

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 388 862**

51 Int. Cl.:  
**C12Q 1/68** (2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **05763597 .1**  
96 Fecha de presentación: **20.06.2005**  
97 Número de publicación de la solicitud: **1766078**  
97 Fecha de publicación de la solicitud: **28.03.2007**

54 Título: **Selección de sonda de longitud variable**

30 Prioridad:  
**18.06.2004 US 581121 P**

45 Fecha de publicación de la mención BOPI:  
**19.10.2012**

45 Fecha de la publicación del folleto de la patente:  
**19.10.2012**

73 Titular/es:  
**Roche NimbleGen, Inc.**  
**1220 N. Market St., Suite 334**  
**Wilmington, DE 19801-2535, US**

72 Inventor/es:  
**ALBERT, Thomas, J.;**  
**NORTON, Jason y**  
**GREEN, Roland**

74 Agente/Representante:  
**de Elzaburu Márquez, Alberto**

ES 2 388 862 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

Selección de sonda de longitud variable

## ANTECEDENTES DE LA INVENCIÓN

- 5 La llegada de la tecnología de micromatrices de ADN hace posible construir una matriz con cientos de miles de secuencias de ADN en una superficie muy pequeña, tal como del tamaño de un portaobjetos para microscopio. Véanse, p. ej., las patentes de los EE.UU. n.º 6 375 903 y n.º 5 143 854. La descripción de la patente de los EE.UU. n.º 6 375 903 permite la construcción de los también denominados instrumentos sintetizadores de matrices sin máscara (MAS, por su nombre en inglés) en los que se utiliza la luz para dirigir la síntesis de las secuencias de ADN, estableciéndose la dirección de la luz con un dispositivo microespejo digital (DMD). Utilizando un instrumento MAS, 10 la selección de secuencias de ADN que constituirán la micromatriz está controlada mediante el programa de ordenador, por lo que se pueden construir matrices a medida para cada caso según se soliciten. En general, la tecnología de síntesis de micromatrices de ADN basada en MAS permite la síntesis en paralelo de unos 800 000 oligonucleótidos únicos en una área muy pequeña de un portaobjetos para microscopio estándar. Las micromatrices se sintetizan de forma general utilizando luz para dirigir qué oligonucleótidos se sintetizan en cada posición concreta en una matriz, denominándose rasgos (*features*) a cada una de estas posiciones. Otra ventaja de los instrumentos de síntesis de micromatrices por MAS es que la denominación de las sondas se encuentra controlada por el programa informático. Esto permite diseñar micromatrices diseñadas a medida y utilizarlas con la máxima flexibilidad en cuanto al diseño de sondas para matrices, ya que se pueden fabricar sondas diferenciales para cada matriz, incluso para diferentes matrices cuyo objetivo sea analizar los mismos elementos genéticos.
- 20 Con la disponibilidad de los genomas enteros de cientos de organismos, para los cuales se ha depositado por lo general una secuencia de referencia en una base de datos pública, se están utilizando las micromatrices para realizar análisis de secuencias del ADN aislado de tales organismos. Una técnica que se puede utilizar para identificar una variante genética es la de secuenciar el ADN genómico de un individuo y luego comparar esta secuencia con la secuencia de referencia de dicho organismo. Se ha encontrado que muchas diferencias en la 25 secuencia del ADN se presentan como variaciones únicas de la secuencia del ADN, que a menudo se denominan polimorfismos mononucleotídicos o SNP (por su nombre en inglés). La comparación de las secuencias entre el genoma problema y el genoma de referencia de una especie se ha realizado con un mecanismo de fuerza bruta, la secuenciación capilar, para identificar los SNP de un individuo en concreto.
