (iii) ADN libre de células de más de una célula individual en suero y/o plasma de pacientes con cáncer o (iv) fragmento(s) o fracción/fracciones secundaria(s) del material según uno cualquiera de los puntos (i) a (iii).
- 55 8. Método según una cualquiera de las reivindicaciones 1 a 7, en el que dicha endonucleasa de restricción es MseI o un isoesquímico de la misma.
9. Método según una cualquiera de las reivindicaciones 1 a 8, en el que dicha secuencia aleatorizada comprende de 3 a 24 nucleótidos.
- 60 10. Método según una cualquiera de las reivindicaciones 1 a 9, en el que dicho primer oligonucleótido tiene la secuencia 5'-TAACTGACdd-3' y/o en el que dicho segundo oligonucleótido tiene la secuencia tal como se muestra en SEQ ID NO: 1 y/o en el que dicho tercer oligonucleótido tiene la secuencia tal como se muestra en SEQ ID NO: 2.
- 65 11. Método según una cualquiera de las reivindicaciones 1 a 10, en el que el último nucleótido en 3' del primer

oligonucleótido es un dd-nucleótido.

- 5
12. Método según una cualquiera de las reivindicaciones 2 a 11, en el que dicho método comprende además la etapa (f') en el que se añade una endonucleasa de guiado en dicha etapa (f').
13. Uso de los fragmentos de ADN secuenciado obtenidos mediante el método según una cualquiera de las reivindicaciones 1 a 12 en métodos para el análisis de secuencias de ADN, generación de árboles de linaje celular o evaluación de números de copias.
- 10
14. Uso según la reivindicación 13, en el que el método para el análisis de secuencias de ADN es secuenciación de genoma completo, secuenciación de exoma completo, secuenciación de reguloma completo, análisis de metilación basado en secuenciación, detección de punto de rotura basado en secuenciación, secuenciación de CHIP o secuenciación dirigida y variaciones de los mismos.
- 15
15. Oligonucleótido de cuatro partes para llevar a cabo el método según la reivindicación 1 que comprende una secuencia fija, secuencia aleatorizada, sitio de reconocimiento de nucleasa de restricción y sitio de restricción, y sitio de unión a cebador, en el que dicha secuencia aleatorizada comprende de 3 a 24 nucleótidos y dicho sitio de reconocimiento de nucleasa de restricción es un sitio de reconocimiento de una endonucleasa de guiado, en el que dicha secuencia fija comprende GTCAGT, en el que dicho sitio de reconocimiento de nucleasa de restricción comprende SEQ ID NO: 3 y en el que dicho sitio de unión a cebador comprende SEQ ID NO: 4.
- 20
16. Oligonucleótido de cuatro partes según la reivindicación 15, que comprende SEQ ID NO: 5 ó 12.
- 25
17. Oligonucleótido de cuatro partes según la reivindicación 15, que comprende SEQ ID NO: 14.

Figura 1

Tipo de oligonucleótido	Número n/m	Secuencia de oligonucleótido
DSL <sub>12</sub>	12	GCT AGG GAT AAC AGG GTA ATG C NNN NNN NNN NNN GTC AGT
DSL <sub>9</sub>	9	GCT AGG GAT AAC AGG GTA ATG C NNN NNN NNN GTC GT
DSS <sub>5</sub>	5	TA ACG ACdd
DSS <sub>6</sub>	6	TA ACT GACdd
DSPCR		GCT AGG GAT AAC AGG GTA ATG C

Figura 2

Medida estadística	2/3 PCR positivas	3/3 PCR positivas
Verdaderos positivos	35	29
Falsos negativos	1	7
Verdaderos negativos	34	36
Falsos positivos	2	0
Sensibilidad	0.97	0.81
Especificidad	0.94	1.0
Valor predictivo positivo	0.95	1.0
Valor predictivo negativo	0.97	0.84

Figura 3

	N	Secuencias detectadas en el ensayo de QC de tres marcadores			
		3/3	2/3	1/3	0/3
WBC individual (sin fijar)	88	80 (90.9%)	3 (3.41%)	0	5 (5.68%)
WBC individual (CellSearch)	189	73 (38.6%)	51 (27.0%)	20 (10.6%)	45 (23.8%)
CTC individual (CellSearch)	510	102 (20.0%)	89 (17.5%)	55 (10.8%)	264 (51.8%)

