

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 605 364**

51 Int. Cl.:

C12Q 1/68

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **05.09.2007** **PCT/EP2007/059299**

87 Fecha y número de publicación internacional: **13.03.2008** **WO08028931**

96 Fecha de presentación y número de la solicitud europea: **05.09.2007** **E 07803261 (2)**

97 Fecha y número de publicación de la concesión europea: **31.08.2016** **EP 2059609**

54 Título: **Código morse genómico**

30 Prioridad:

07.09.2006 US 516673

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

14.03.2017

73 Titular/es:

**INSTITUT PASTEUR (50.0%)
25-28 rue du Docteur Roux
75724 Paris Cédex 15, FR y
GENOMIC VISION (50.0%)**

72 Inventor/es:

**LEBOFSKY, RONALD;
BENSIMON, AARON y
WALRAFEN, PIERRE**

74 Agente/Representante:

CURELL AGUILÁ, Mireia

ES 2 605 364 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

Código morse genómico.

5 **Campo de la invención**

La presente invención se refiere a un método de detección de la presencia de por lo menos un dominio de interés en una macromolécula que se debe someter a prueba.

- 10 Más particularmente, también se describe un kit para la detección de la presencia de por lo menos dos dominios de interés en una macromolécula, que comprende por lo menos dos conjuntos de tres sondas, pudiendo cada conjunto de sondas unirse específicamente a un dominio, en el que el dominio de un conjunto de por lo menos tres sondas en la macromolécula forma una secuencia de por lo menos dos códigos seleccionados de entre un grupo de por lo menos dos códigos diferentes (por ejemplo, "puntos" y "rayas"), siendo dicho grupo idéntico para todos los dominios,
- 15 siendo dicha secuencia de códigos específica de un dominio y siendo la firma de dicho dominio.

Antecedentes

- 20 El estudio de las macromoléculas, en particular de las biológicas (más específicamente, el ADN), requiere a menudo marcar de forma precisa algunos dominios, ya sea para fines "cartográficos", es decir, para estudiar la organización espacial de estos dominios, o para el fin de localizar la posición, en la macromolécula, de una reacción o un conjunto de reacciones químicas o bioquímicas.

- 25 En 1997, Michalet et al (Science 1997) propusieron usar el peinado molecular dinámico a fin de llevar a cabo estudios de alta resolución. Más tarde, Anglana M. et al. (Cell, 2003) analizaron un locus amplificado que contiene OriGNAI3 usando el análisis tricolor de moléculas de ADN peinadas. Finalmente, Lefobsky et al. establecieron y compararon parámetros clave de los programas de replicación del gen de ARNr humano (Mol. and Cellular Biology, 2005), y estudiaron el origen de la replicación del ADN en células humanas (Mol. Biology, of the Cell, 2006).

- 30 Los métodos que permiten la observación de la organización espacial de secuencias de ARN (FISH sobre fibras, peinado molecular, etc.) requieren en numerosas aplicaciones que se marquen como puntos de referencia algunas regiones, es decir, que se marquen de tal manera que permita la identificación de regiones específicas a través de alguna técnica de detección. Este es el caso para las aplicaciones cartográficas, en las que el asunto principal abordado es la posición relativa de varias regiones, así como aplicaciones en las que se estudia un fenómeno biológico en un locus específico (o varios locus). Los dominios se pueden identificar entonces mediante marcadores
- 35 específicos, de hecho generalmente sondas de ADN complementarias a las secuencias de interés (denominadas dominios de interés). Estos marcadores se pueden detectar, por ejemplo, mediante microscopía de fluorescencia, autorradiografía, etc.

- 40 Cualquiera que sea la técnica de detección usada, el método habitual para marcar un punto de referencia consiste en sintetizar sondas, es decir, secuencias complementarias a las regiones de interés acopladas a elementos que permiten la detección (fluorocromos, radioelementos, etc.). Para distinguir varias regiones, el método principal es diseñar sondas que se pueden distinguir en base a sus propiedades intrínsecas tales como su longitud o la naturaleza de los elementos detectables (típicamente, fluorocromos de diferentes colores). En este caso, las sondas
- 45 son distinguibles debido a que, incluso cuando se detectan individualmente, se pueden identificar: son de naturaleza diferente.

- Este enfoque alcanza rápidamente su límite cuando se trata de distinguir un gran número de regiones: la longitud de las sondas sufre restricciones técnicas, y el número de elementos detectables diferentes es a menudo limitado.