- 30 Una etapa clave y una estrategia más reciente para identificar las variaciones genéticas asociadas a las enfermedades es resecuenciar los genes candidatos u otras regiones genómicas de interés en los pacientes y en los controles para identificar los SNP asociados a un fenotipo determinado (véase Sakai et al., (1989) *PNAS* 86: 6230-6234). Una estrategia de resecuenciación que ha demostrado tener resultados significativos utiliza la tecnología de micromatrices con oligonucleótidos (Hacia et al., (1999) *Nature Genetics*, 21 (1 Suppl): 42-7).
- 35 En particular, este tipo de estrategia de resecuenciación con matrices (ABR, por su nombre en inglés) depende de la hibridación diferencial de fragmentos genómicos sobre oligonucleótidos cortos de concordancia (coincidencia) perfecta (PM, por su nombre en inglés) y con discordancias (MM, por su nombre en inglés). Cada nucleótido a consultar se encuentra localizado en la posición central de un oligonucleótido. Para cada oligonucleótido PM también se sintetizan en la matriz las sondas que representan los tres nucleótidos discordantes posibles, cada uno para representar cada posible SNP en la misma posición central. Las diferencias de la intensidad de las señales de 40 hibridación entre las secuencias que se fijan fuertemente a los oligonucleótidos PM y las que unen mal a los correspondientes oligonucleótidos MM hacen posible diferenciar la base correcta en una posición de secuencia dada. Así pues, en teoría, siempre que haya un SNP, la sonda con discordancias que representa este SNP tendrá una intensidad de señal más alta que la correspondiente sonda que concuerda con la secuencia de referencia.
- 45 Sin embargo, debido a que no somos capaces de predecir la fuerza de la señal, al variar la eficacia de hibridación, y otras diferentes fuentes de ruido, este método da lugar típicamente a muchas posiciones de bases cuyas identidades se predicen de modo incorrecto. Por ejemplo, dado que todas las sondas para matrices se deben hibridar en las mismas condiciones de temperatura y de rigor de la hibridación, pueden aparecer problemas con sondas que tienen temperaturas de fusión ( $T_m$ ) que divergen significativamente de la temperatura a la que se hibrida la matriz. En el caso de las sondas con una  $T_m$  baja, las dianas hibridadas se despegarían significativamente de la 50 superficie de la matriz, lo que produciría poca o ninguna señal. En el caso de las sondas con una  $T_m$  elevada, la discordancia de una única base podría no desestabilizarla significativamente y no proporcionaría una discriminación adecuada para que la deducción de la base fuera robusta.
- 55 Otro problema con la discriminación de una única base es que la posición de la discordancia puede alterar significativamente la capacidad que una sonda tiene para proporcionar una discriminación significativa y no permitiría que deducción de la base fuera robusta. Por ejemplo, si la  $T_m$  del trozo de la sonda a uno de los lados de la discordancia es muy alta, este trozo de sonda puede mostrar una hibridación robusta independientemente de la discordancia y, por lo tanto, la discordancia no proporcionaría suficiente discriminación para la deducción de la base. Así pues, una contribución deseable a la técnica serían las estrategias alternativas para incrementar la eficacia y la exactitud de los análisis basados en matrices, tales como la resecuenciación de ADN, para identificar mutaciones en

el genoma de los organismos.

Las patentes de los EE.UU. n.º 6 468 744 y n.º 6 228 575 describen (i) el análisis de polimorfismos genéticos y del número de copias génicas y (ii) la identificación de especies mediante chips y la caracterización fenotípica de microorganismos utilizando micromatrices de oligonucleótidos, respectivamente.

- 5 Anderson et al (1995) *Gene Probe: A Practical Approach*, páginas 1-29 está relacionado con la estrategia de hibridación. Nuwaysir et al (2002) *Genome Research*, páginas 1749-1755 está relacionado con el análisis de la expresión génica mediante matrices de oligonucleótidos producidas mediante fotolitografía sin máscara.