Figura 4

Ensayo molecular	Células analizadas	Índice de integridad genómica (GI)				Valor de p chi cuadrado	
		GII 0	GII 1	GII 2	GII 3 GII 4		
PIK3CA HS1	n = 383	7/23 (30.4%)	14/25 (56.0%)	48/62 (77.4%)	102/117 (87.2%)	146/156 (93.6%)	<0.0001
PIK3CA HS2	n = 383	8/23 (34.8%)	16/25 (72.0%)	55/62 (88.7%)	109/117 (93.2%)	149/156 (95.5%)	<0.0001
PIK3CA completo	n = 383	4/23 (17.4%)	12/25 (48.0%)	45/62 (72.6%)	97/117 (82.9%)	141/156 (90.4%)	<0.0001
HER2 qPCR	n = 351	3/12 (25.0%)	8/18 (50.0%)	41/61 (67.2%)	95/112 (84.8%)	136/148 (91.9%)	<0.0001
aCGH	n = 50	no evaluado	no evaluado	4/5 (80.0%)	7/9 (77.8%)	36/36 (100%)	0.016

Figura 5

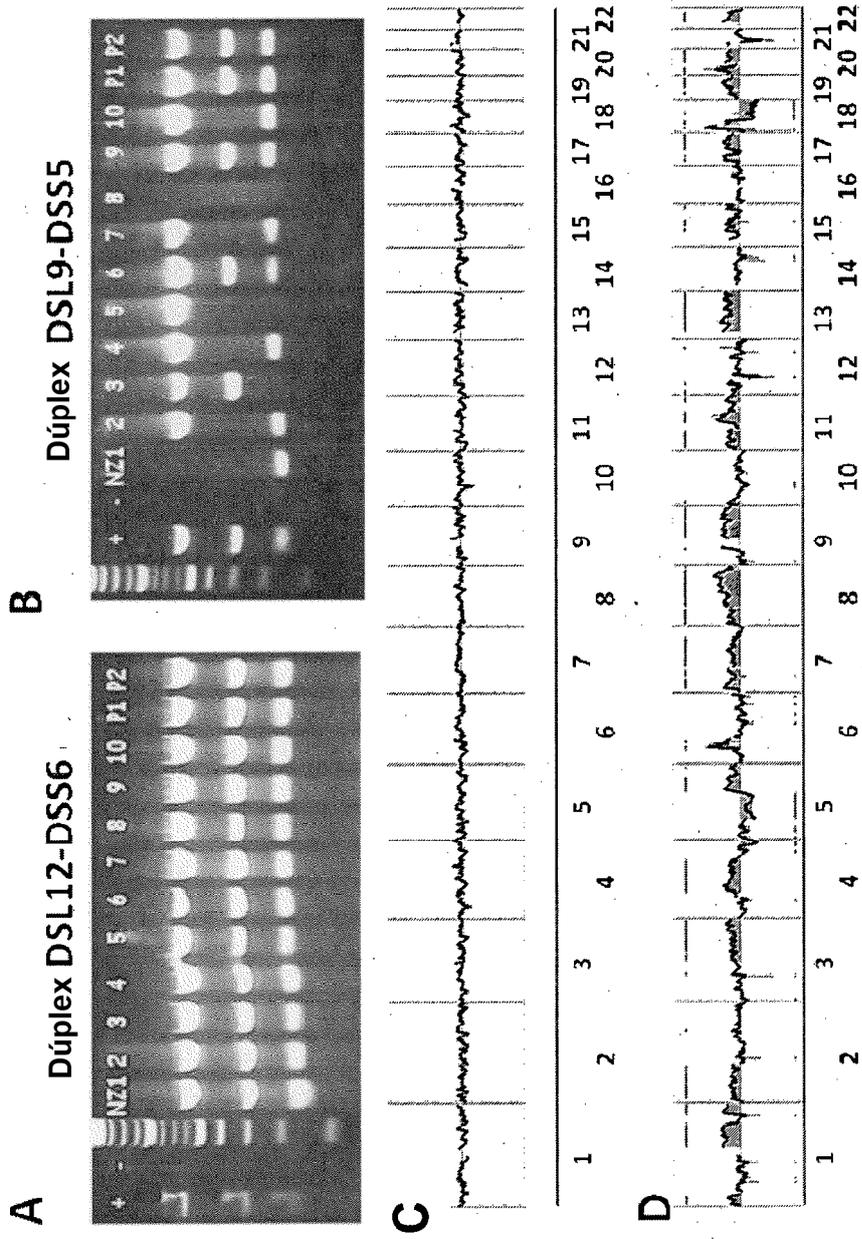


Figura 6

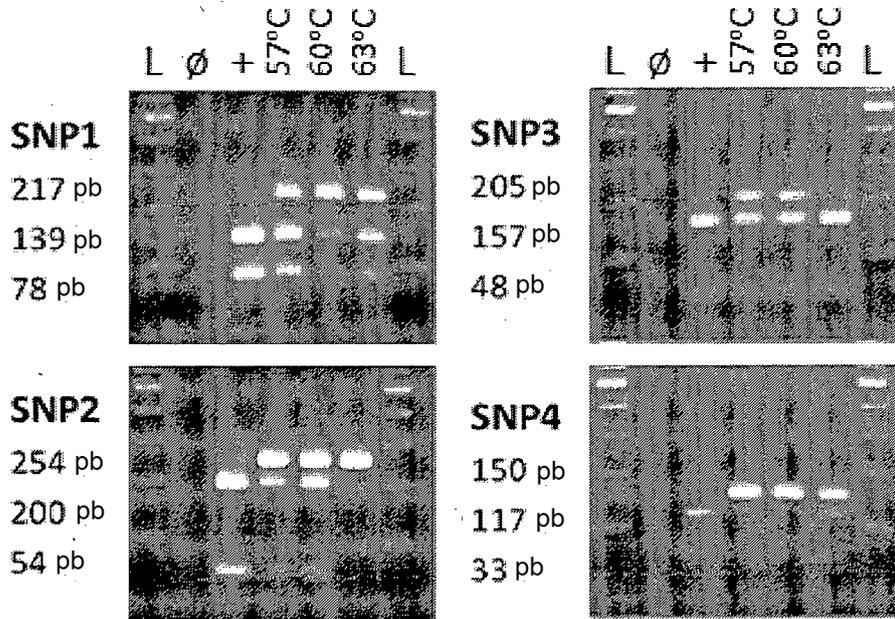


Figura 7

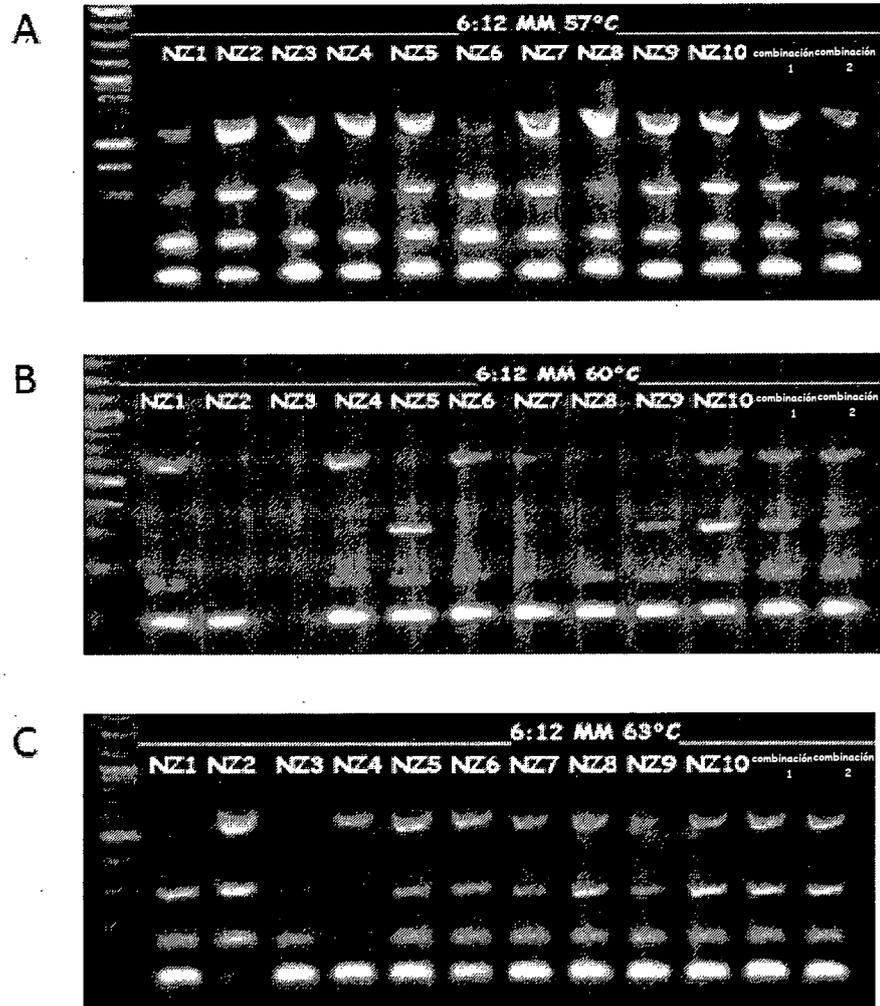
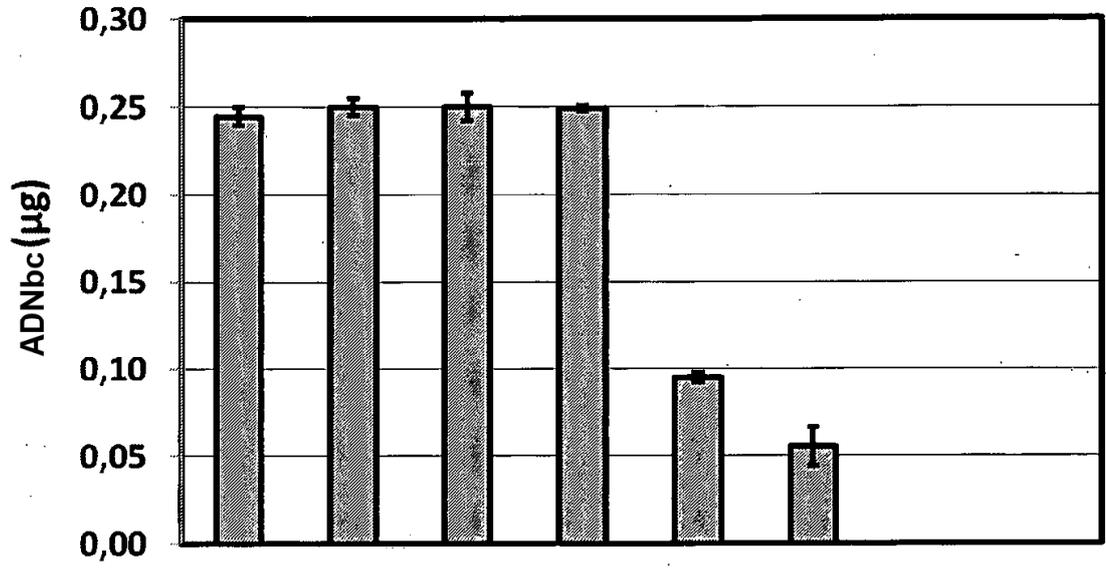