- 50 De hecho, a menudo es necesario marcar varios dominios simultáneamente de una manera diferenciable. Los esfuerzos para desarrollar marcadores diferenciables consistieron principalmente en multiplicar los marcadores de "naturaleza" diferente, es decir, que se pueden diferenciar individualmente por el método de detección usado: acoplamiento de fluorocromos de diferentes espectros en microscopía con fluorescencia, uso de diferentes
- 55 longitudes de sondas en microscopía con fluorescencia y autorradiografía. El límite principal con estos enfoques es que el número de marcadores diferentes así obtenidos es limitado. Además, la detección simultánea de marcadores de naturaleza diferente obliga generalmente a usar métodos de adquisición delicados y largos de implementar.

Sumario de la invención

- 60 La invención se refiere a un método de detección de la presencia y/o posición de por lo menos un dominio de interés en una macromolécula que se debe someter a prueba, en el que dicho método comprende las siguientes etapas:

- 65 a) determinar previamente por lo menos tres regiones diana en el dominio de interés, diseñar y obtener sondas marcadas correspondientes de cada región diana, denominadas conjunto de sondas del dominio de interés, escogiéndose la posición de estas sondas, unas comparadas con las otras, y formando una secuencia de por

lo menos dos códigos escogidos entre un grupo de por lo menos dos códigos diferentes, siendo dicha secuencia de códigos específica de dicho dominio de interés en el ácido nucleico que se debe someter a prueba,

- b) tras extender el ácido nucleico que se debe someter a prueba sobre el que se unen las sondas obtenidas en la etapa a), detectar las señales correspondientes a cada una de las sondas marcadas unidas sobre el ácido nucleico linealizado,
- c) transcribir dichas señales en un código de secuencia, en el que dicho código resulta de la lectura del espaciamiento entre señales consecutivas de las sondas, indicando la detección de la secuencia de códigos de un dominio de interés la presencia de dicho dominio de interés en el ácido nucleico que se debe someter a prueba, y al contrario, indicando la ausencia de detección de secuencia de códigos o parte de secuencia de códigos de un dominio de interés la ausencia de dicho dominio o parte de dicho dominio de interés en el ácido nucleico que se debe someter a prueba.

En una forma de realización, la macromolécula es ácido nucleico, particularmente ADN, más particularmente ADN bicatenario.

En una forma de realización adicional, el conjunto de sondas comprende además dos sondas (sonda 1 o sonda 2), cada sonda capaz de unirse en un extremo diferente del dominio de interés, permitiendo la lectura de la señal de una de dicha sonda 1 o sonda 2 asociada con su sonda consecutiva en el dominio de interés, denominado "pareja de sondas de extremo de comienzo o final", obtener una información de comienzo o final de la lectura.

También se describe un kit para la detección de la presencia de por lo menos dos dominios de interés en una macromolécula, que comprende por lo menos dos conjuntos de tres sondas, siendo capaz cada conjunto de sondas de unirse específicamente a un dominio, en el que la unión de un conjunto de por lo menos tres sondas en la molécula forma una secuencia de por lo menos dos códigos escogida entre un grupo de por lo menos dos códigos diferentes (por ejemplo, "puntos" y "guiones"), siendo dicho grupo idéntico para todos los dominios, siendo dicha secuencia de códigos específica de un dominio y siendo la firma de dicho dominio.

Breve descripción de los dibujos

Figura 1: representa el cartografiado de iniciación de la replicación del ADN en 1,5 Mb en el cromosoma humano 14q11.2.

(A) Tres tipos de señales de replicación en ADN peinado que indican un origen. La iniciación se produce antes de los períodos de marcaje, durante el pulso de IdU (azul), y durante el pulso de CldU (rojo) que da lugar a las señales mostradas en (i), (ii), y (iii) respectivamente. En los tres casos, se supone que el punto medio de las tres pistas es el sitio de iniciación.

(B) Estrategias de hibridación en ADN peinado. Dos sondas de igual longitud pero detectadas con diferentes colores (i), o dos sondas de diferente longitud detectadas con el mismo color (ii), se hibridan para visualizar una región genómica de interés. Como alternativa, se pueden usar espacios entre los conjuntos de sondas para proporcionar la misma información. En (iii), cuatro sondas cortas se hibridan dando lugar a tres espacios informativos. El espacio 3 permite que la molécula se oriente durante la ruptura (iv, v). Las rupturas de ADN se representan mediante un par de líneas continuas verticales.