### BREVE RESUMEN DE LA INVENCION

10 La presente invención se resume como un nuevo método para incrementar la eficacia y la exactitud del mapeo de mutaciones de alto rendimiento y de la resecuenciación de genomas mediante un algoritmo de selección de sondas de longitud variable para seleccionar racionalmente las sondas a utilizar en el diseño de matrices de oligonucleótidos. Así pues, la presente invención da a conocer un método y un algoritmo para variar de modo racional la longitud de las sondas oligonucleotídicas que permiten discriminar con exactitud las discordancias de una única base en los proyectos de resecuenciación a gran escala.

15 En un aspecto, la invención da a conocer un método para producir una micromatriz que tiene una serie de rasgos que comprenden una multitud de sondas oligonucleotídicas de longitud variable que tienen una discordancia nucleotídica respecto a la secuencia de referencia, y el método comprende las etapas de:

20 seleccionar una multitud de sondas de longitud variable para matrices de oligonucleótidos, en donde las sondas de longitud variable se deducen introduciendo una longitud de sonda mínima, una longitud de sonda máxima y una temperatura de fusión teórica para una multitud de sondas para matrices de oligonucleótidos en un algoritmo de selección de sondas; en donde el algoritmo es capaz de calcular longitudes de sondas oligonucleotídicas variables, la  $T_m$  teórica de cada sonda cuando se divide por la mitad, por lo que se puede variar la longitud de la porción de la sonda a cada lado de la discordancia de modo que permanezca dentro de los parámetros de longitud especificados para alcanzar la mitad de la  $T_m$  teórica especificada, en donde las sondas muestran las características siguientes: (i) temperatura de fusión similar y (ii) la posición de la discordancia en cada sonda está localizada en el centro termodinámico aproximado de cada sonda; y

25 sintetizar las sondas oligonucleotídicas inmovilizadas en un sustrato que alcanzan la mitad de la  $T_m$  teórica especificada, en donde todas las sondas muestran las características siguientes: (i) temperatura de fusión similar y (ii) la posición de la discordancia en cada sonda está localizada en el centro termodinámico aproximado de cada sonda; y

30 sintetizar las sondas oligonucleotídicas inmovilizadas en un sustrato para producir la micromatriz, en donde la longitud de las sondas se determina mediante la etapa de selección para que la longitud de las sondas en algunos rasgos sea diferente de la longitud de las sondas en otros rasgos. En particular, tal micromatriz se puede producir mediante la síntesis de matrices sin máscara (MAS, por su nombre en inglés).

35 Este método puede proporcionar una mayor discriminación de las discordancias para la deducción de bases que las sondas que no tienen las características anteriores.

Además, la  $T_m$  teórica se puede alcanzar con el incremento de la longitud de las sondas que tienen una  $T_m$  baja y la disminución de la longitud de las sondas que tienen una  $T_m$  alta. En consecuencia, podrían alargarse las sondas oligonucleotídicas con una  $T_m$  baja y podrían acortarse las sondas con una  $T_m$  alta.

40 Las sondas de longitud variable para matrices de oligonucleótidos tienen temperaturas de fusión similares y una posición discordante en cada sonda que se mantiene en el centro termodinámico aproximado de cada oligonucleótido.

La invención utiliza un algoritmo de selección de sondas que es capaz de aceptar datos de longitud de sondas mínimos y máximos junto con una  $T_m$  teórica para todas las sondas de la matriz.

45 La  $T_m$  teórica de una sonda de interés se puede dividir por la mitad y a cada trozo de la sonda de cada lado de la discordancia se le puede variar la longitud, de modo que permanezca dentro de los parámetros de longitud especificados y alcance la mitad de la  $T_m$  teórica especificada.

La  $T_m$  de la sonda se puede calcular con la fórmula:  $T_m = 5 * (G_n + C_n) + 1 * (A_n + T_n)$ , en donde n es el número de cada base específica presente en la sonda.

50 También se refiere una micromatriz que comprende

un sustrato;

una multitud de rasgos que comprenden copias de un nucleótido monocatenario construido en el lugar, inmovilizado

por un extremo al sustrato, conteniendo la sonda un nucleótido discordante con respecto a la secuencia de referencia, con la discordancia en el centro termodinámico aproximado de la sonda, siendo la longitud de las sondas de un rasgo diferente de la longitud de las sondas de otros rasgos de la micromatriz, y las sondas tienen temperaturas de fusión similares.