Figura 8



ADNbc	+	+	+	+	-	-	-	-
Adaptador de PCR	-	-	+	+	+	+	-	-
Exonucleasa	-	+	-	+	-	+	+	-

Figura 9

**A** GCTA**GGG**ATAACAGGGTAAATCCGTGGTCAGGGGGTCAGTTAAAGTAGTTTGGCAAAATGGATAAAAGTCA  
AGACCCATCAGTGTGCTGTATTCAGGAAACCCATCTCATGTGCAGAGACACACATAGGCTCAAAAATAAAAGGA  
TGGAGGAAGATCTACCAAGCAAATGGAAAAAACAAAAAAGGCAGGGTTGCAATCCTAGTCTCTGATAAAACAG  
ACTTTATACCAACAAAAGATCAAAGAGACAAAGAGCCATTACATAATGGTAAAGAGATCAATTCAACAAGA  
AGAGCTAACTATCCTAAATATATATGCACCCAATACAGGAGCACCCAGATTCATAAAGCAAGTCCCTGAGTGAC  
CTACAAAAGAGACTTAGACTCCCAGACATTAAGTGACTCTCGTCTTTCATTACCCGTTTATCCCTAGC

Anotación del fragmento: chr10, posición de 1811253 a 1811607

**B** GCTA**GGG**ATAACAGGGTAAATGCCGGAAAGTAGGCTCAGTTAACCAAGAAAAGAGGAGAGAAAATACAAATAA  
GCTCAATAGGAACAAACAGGAGACATTACAACTGACACTATATAAATACAAAAGATCATTCAAGGCTACTG  
TGAAACACCTTTACATGCATAAACTAGAAAACCTAAAAGACATGATAAATCTCTGAAAAGTAGAGAGGCATCA  
CATTACCCAACTTCAACTATACTATAAGGCTATAGTCACCAACTCAGCATGGCACTGGTATAAAAATAGGCA  
TGTAGACCAATGAAACAGAAAAGAGAACCAGACATAAAGCCAAATACAGCCAACCAATCTTTGACAAAGCAA  
ACAAAACATAAAGTGGGAAAAGGACCGCTCTATTCACAAAATGGTCTGGGATAATTGGCAAGCCACATGTAG  
AAGATGAAACTGGATCCTCATCTCCCTTTATACAAAATCAACACAAAATAGATAAAAAGCTTAAGTGAGCC  
ACAAGGATAGGCATTACCCTGTATCCCTAGC