(C) Código Morse Genómico (GMC) que cubre 1,356 Mb en el cromosoma humano 14q11.2. Se proporcionan los patrones lineales de los primeros cuatro símbolos en Código Morse, A, B, C y D. El GMC comprende estos cuatro símbolos, cada símbolo representado por un conjunto de sondas. Las sondas se muestran en verde. Los espacios codificantes son espacios cortos y espacios largos. Se incluyeron el espacio de comienzo y el espacio de final para ayudar a orientar los símbolos cuando el ADN se rompe. Los símbolos están separados por espacios. Los tamaños de la sonda y del espacio en kb se dan encima de cada uno, respectivamente.

(D) Ejemplos de sucesos de iniciación observados en la región de GMC. Las flechas blancas indican el sitio de iniciación. Para las fibras 2, 7, 8, y 13, el GMC es todavía descodificable, incluso aunque todas las sondas que pertenecen a un símbolo no estén presentes. Los sucesos de iniciación que flanquean a los símbolos se cartografiaron cuando uno de los símbolos adyacentes se descodificó (moléculas 1, 5, 11, y 14) o cuando la información del espacio estuvo disponible (moléculas 4 y 10). (bar = 100 kb).

Figura 2: muestra la identificación de la zona de iniciación mediante análisis de agrupamientos.

(A) Definiendo agrupamientos ideales. Las ecuaciones para dentro de la varianza (W) y entre la

varianza (B) se muestran en el recuadro. Para W, N es el número de agrupamientos y V_i es la varianza del agrupamiento i . Para B, C_i es el centroide del agrupamiento i , y C es la media de todos los centroides en un conjunto de agrupamientos. La representación de B – W como función del número de agrupamientos reveló valores máximos cuando el dato se dividió en 9 (círculo azul), 22 (círculo verde), y 45 (círculo rojo) grupos.

(B) La ruptura de un agrupamiento en sus componentes cuando se aplican 9 (línea azul), 22 (línea verde) y 45 (línea roja) particiones a los datos. Las regiones por debajo de las líneas rojas representan zonas de iniciación, que se ilustran mediante las cajas blancas encima de las moléculas. Las flechas blancas indican el sitio de iniciación. (bar = 100 kb).

(C) Distribución de zonas de iniciación en la región de GMC. Las líneas rojas horizontales representan el tamaño y la posición de la zona. Las líneas blancas verticales designan las posiciones de los 307 sitios de iniciación cartografiados. Las zonas que contienen menos de 4 puntos de datos se marcan mediante una línea gris horizontal en lugar de una roja, y no son considerados en los análisis subsiguientes. El agrupamiento en el panel B de esta figura está indicado por un asterisco. (bar = 100 kb).

Figura 3: muestra el análisis espacio-temporal de orígenes funcionales.

(A) Señales de replicación que proporcionan distancias entre orígenes (X). En (i), las pistas de replicación de dos sitios de iniciación permanecen separadas. En (ii) y (iii), las horquillas que se aproximan se funden durante los pulsos de IdU y CldU, respectivamente.

(B) Histograma que muestra la frecuencia de distancias entre orígenes medidas.

(C) Ejemplos de moléculas con por lo menos dos sucesos de iniciación en los símbolos A, B (i) y C, D (ii). Las flechas blancas indican el sitio de iniciación. Las zonas de iniciación están marcadas mediante líneas rojas horizontales. Para moléculas individuales, la zona de iniciación a partir de la que arranca un origen está indicada por una caja blanca. Las cajas oscuras designan zonas de iniciación silenciosas. (bar = 100 kb).

Figura 4: muestra la interferencia del origen en base a la extensión de la horquilla.

(A) Las horquillas que se alargan desde el origen activo en la zona (v) cubren la región unida por los pares de líneas verticales. Se observa la terminación de la horquilla que se mueve hacia la izquierda (triángulos en blanco invertidos). Las zonas de iniciación (i) y (ii) no se interfieren, ya que la horquilla desde el origen en cuestión no se extiende hasta sus fronteras. Se suprimen las zonas (iii), (iv) y (vi), ya que la horquilla replica de forma pasiva todas sus longitudes. La horquilla que se mueve hacia la derecha penetra en la zona (vii), pero no alcanza su centroide (punto negro). Esta zona no está incluida en el dato de interferencia del origen. Los rectángulos grises designan zonas de iniciación, y el rectángulo gris oscuro designa zonas de iniciación suprimidas.

(B) Ejemplos de moléculas que presentan interferencia del origen. Las zonas de iniciación relevantes para esta figura se ilustran mediante las líneas rojas horizontales. La caja blanca marca la zona de iniciación a partir de la que arrancan los orígenes (flechas blancas). Las cajas negras indican zonas de iniciación que están suprimidas debido a la extensión de la horquilla (bar = 100 kb).