- 5 En la micromatriz a la que se hace referencia, la longitud de los rasgos de la micromatriz se varía para alcanzar una temperatura de fusión determinada de un complejo formado entre la sonda y un ácido nucleico que se unirá a la sonda.

Otros objetos, ventajas y rasgos de la presente invención serán evidentes a partir de la especificación que viene a continuación.

## 10 BREVE DESCRIPCIÓN DE LAS DISTINTAS PERSPECTIVAS DE LOS DIBUJOS

No aplicable

## DESCRIPCIÓN DETALLADA DE LA INVENCION

15 La presente invención da a conocer ampliamente un nuevo método para incrementar la eficacia y la exactitud del mapeo de mutaciones de alto rendimiento y la resecuenciación de genomas mediante el uso de un algoritmo de selección de sondas de longitud variable para seleccionar de modo racional las sondas utilizadas en el diseño de matrices de oligonucleótidos. Específicamente, la invención da a conocer una solución al problema de la discriminación insuficiente de discordancias para la deducción de bases en la resecuenciación a gran escala mediante matrices al variar racionalmente la longitud de las sondas oligonucleotídicas. Para hacer esto, desarrollamos un algoritmo de selección de sondas de longitud variable que desempeña las funciones que permiten  
20 que se alarguen las sondas con una  $T_m$  baja y que se acorten las sondas con una  $T_m$  alta y, al mismo tiempo, se mantenga la posición de la discordancia de cada sonda en el centro termodinámico del oligonucleótido. Esta estrategia difiere de las estrategias anteriores que típicamente varían la secuencia de las sondas, en la que todas las sondas tienen la misma longitud.

25 De acuerdo con la invención, una realización da a conocer un método para producir una micromatriz que tiene una serie de rasgos que comprenden una multitud de sondas oligonucleotídicas de longitud variable que tienen un nucleótido discordante respecto a una secuencia de referencia, comprendiendo el método las etapas de:

30 seleccionar una multitud de sondas de longitud variable para matrices de oligonucleótidos, en donde las sondas de longitud variable se obtienen con la introducción de una longitud de sonda mínima, una longitud de sonda máxima y una temperatura de fusión teórica para una multitud de sondas para matrices de oligonucleótidos en un algoritmo de selección de sondas; en donde el algoritmo es capaz de calcular longitudes de sondas oligonucleotídicas variables, la  $T_m$  teórica de cada sonda se divide por la mitad, para que cada porción de la sonda a cada lado de la discordancia pueda variar de longitud para que permanezca dentro de los parámetros de longitud especificados que pueden alcanzar la mitad de la  $T_m$  teórica especificada, en donde todas las sondas muestran las características siguientes: (i) temperatura de fusión similar y (ii) la posición de la discordancia en cada sonda se localiza en el centro termodinámico aproximado de cada sonda; y  
35

40 sintetizar las sondas oligonucleotídicas inmovilizadas en un sustrato para producir la micromatriz, en donde la longitud de las sondas se determina mediante la etapa de selección de tal forma que la longitud de las sondas en algunos rasgos es diferente de la longitud de las sondas en otros rasgos. Tal matriz se fabrica mejor mediante un instrumento de síntesis de matrices sin máscara (MAS, por su nombre en inglés) que permite producir una micromatriz con una serie de rasgos que contienen sondas oligonucleotídicas de longitud variable.

45 Para utilizar una terminología común de principio a fin, utilizaremos sonda para referirnos a las moléculas de ácido nucleico normalmente monocatenarias sintetizadas como parte de la micromatriz y típicamente inmovilizadas de la misma forma a un sustrato. Las sondas tienen secuencias definidas y todas las sondas en un región definida tienen una secuencia común, denominándose dicha región de sondas de la misma secuencia común un rasgo de la micromatriz. Obviamente, de acuerdo con la presente invención, las sondas serán de longitud diferente, pero normalmente las sondas en un mismo rasgo tendrán la misma longitud. Utilizando el instrumento de MAS, es práctico construir micromatrices con sondas que llegan hasta los 75 a 100 nucleótidos de longitud con una fidelidad utilizable.