Anotación del fragmento: chr5, posición de 96525427 a 96525889

**C** GCTA**GGG**ATAACAGGGTAAATGCTAGGAGAAAATGTCAGTTAAATCTATAAATTGTGTGCCACGTGCCGT  
GGCTCATGACTGTAAATCCCAGCACTTTGGGAGTCCGAGGTGGGCAGATCACGAGGTCAGGAGATTGAGAC  
CATCTTGCTAACACGGTGAAACCCTGTCTCTACTAAAAATACAAAACAAAAATAGCCGGTGAGGTGG  
CAGGCACCTGTAGTCCCAGCTACTCCAGAGGCTGAGGTGAGAGAATGGCGTGAACCTGGGAGGCAGAGCT  
TGCAGTGAGCTGAGATTGTGCCACTGCACTCCAGTCTGGGTGACAGAGCAAGACTCTGTCTCAAAAAAAA  
AAAAAAAAAAAAATCTATAAATTGTGTGGGCAGTATGGCCTTTTTAGTGACACTAAAGTCACAGC  
ATTACCCCTGTATCCCTAGC

Anotación del fragmento de restricción: chrX, posición de 94553619 a 94553978

**Figura 10**

Etapa	Temp.	Tiempo
Digestión	42°C	10 h
Inactivación	80°C	10 min
	4°C	∞

**Figura 11**

Etapa	Temp.	Tiempo
Digestión	37°C	3 h
Inactivación	65°C	5 min
	4°C	∞

**Figura 12**

Temp.	Δ de temp. / ciclo
65°C → 15°C	-1°C/min
15°C	

**Figura 13**

Etapa	Temp.	Tiempo
Ligación	15°C	durante la noche

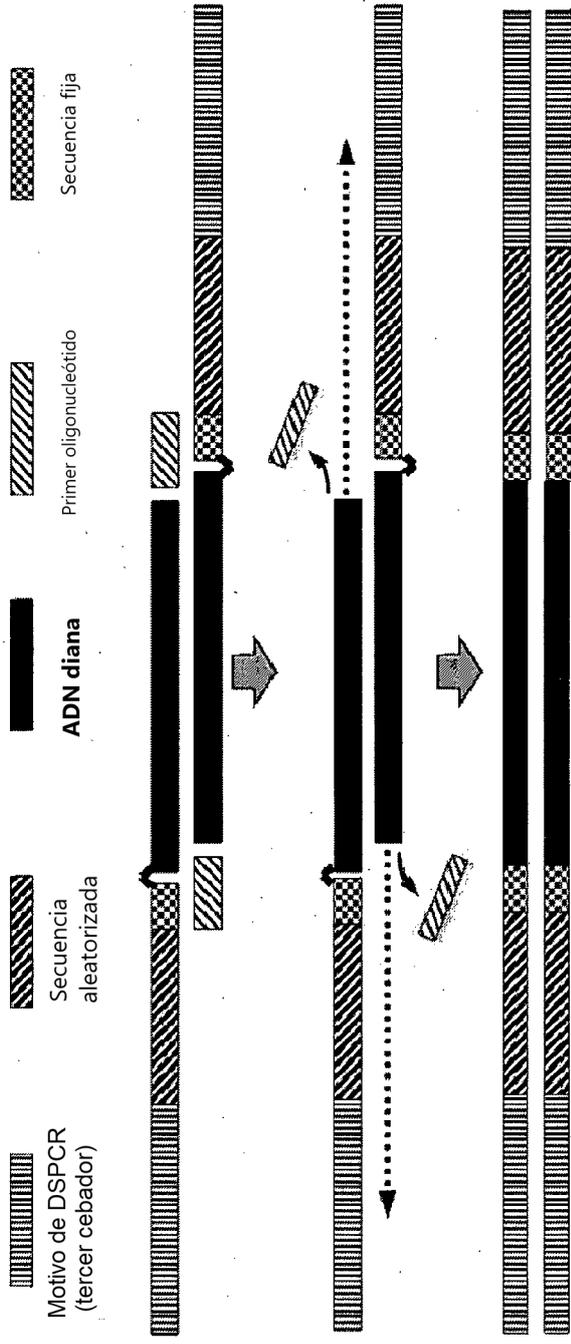
**Figura 14**

Etapa	Temp.	Tiempo
Relleno	68°C	3 min
1. Ciclación de PCR 8x	94°C	40 seg
	57/60/63°C	30 seg
	68°C	1min 30seg +1 seg/ciclo
2. Ciclación de PCR 40x	94°C	40 seg
	57/60/63°C	30 seg
	68°C	1 min 45 seg +1seg / ciclo
Extensión final	68°C	3 min 40 seg
	4°C	∞