(C) Histograma que muestra la frecuencia de distancias entre un suceso de iniciación y zonas con las que interfiere. Las zonas que se suprimieron por horquillas que se mueven centroméricas y teloméricas se representan mediante valores negativos y positivos, respectivamente.

Figura 5: muestra tres tipos de códigos

A: código basado en el espaciado entre sondas consecutivas

B: código basado en el marcaje (más particularmente el color)

C: código redundante basado en el marcador (color) + espaciado entre sondas consecutivas

Tres tipos de codificación. Las sondas se ilustran mediante un segmento corto de color. Estas sondas pueden medir, por ejemplo, 5 a 10 Kb. Los "colores" representados en la presente memoria se deben de incluir/entender como sondas de diferente naturaleza, pueden diferir por el fluorocromo asociado, pero también por su tamaño o cualquier otro carácter o combinación de caracteres – en particular, en sí no es una cuestión de sondas fluorescentes -. El espaciado entre las sondas puede ser, por ejemplo, de aproximadamente 20 KB para espacios pequeños, 40 KB para los grandes, y menos de

10 KB para los dobletes (en el primer caso). La codificación redundante es una combinación de los otros dos.

Figura 6: representa el interés del código redundante

Interés de codificación redundante. Una supresión en la parte de la secuencia marcada de una flecha azul se ilustra para los 3 tipos de codificación 1 descrita. En el primero de ambos casos, la modificación de la razón relacionada con la supresión implica confusión con otra secuencia de códigos. En el tercer caso, la secuencia de códigos obtenida es incoherente y de este modo parece claramente que hubo un reordenamiento. La naturaleza de este reordenamiento (una supresión) se identificará posiblemente según el contexto.

Figura 7: muestra código de solapamiento

Codificación de solapamiento. Tres naturalezas de sondas se representan mediante las letras A, B y C. Se muestra en la parte superior cómo, ordenando las diversas secuencias de códigos formadas de 7 sondas consecutivas de manera que las 6 últimas sondas de la primera corresponden a las 6 primeras de la siguiente, es posible obtener una sucesión de sondas en la que cualquier sucesión de 7 letras es única. En la parte inferior, una sucesión de 2193 letras A, B y C, de manera que todas las posibles razones para 7 letras son representadas solamente una vez. Así, si por ejemplo se detecta un fragmento que incluye/comprende la secuencia de códigos abaaccb, se sabe que resulta del área en rojo.

Descripción detallada

Los inventores han descubierto un nuevo enfoque para el diseño de sondas, que permite la distinción de un número virtualmente infinito de regiones, usando muy pocas sondas de diferente naturaleza – o incluso una única naturaleza de sondas. En este enfoque, se realizan puntos de referencia de varias sondas que juntas constituyen un símbolo único para cada punto de referencia, que es una secuencia de códigos, obteniéndose un código mediante la lectura de una sucesión de dos sondas. Este motivo (secuencia de códigos) puede combinar la naturaleza y la posición de las sondas.

La presente invención proporciona varios ejemplos de aplicaciones de este enfoque, uno de los cuales se ha usado con éxito para el estudio de la replicación en un locus dado. En este último caso, se ha usado una única naturaleza de sonda. También se muestra cómo es posible, con muy pocas naturalezas diferentes de sondas, señalar con una resolución elevada regiones del tamaño de todos los cromosomas (por ejemplo, 3 naturalezas de sondas permiten señalar 600 Mb con una resolución de 40 kb).

La presente invención consiste en marcar dominios de una forma diferenciable usando marcadores de solamente una naturaleza, pero basada espacialmente en el dominio de interés de manera que forman una única sucesión de sondas para cada dominio, que forman una única secuencia de códigos para cada dominio. Debido a que una de las posibles implementaciones de este método consiste en adaptar el código Morse usando una letra (un código) por dominio en una porción del genoma, se le proporcionó el nombre de Código Morse Genético, sustituyéndose los "puntos" y "guiones" del código Morse por marcadores separados respectivamente por intervalos cortos o largos.

Puede ser necesario, o más eficaz, localizar de una forma inequívoca el comienzo y el final de la secuencia específica de códigos de un dominio. Esta localización se puede obtener por ejemplo usando un "doblete" de marcadores (pareja de sondas de extremo de comienzo o final, marcándose dichas sondas con marcadores (etiquetas), es decir, dos marcadores separados por un intervalo más corto que todos aquellos usados dentro de la secuencia específica de códigos de un dominio, para marcar el comienzo y el final de la lectura de una secuencia de códigos. Los intervalos cortos en los "dobletes" que marcan el comienzo y el final de la secuencia de códigos pueden ser de diferentes longitudes, permitiendo la orientación de la secuencia de códigos.