50 También se hace referencia a un método para solventar la discriminación insuficiente de las discordancias para la deducción de bases variando de modo racional la longitud de las sondas oligonucleotídicas mediante un algoritmo de selección de sondas de longitud variable que selecciona sondas que tienen una  $T_m$  similar y, al mismo tiempo, se mantiene la posición de la discordancia de la sonda en el centro termodinámico del oligonucleótido. Para realizar estas dos funciones simultáneamente, el método utiliza un algoritmo que es capaz de aceptar los datos de entrada del usuario para especificar las longitudes mínima y máxima de las sondas, y la  $T_m$  teórica para todas las sondas de  
55 la matriz. A continuación, la  $T_m$  teórica de la sonda se divide por la mitad, y a cada trozo de sonda de cada lado de la discordancia se le varía su longitud, de manera que permanezca dentro de los parámetros de longitud especificados y que alcance la mitad de la  $T_m$  teórica especificada. Para calcular la  $T_m$  de la sonda se utiliza la

ecuación siguiente:

$$T_m \text{ de la sonda} = 5 * (G_n + C_n) + 1 * (A_n + T_n),$$

en la que n es el número de cada base oligonucleotídica específica (G, C, A o T) presente en la sonda. Sin embargo, se contempla que cualquier ecuación para determinar la temperatura de la sonda podría utilizarse junto con la nueva estrategia para variar la longitud de la sonda para mejorar la discriminación de discordancias para la deducción de bases.

También se hace referencia a un algoritmo de selección de sondas de longitud variable para el diseño de matrices de oligonucleótidos para los proyectos de resecuenciación de genomas. El algoritmo de selección de sondas de longitud variable se usa para producir sondas para matrices que tienen todas una  $T_m$  similar y se mantiene la posición de la discordancia en el centro termodinámico aproximado de cada oligonucleótido, donde provoca más desestabilización desde el punto de vista teórico.

Los ejemplos siguientes se dan a conocer como ilustraciones adicionales no limitantes de realizaciones particulares de la invención.

### Ejemplos

#### 15 Secuenciación y resecuenciación de 10 000 bases del genoma del coronavirus del SARS

Para resecuenciar aproximadamente 10 000 bases del genoma del coronavirus del SARS, el ADN genómico del virus del SARS se amplificó, se digirió y se marcó como se describe a continuación. Por ejemplo, 1  $\mu\text{g}$  de S SARS se amplificó en varias reacciones mediante PCR múltiplex, lo que dio lugar a 1  $\mu\text{g}$  de ADN amplificado. Para cada hibridación de matrices, se digirieron 100 ng del ADN genómico amplificado con 0,1 U de ADNasa I (Amersham Biosciences, Piscataway, NJ) y tampón One-Phor-All (Amersham) a 1 X en un volumen total de 20  $\mu\text{l}$ . La ADNasa I se inactivó mediante incubación a 97 °C durante 15 minutos. La muestra se marcó en el extremo con 1  $\mu\text{l}$  de biotina- $\text{N}^6$  ddATP (Perkin Elmer, Wellesley, MA) y 25 U de la desoxinucleotidil-transferasa terminal (Promega, Madison, WI) a 37 °C durante 90 minutos, y la transferasa terminal se inactivó mediante incubación a 97 °C durante 15 minutos.