**Figura 15**

Etapa	Temp.	Tiempo
Relleno	68°C	3min
	80°C	2 min
Adición de la exonucleasa I	4°C	PAUSA
	37°C	30 min
	85°C	15 min
Adición de cebador de DSPCR		PAUSA
1. Ciclación de PCR 15x	94°C	40 seg
	57°C	30 seg
	68°C	1min 30 seg +1 seg / ciclo
2. Ciclación de PCR 9x	94°C	40seg
	57°C +1°C/Z	30seg
	68°C	1min 45seg +1 seg / ciclo
3. Ciclación de PCR 23x	94°C	40 seg
	65°C	30 seg
	68°C	1min 53 seg +1 seg / ciclo
Extensión final	68°C	3min 40seg
	4°C	∞

Figura 16



DSS6: 5' - TA ACT GACdd - 3'  
 (Primer cebador)

DSL12: 5' - GCT AGG GAT AAC AGG GTA ATG C NNN NNN NNN NNN GTC AGT-3'  
 (Segundo cebador)

DSPCR: 5' - GCT AGG GAT AAC AGG GTA ATG C-3'  
 (Tercer cebador)

Figura 17

Figura 17

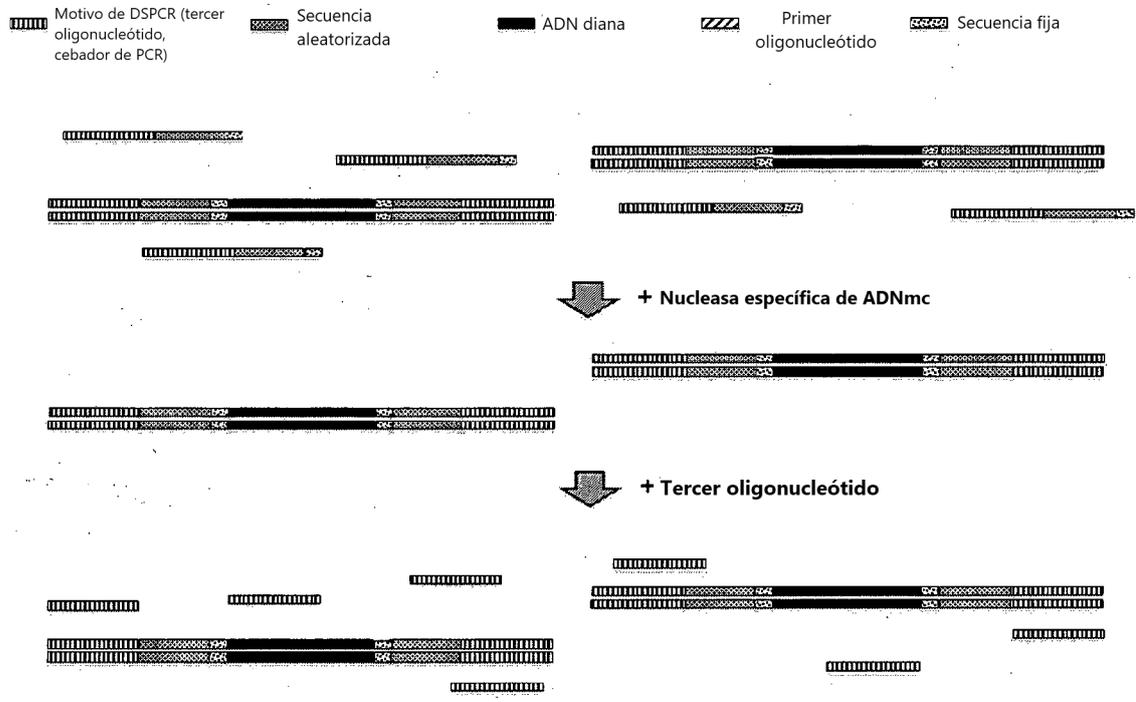


Figura 18

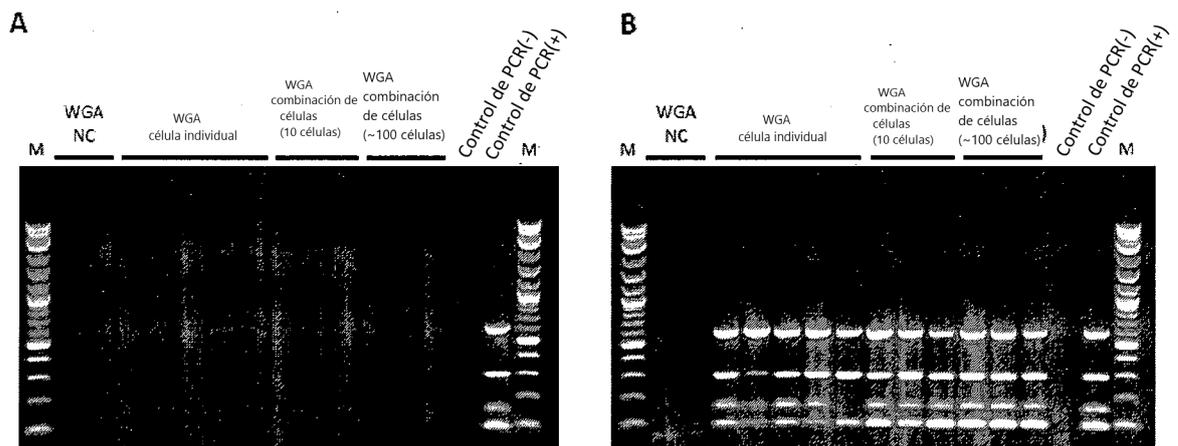


Figura 19a

ID de lectura	Secuencia aleatorizada	chr6_124,934,321_124,934,396
Referencia	TGGGAAG-T-GAGGCTGACTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	X X X
IQ4WJ2H01EBTDI_GAAGAGGGG-AG	TGGGAAGGTAGAGGCTGACTGATTTAGGTATTTCCAATAATCAATAGCAAGACCCTGTGAGCTGAATACATCGGGTTATC	
IQ4WJ2H01DJ2RX_GAAGAGGGG-AG	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCA-TAATCAATAGCAAGACCCTGTGAGCTGAATACATCGGGTTATC	
IQ4WJ2H01EV3C9_GAAGAGGGG-AG	TGGGAAG-TAGAGGCTGAGTGATTTAGGTATTTCCAATAATCAATAGCAAGACCCTGTGAGCTGAATACATCGGGTTATC	
IQ4WJ2H01D9VJ6_GAAGAGGGG-AG	TGGGAAG-T-GAGGCTGACTGATTTAGGTATTTCCAATAATCAATAGCAAGACCCTGTGAGCTGAATACATCGGGTTATC	
IQ4WJ2H01BDRUD_GAAGAGGGGAG	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCAATAATCAATAGCAAGACCCTGTGAGCTGAATACATCGGGTTATC	
***** **	***** * *****	
IQ4WJ2H01DC2CL_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGACTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	X X X
IQ4WJ2H01AV5Z7_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H01BN7XB_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H01AH386_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGACTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H01D63DR_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H01BV1TH_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H01BI7HX_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGACTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H02IESZE_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGAGTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