De forma general, el presente método consiste en puntuar una macromolécula de puntos de referencia constituidos de un cierto número de marcadores de solamente una naturaleza dispuestos espacialmente de manera que forman una secuencia de códigos específicos para cada punto de referencia, estando los marcadores relacionados directamente con la macromolécula estudiada.

En el caso en el que los marcadores sean todos idénticos, son posibles muchos modos de detección de los marcadores (autorradiografía, centelleo, fluorescencia, quimioluminiscencia, etc.). Se puede así obtener un número prácticamente infinito de secuencias de códigos, cada una específica de un dominio, siendo el límite solamente la resolución espacial del método usado y la latitud en el posicionamiento específico de los marcadores.

Obviamente, según la presente invención, si se pueden usar marcadores de diferentes naturalezas, incluso en un número limitado, la secuencia de códigos se puede obtener mediante combinación de estos marcadores diferentes. Las posibilidades de codificar los puntos de referencia son entonces incluso más numerosas.

El Ejemplo 1 de la presente solicitud se refiere al estudio de un fenómeno biológico (de hecho, la replicación del ADN) en un locus preciso delimitado por estos puntos de referencia. Se ha marcado un dominio de 1,5 Mb usando 4 secuencias de códigos, directamente inspiradas por el Código Morse. Cada uno de estas cuatro secuencias de

5 códigos está constituida de 5 o 7 sondas de solamente una naturaleza (solamente un fluorocromo y longitudes idénticas) en un número variable y espaciadas de una manera variable, detectadas mediante fluorescencia (figura 5a). Cada comienzo de un dominio de interés se marca mediante una "pareja de sondas de extremo de comienzo" separada, estando cada sonda de cada pareja espaciada por 17 kb, mientras que cada final de un dominio de

10 interés está marcado por una "pareja de sondas de extremo de final" separada, estando cada sonda de cada pareja espaciada por 6 kb. Los códigos que corresponden a las primeras 4 letras del código Morse, consistiendo los puntos en sondas separadas por pequeñas distancias (aprox. 25-30 kb), y los guiones por distancias largas (55-70 KB).

También es posible codificar las balizas mientras se sigue el Código Morse usando marcadores de dos naturalezas diferentes, espaciados regularmente (figura 5b). Por ejemplo, en fluorescencia, se puede dar un color específico a los "puntos" del Morse, y otro color específico para los "guiones". Este método tiene la ventaja de ser más "compacto", es decir, toda la secuencia que forma el código está contenida en una secuencia global más corta, que puede tener ventajas.

Como en el caso anterior, se puede desear marcar el comienzo y el final de una secuencia de códigos específica de un dominio. Se puede usar un doblete (o pareja) de dos sondas marcadas cada una con un color diferente, que se puede dirigir sistemáticamente en la misma dirección. Este doblete da entonces la dirección de lectura de la secuencia de código.

También se pueden combinar los dos sistemas previamente descritos (figura 5c): los "puntos" están formados por dos sondas del mismo color separadas por un espacio corto, los "guiones" están formados por dos sondas de diferente color separadas por una distancia larga. La redundancia de los dos sistemas de codificación permite una mayor robustez.

En particular, en aplicación cartográfica de la invención, por ejemplo diagnóstico molecular, se esperan variaciones de secuencias, que pueden modificar la secuencia de códigos formada por las sondas: pérdida o repetición o espacios entre las sondas modificadas. Si se modifica así una de las secuencias de códigos, con solamente un sistema de codificación en el mejor de los casos se haría imposible descodificar la secuencia de códigos, y en el peor de los casos podría confundirse con otra. Si se usan los dos sistemas redundantes, la confusión con otra secuencia de códigos se hace muy poco probable, y en el mejor de los casos se puede identificar la secuencia de

35 códigos y la variación de secuencias se puede definir claramente (figura 6).

Una de las desventajas del enfoque del diseño de sondas descrito en esta invención es la posibilidad de crear un número muy grande de sondas con marcadores de unos pocos "colores" diferentes.

Se describen algunas implementaciones que destacan esta característica: así, con 7 sondas creadas usando solamente 3 colores, según la presente invención, es posible crear 2187 ($= 3^7$) códigos diferentes, con una longitud y espaciamiento uniformes de las sondas. Además, es posible ordenar estos 2187 códigos de manera que las 6 últimas sondas que constituyen un código corresponden a las 6 primeras sondas del siguiente código.