Antes de aplicar la muestra marcada en la matriz, las matrices de resecuenciación se prehibridaron con tampón de resecuenciación de NimbleGen a 1 X (NimbleGen Systems, Madison, WI). Las muestras se calentaron a 95 °C durante 5 minutos, se calentaron a 45 °C y se centrifugaron durante 5 minutos a  $> 12\ 000\ \text{g}$ . A continuación, cada muestra de ADN marcado fue aplicada a las matrices, que se colocaron en una cámara de hibridación a medida y se incubaron a 45 °C durante 14 a 16 horas en un horno giratorio. Luego, las matrices se lavaron con el tampón de lavado no riguroso (SSPE a 6 X, Tween-20 al 0,01% [v/v]), seguido de dos lavados de 5 minutos en el tampón de lavado riguroso (MES a 100 mM, NaCl a 0,1 M, Tween-20 al 0,01% [v/v]) a 47 °C. Las matrices se tñieron con una solución que contenía el conjugado Cy3-estreptavidina (Amersham Biosciences, Piscataway, NJ) durante 10 minutos, y se lavó de nuevo con el tampón de lavado no riguroso. La señal de Cy3 se amplificó mediante la marcación secundaria del ADN con anti-estreptavidina de cabra biotilado (Vector Laboratories, Burlingame, CA). El anticuerpo secundario se retiró por lavado con el tampón de lavado no riguroso y la matriz se volvió a teñir con la solución de Cy3-estreptavidina. Finalmente, se retiró la solución de tinción y la matriz se lavó en un tampón de lavado no riguroso seguido de un lavado de 30 s en SSC a 0,5 X y un lavado de 15 s en etanol al 70%. Las matrices se secaron por centrifugación en una centrífuga a medida y se conservaron antes de escanearlas.

A continuación, las micromatrices se escanearon a una resolución de 5  $\mu\text{m}$  con el escáner Genepix<sup>®</sup> 4000b (Axon Instruments, Inc., Union City, CA). La imagen se interpoló y el tamaño se escaló 2,5x con el programa informático NIH Image (<http://rsb.info.nih.gov/nih-image/>), y la intensidad de los píxeles se extrajo con el programa informático NimbleScan<sup>™</sup> (NimbleGen Systems, Inc. Madison, WI). La deducción de la secuencia se realizó sobre la base de un análisis estadístico de las intensidades de hibridación combinando los datos de ambas hebras con un algoritmo nuevo preferente de aprendizaje automatizado que utiliza una técnica de clasificación denominada de los K vecinos más cercanos. Las matrices se escanearon luego y las imágenes se extrajeron como se describe más arriba.

#### 45 Comparación de las sondas de longitud única frente a variable

El algoritmo de selección de sondas de longitud variable se analizó al comparar longitud de sonda única frente al algoritmo de selección de sondas de longitud variable variando las  $T_m$  de las sondas diana al diseñar las sondas para resecuenciar aproximadamente 10 000 bases del genoma del coronavirus del SARS.

El ADN genómico del SARS se amplificó, se digirió, se marcó y se resecuenció como se describe más arriba. Luego se hibridó la muestra a una matriz. Téngase en cuenta que el número de conjuntos de matrices en un experimento concreto puede variar en consonancia con el tamaño del genoma que se analiza.

Para determinar la mejor técnica de selección de sondas, se calculó el número de bases no amoldables para cada criterio de selección de sondas. Tal y como se hace referencia en la presente memoria, una «base amoldable» es una base que, mediante la matriz, se deduce o asigna que es idéntica a la de la base de la secuencia del virus de referencia. De igual forma, tal y como se utiliza en la presente memoria una «base no amoldable» es una base que

se deduce o asigna que es diferente de la correspondiente base en la secuencia de referencia. Por lo tanto, si el sistema se comporta perfectamente, todas las bases serán amoldables con respecto a la referencia. El algoritmo de selección de sondas que produzca el número más bajo de bases no amoldables producirá los mejores resultados experimentales. La siguiente tabla recoge los resultados de resecuenciación reales utilizando varios criterios de selección de sondas:

5

TABLA 1

Longitud de la sonda	Deducción de bases no amoldables
29-meros	203
33-meros	45
37-meros	37
29- a 39-meros Tm92	14
29- a 39-meros Tm86	12
29- a 39-meros Tm78	8
29- a 39-meros Tm72	7

La tabla 1 muestra claramente que el algoritmo de selección de sondas de longitud variable es mejor que las sondas de longitud uniforme. Además, estableciendo la Tm teórica a 72 °C, se observó el número más bajo de deducción de bases no amoldables, lo que da lugar a la mejor o más exacta discriminación de discordancias para la deducción de bases de los criterios analizados.