IQ4WJ2H02G69J5_ATGAGTGGGATA	TGGGAAG-T-GAGGCTGACTGATTTAGGTATTTCCA-TA-TCAATAGCAAGACCCTGTGAGCTGAATACATCAGGTTATC	
***** **	***** * *****	X X X

Figura 19b

ID de lectura	Secuencia aleatorizada	chr22:39,277,551-39,277,608
Referencia	CACTGCACGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	X
I935FME02I8EWU_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02HST0F_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02IS3DB_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02GVA1A_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02GV56X_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02HT9UP_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02IPW2U_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02HNKYA_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02IOS0E_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02HXEG8_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02JFTGT_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02GBSUA_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02GR861_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME02IPFUP_GATAAATATGAA	CACTGCACCGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
***** **	***** * *****	
I935FME01C1GTK_AGACGTTTTAAG	CACTGCACGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	X
I935FME01CN901_AGACGTTT-AAG	CACTGCACGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME01BAK2C_AGACGTTTTAAG	CACTGCACGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
I935FME01AXYNN_TAAATAGGGTTT	CACTGCACGTGAAATTCGTGCCTCCGAAGACAGGCTCGCTATTTTTAGAAAGAGGG	
***** **	***** * *****	X

La lectura "I935FME01AXYNN" se sometió a mapeo inverso, es decir, se originó a partir de la cadena opuesta de las otras lecturas dentro de su grupo y por tanto tiene una secuencia aleatorizada diferente.