Si estos códigos se superponen, es posible obtener una sucesión de 2193 sondas en la que 7 sondas consecutivas definen un código único, por lo tanto un sitio único en la secuencia en la que están colocadas estas sondas, y por lo tanto un dominio específico de interés en la macromolécula.

De este modo, es posible, por ejemplo, marcar más de 100 Mb mientras se coloca una sonda cada 50 kb (por ejemplo sondas de 5 kb, en tres colores o pseudo-colores). La posición original (posición en la secuencia de 100 Mb) de cada fragmento de ADN en el que será posible detectar por lo menos 7 sondas (de este modo, todos los fragmentos de más de 400 kb) se identificará claramente. La secuencia en la figura 7, compuesta de 3 letras A, B y C que corresponde a 3 colores de sondas asciende a 2193 letras, y cada sucesión de 7 letras es única.

Las ventajas de esta implementación son por un lado que optimiza la relación entre el número de sondas, la resolución y la proporción de los fragmentos explotables (es decir, aquellos que se pueden relocalizar de una manera no ambigua en la secuencia total), y por otro lado, que esta implementación es fácilmente automatizable, siendo la sucesión de las sondas fácil de localizar mediante procesamiento de datos.

Una solución con el problema de orientación de los fragmentos de secuencias simétricas es periódicamente introducir (por ejemplo cada 7 sondas) una sonda "orientada" en la sucesión de sondas descrita anteriormente, independientemente de la simetría. Por ejemplo, es posible sustituir cada séptima sonda mediante un doblete que da al mismo tiempo el "color" de la sonda y su orientación. Un doblete estaría formado por dos sondas muy cercanas en el espacio. De este modo, es posible sustituir una "A" por un doblete "AB", una "B" por un doblete "BC", y "C" por un doblete "CA".

Existen numerosas ventajas para la presente invención. De hecho, en la técnica anterior, los esfuerzos de desarrollo consistieron hasta entonces en multiplicar el número de diferentes elementos detectables. En particular, esto condujo a usar detección por fluorescencia como el método de detección principal, debido a que solamente la fluorescencia permite el uso de muchos colores diferentes para etiquetar las sondas. En este contexto, el uso de un conjunto limitado de colores permite reducir significativamente los tiempos de adquisición y análisis. Además, la posibilidad de mantener solamente uno o dos elementos detectables diferentes permite el uso de técnicas de detección mucho más rápidas y más fácilmente automatizables. Finalmente, incluso para la detección por fluorescencia, el número de diferentes naturalezas posibles de sondas sigue siendo limitado, mientras que nuestro método permite generar un número virtualmente infinito de puntos de referencia.

En consecuencia, también se describe un método de detección de la presencia y/o la posición de por lo menos un dominio de interés en una macromolécula que se debe someter a prueba, en el que dicho método comprende las siguientes etapas:

- a) determinar previamente por lo menos dos regiones diana en el dominio de interés, diseñar y obtener sondas marcadas correspondientes de cada región diana, denominadas conjunto de sondas del dominio de interés, escogiendo la posición de estas sondas una en comparación con las otras y formando la firma específica de cada dominio de interés en la macromolécula que se debe someter a prueba;
- b) tras extender la macromolécula que se debe someter a prueba en la que están unidas las sondas obtenidas en la etapa a), detectar la posición, una en comparación con las otras, de las sondas unidas a la macromolécula linealizada, detectar la firma de un dominio de interés que indica la presencia de dicho dominio de interés en la macromolécula que se debe someter a prueba, y por el contrario, indicando la ausencia de detección de la firma o parte de la firma de un dominio de interés la ausencia de dicho dominio o parte de dicho dominio de interés en la macromolécula que se debe someter a prueba.

De forma más precisa, la presente invención se refiere a un método de detección de la presencia de por lo menos un dominio de interés en un ácido nucleico que se debe someter a prueba, en el que dicho método comprende las siguientes etapas:

- a) determinar previamente por lo menos tres regiones diana en el dominio de interés, diseñar y obtener sondas marcadas correspondientes de cada región diana, denominadas conjunto de sondas del dominio de interés, escogiéndose la posición de estas sondas una en comparación con las otras y formando una secuencia de por lo menos dos códigos escogidos entre un grupo de por lo menos dos códigos diferentes, siendo dicha secuencia de códigos específica de dicho dominio de interés en el ácido nucleico que se debe someter a prueba;
- b) tras extender el ácido nucleico que se debe someter a prueba en el que están unidas las sondas obtenidas en la etapa a), detectar las señales correspondientes a cada una de las sondas marcadas unidas al ácido nucleico linealizado,
- c) transcribir dichas señales en un código de secuencia, en el que dicho código resulta de la lectura del espaciamiento entre señales consecutivas de las sondas,
- indicando la detección de la secuencia de códigos de un dominio de interés la presencia de dicho dominio de interés en el ácido nucleico que se debe someter a prueba, y por el contrario, indicando la ausencia de detección de secuencia de códigos o parte de secuencia de códigos de un dominio de interés la ausencia de dicho dominio o parte de dicho dominio de interés en el ácido nucleico que se debe someter a prueba.