10

Además, los solicitantes contemplan que el algoritmo de selección de sondas de longitud variable para el diseño de matrices de oligonucleótidos se pueda utilizar en cualquier proyecto de resecuenciación, tal como la resecuenciación genómica comparativa, utilizado para identificar mutaciones en especies haploides o en una preparación haploide de un genoma diploide. La invención también puede ser útil para los grandes proyectos de resecuenciación que utilizan BAC (cromosomas artificiales bacterianos, que tienen un tamaño medio de 100 kb), plásmidos, fósidos, YAC (cromosomas artificiales de levadura) o cualquier otra genoteca o preparación que da lugar al aislamiento de trozos haploides de un genoma.

15

20

## REIVINDICACIONES

1. Método para producir una micromatriz que tiene una serie de rasgos que comprenden multitud de sondas oligonucleotídicas de longitud variable que tienen una secuencia nucleotídica idéntica a una secuencia de referencia excepto para una discordancia respecto a la secuencia de referencia, comprendiendo el método las etapas de:
- 5           seleccionar multitud de sondas de longitud variable para matrices de oligonucleótidos, en donde las sondas de longitud variable se obtienen introduciendo una longitud de sonda mínima, una longitud de sonda máxima y una temperatura de fusión teórica para las sondas para matrices de oligonucleótidos en un algoritmo de selección de sondas; en donde el algoritmo es capaz de calcular sondas oligonucleotídicas de longitud de variable, la  $T_m$  teórica se divide por la mitad, por lo que la longitud de cada porción de la sonda a cada lado de la discordancia se puede
- 10          variar de tal forma que permanezca dentro de los parámetros de longitud especificados para alcanzar la mitad de la  $T_m$  teórica especificada, en donde todas las sondas muestran las características siguientes: (i) temperatura de fusión similar y (ii) posición de discordancia en cada sonda localizada en el centro termodinámico aproximado de cada sonda; y
- sintetizar las sondas oligonucleotídicas inmovilizadas a un sustrato para producir la micromatriz, en donde la
- 15          longitud de las sondas se determina seleccionando la etapa de tal forma que la longitud de las sondas de algunos rasgos es diferente de la longitud de las sondas de otros rasgos.
2. Método de acuerdo con la reivindicación 1, en el que la temperatura de fusión ( $T_m$ ) teórica se consigue al incrementar la longitud de las sondas que tienen una  $T_m$  baja y al disminuir la longitud de las sondas que tienen una  $T_m$  alta respecto a una  $T_m$  media de las sondas.
- 20          3. Método de acuerdo con la reivindicación 1 o 2, en el que la  $T_m$  teórica para cada una de la multitud de sondas para matrices se divide por la mitad, y cada mitad de sonda a cada lado de la posición de la discordancia varía su longitud de forma independiente, de tal forma que permanece dentro de los parámetros de longitud máxima y mínima de la sonda y alcanza la mitad de la  $T_m$  teórica.
- 25          4. Método de acuerdo con la reivindicación 1, 2 o 3, en el que la  $T_m$  de la sonda se calcula con la fórmula:  $T_m = 5 * (G_n + C_n) + 1 * (A_n + T_n)$ , donde la n es el número de cada base específica presente en la sonda.
5. Método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que la multitud de sondas de longitud variable para matrices de oligonucleótidos se utiliza en la resecuenciación de genomas basada en matrices.
6. Método de acuerdo con una cualquiera de las reivindicaciones anteriores, en el que la síntesis de la micromatriz se realiza en un instrumento de síntesis de matrices sin máscara.

30