Figura 19c

ID de lectura	Secuencia aleatorizada	chr22:18,455,108-18,455,137	chr22:18,455,218-18,455,247
Referencia	AGGTAAGTGGTG-ACAACGGGTTTCCCAA	X	ACAGGTGAACGGACGTACGGGAGTAGGGAC
		X	X
I935FME02G3JAX_AGC	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		-----
I935FME02HPZCG_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02HEGFO_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACA-----
I935FME02FY2OL_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGA-----
I935FME02FUAIM_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02HNVDL_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02GP6YE_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02IAAW2_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02I4UCC_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02FVCO0_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGA-----
I935FME02JTEOX_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACA-----
I935FME02GYCTS_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		-----
I935FME02HYTON_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTA----
I935FME02JBAVG_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		-----
I935FME02FX9GA_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02GYILN_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		-----
I935FME02ID4XH_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02HVJJE_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02H32G2_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGA-----
I935FME02F6QOH_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02HNHHY_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTA----
I935FME02FOIUY_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02GSN70_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		-----
I935FME02FR6XR_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCC---		-----
I935FME02IP2KO_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
I935FME02JKE75_AGC	AGGTAAGTGGTG-ATAACGGGTTTCCCAA		ACAGGTGAACGGACGTACGGGAGTAGGGAC
	***** **	***** **	***** **
		X	X
I935FME01AGM6X_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01AMWUD_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATAC-----
I935FME01EEHFG_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01CP0EN_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGA-TAGGGAC
I935FME01CAYQ8_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01CT2IH_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01BLNIO_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01EW5VF_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01BR9EW_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01D9HRR_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGA-TAGGGAC
I935FME01DYFB5_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01D6FUV_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01BB15Y_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02IGF97_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02FLW03_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02FOXSW_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02GUX6V_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02GI4Z4_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02H2M7T_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02JHRO7_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01DYW8K_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME01BFXU5_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02IQWMO_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02HQ894_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAA		ACAGGTGAACGGACATACGGGAGTAGGGAC
I935FME02JROVV_ATAGACTAGGCG	AGGTAAGTGGTG-ACAACGGGTTTCCCAGA		ACAGGTGAACGGACATACGGGAGTAGGGAC
	***** **	***** **	***** **
		X	X

Figura 19d

ID de lectura	Secuencia aleatorizada	chr22:42,820,701-42,820,754
Referencia	AAT-GGGCACTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGGGAGGGG	
	X XXX XX	X X X
I935FME01DKBPG_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGGGAGGGG	
I935FME01C29WR_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGGGAGGGG	
I935FME01BLAFB_ATATTTAAAGCT	-----GGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01AN93Y_ATATTTAAAGCT	----CCCCTGTGTGCCTTGGTGGAGTTGGA-GGAATGA-TTTGTCTCTGGGAGGGG	
I935FME01B80BO_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGA-GGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01DFR9W_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGA-GGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01BNPCR_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGA-GGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01E06UA_ATATTTAAAGCT	TATCCCCTGTGTGCCTTGGTGGAGTTGGA-GGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01D7FI6_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01B4QJQ_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01AP8OU_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01EGAYX_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01C0Z0O_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01A84M0_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01CNOA1_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01A1C15_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTG-A-GGAATGA-TTTGTCTCTGGGAGGGG	
I935FME01AS8D7_ATATTTAAAGCT	----CCCCTGTGTGCCTTGGTGGAGTTG-A-GGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01BJ137_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTG-A-GGAATGA-TTTGTCTCTGG-AGGGG	
I935FME01A46D0_ATATTTAAAGCT	TAT-CCCCTGTGTGCCTTGGTGGAGTTG-A-GGAATGA-TTTGTCTCTGG-AGGGG	
	*****	*****
	X XXX XX	X X X
I935FME01ENKQ_ATGAATGCAGTT	-----TGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGAGGGG	
I935FME01A5DNG_ATGAATGCAGTT	-----TGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01C70XZ_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01DB053_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01CU25R_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01AJLKE_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01B8P5R_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01BSJKW_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01DZ64B_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGGGA-GGG	
I935FME01EPZER_ATGAATGCAGTT	-----TTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-A-GGG	
I935FME01ASAI7_ATGAATGCAGTT	-----TTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-A-GGG	
I935FME01C80EK_ATGAATGCAGTT	-----TTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01CN455_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01C304D_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01CGOON_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01EJQ9M_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01D218C_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01DCUU6_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01AI0RD_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01B6QGC_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01B6P77_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01BRGZH_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
I935FME01A2VKU_ATGAATGCAGTT	TAT-CCCCTGTGTGCCTTGGTGGAGTTGGAAGGAGTGA-TTTGCGTCTGG-AGGGG	
	*****	*****
	X XXX XX	X X X

Todas las lecturas mostradas se sometieron a mapeo inverso con la referencia.

Figura 20