La posición de estas sondas, una en comparación con las otras, se escoge arbitrariamente y de forma específica para formar la firma específica de dicho dominio de interés en la macromolécula que se debe someter a prueba.

Por arbitrariamente se quiere decir, según la presente invención, que el experto principal en la materia puede seleccionar la posición de las sondas en el dominio de interés, de una manera independiente de cualesquiera condiciones, en particular la secuencia del dominio.

Por específicamente se quiere decir, según la presente invención, que el experto en la materia puede escoger la posición de las sondas en el dominio de interés, de una forma para formar una firma específica del dominio de interés, que permite la determinación de su presencia o ausencia en la macromolécula. Esta firma se escoge por el experto en la materia, siendo la única condición atribuir diferentes firmas para diferentes dominios de interés.

El espaciamiento (espacios) entre dos señales consecutivas se mide por medición directa si se usa por ejemplo microscopía o autorradiografía, o cualquier otro método adecuado, dependiendo del método de detección.

En una forma de realización, la detección de la presencia de un dominio de interés puede ser una detección de firma parcial, por ejemplo con una o más sondas perdidas, o que se unen en una posición errónea (por ejemplo con un

espaciamiento más grande o más pequeño con su sonda consecutiva en el dominio de interés), siendo esta firma parcial la indicación de una alteración del dominio de interés de las macromoléculas. En el caso de que la macromolécula sea ADN, esta alteración puede ser una mutación tal como una supresión, inserción o sustitución de uno o varios nucleótidos en el dominio de interés en la macromolécula.

5 En una forma de realización, el método es preferentemente para la determinación de la presencia y/o posición de por lo menos dos dominios de interés, y comprende en la etapa a) determinar previamente por lo menos tres regiones diana en el dominio de interés.

10 En una forma de realización, la firma de un dominio de interés según la presente invención resulta de la sucesión del espaciamiento entre sondas consecutivas.

En una forma de realización, la posición del dominio de interés según la presente invención se usa como referencia para localizar una reacción química o bioquímica.

15 En una forma de realización, la posición del dominio de interés según la presente invención se usa para establecer un mapa físico en la macromolécula que engloba la región diana.

20 En una forma de realización, la firma del dominio de interés según la presente invención consiste en una sucesión de diferentes sondas marcadas.

En una forma de realización, algunas de las sondas de la región diana según la presente invención también son parte de la firma de por lo menos algún otro dominio de interés situado cerca en la macromolécula.

25 En una forma de realización, la macromolécula es ácido nucleico, proteína, polímero, o hidrato de carbono. Particularmente, la macromolécula es ADN, ADN bicatenario o ADN monocatenario, más particularmente ADN bicatenario en el caso de que se use peinado molecular para la linealización del ADN.

30 Como se usa de forma intercambiable en la presente memoria, las expresiones "oligonucleótidos", "ácidos nucleicos" y "polinucleótidos" incluyen secuencias de ARN, ADN, o secuencias híbridas de ARN/ADN de más de un nucleótido en forma monocatenaria o de dúplex.

35 Como se usa en la presente memoria, la expresión "ácidos nucleicos" y "molécula de ácido nucleico" pretende incluir moléculas de ADN (por ejemplo, ADNc o ADN genómico) y moléculas de ARN (por ejemplo, ARNm), y análogos del ADN o ARN generados usando análogos nucleotídicos. La molécula de ácido nucleico puede ser monocatenaria o bicatenaria, pero preferentemente es ADN bicatenario. A lo largo de la presente memoria descriptiva, la expresión "secuencia nucleotídica" se puede utilizar para designar indistintamente un polinucleótido o un ácido nucleico. Más precisamente, la expresión "secuencia nucleotídica" comprende el propio material nucleico, y de este modo no está restringida a la información de la secuencia (es decir, la sucesión de letras escogidas entre las cuatro letras de bases) que caracteriza bioquímicamente una molécula de ADN o ARN específica. También, se usan intercambiablemente en la presente memoria las expresiones "ácidos nucleicos", "oligonucleótidos", y "polinucleótidos".

45 En el caso de que las macromoléculas sean ADN monocatenario y las sondas sean oligonucleótidos, el término "unión" significa "hibridación".

50 Como se usa en la presente memoria, el término "hibridación", "se hibrida a" o "que se hibrida" pretende describir condiciones para hibridación de restricción moderada o de alta restricción, preferentemente en el que las condiciones de hibridación y de lavado permiten secuencias nucleotídicas por lo menos 60% homólogas entre sí para permanecer hibridadas entre sí.

55 Preferentemente, las condiciones son tales que las secuencias por lo menos de aproximadamente 70%, más preferentemente por lo menos de aproximadamente 80%, incluso más preferentemente por lo menos de aproximadamente 85%, 90%, 95% o 98% homólogas entre sí típicamente permanecen hibridadas entre sí. Las condiciones restrictivas son conocidas por los expertos en la materia, y se pueden encontrar en Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6.

60 Por secuencias nucleicas que tienen un porcentaje de identidad de por lo menos 80%, preferentemente 85%, 90%, 95% y 98%, tras el alineamiento óptimo con una secuencia preferida, se pretende indicar las secuencias nucleicas que tienen, con respecto a la secuencia nucleica de referencia, ciertas modificaciones tales como, en particular, una supresión, un truncamiento, una elongación, una fusión química y/o una sustitución, especialmente sustitución de punto. Preferentemente se refiere a secuencias en las que las secuencias codifican las mismas secuencia de aminoácidos que la secuencia de referencia, estando esto conectado con la degeneración del código genético, o secuencias complementarias que son capaces de hibridarse específicamente con las secuencias de referencia, preferentemente en condiciones de gran restricción, especialmente tales como se definen a continuación.

Una hibridación en condiciones de gran restricción significa que las condiciones de temperatura y las condiciones de fuerza iónica se escogen de tal manera que permiten el mantenimiento de la hibridación entre dos fragmentos de ADN complementario. A título ilustrativo, las condiciones de gran restricción de la etapa de hibridación con el fin de definir los fragmentos polinucleotídicos descritos anteriormente son ventajosamente las siguientes.

La hibridación de ADN-ADN o ADN-ARN se lleva a cabo en dos etapas: (1) prehibridación a 42°C durante 3 horas en amortiguador de fosfato (20 mM, pH 7,5) que contiene 5 x SSC (1 x SSC corresponde a NaCl 0,15 M + disolución de citrato sódico 0,015 M), 50% de formamida, 7% de dodecilsulfato de sodio (SDS), 10 x Denhardt, 5% de sulfato de dextrano, y 1% de ADN de esperma de salmón; (2) hibridación real durante 20 horas a una temperatura que depende del tamaño de la sonda (es decir: 42°C, para un tamaño de sonda > 100 nucleótidos) seguido de 2 lavados de 20 minutos a 20°C en 2 x SSC + 2% de SDS, 1 lavado de 20 minutos a 20°C en 0,1 x SSC + 0,1% de SDS. El lavado final se lleva a cabo en 0,1 x SSC + 0,1% de SDS durante 30 minutos a 60°C para un tamaño de sonda de > 100 nucleótidos. Las condiciones de hibridación de gran restricción descritas anteriormente para un polinucleótido de tamaño definido se pueden adaptar por la persona experta en la técnica a oligonucleótidos de mayor o menor tamaño, según la enseñanza de Sambrook et al., (1989, Molecular cloning: a laboratory manual. 2ª Ed. Cold Spring Harbor).

En una forma de realización, las sondas son oligonucleótidos de por lo menos 15 nucleótidos, preferentemente por lo menos 1 Kb, más preferentemente entre 1 a 10 kb, incluso más preferentemente entre 4 a 10 kb.

Puesto que la resolución máxima en ADN peinado es 1-4 kb, las sondas según la presente invención son preferentemente de por lo menos 4 kb.

En una forma de realización, la linealización de las macromoléculas se realiza antes o después de la unión de las sondas a las macromoléculas.

En una forma de realización, la linealización de la macromolécula se realiza mediante peinado molecular o Fish sobre fibras.

El peinado molecular se realiza según métodos publicados (Lebofsky y Bensimon, 2005). La caracterización física de genomas individuales a lo largo de grandes regiones genómicas es posible con tecnología de peinado molecular. Una matriz de moléculas de ADN individuales peinadas se prepara estirando moléculas unidas por sus extremidades a una superficie de vidrio silanizado con un menisco de aire-agua cóncavo. Al llevar a cabo la hibridación fluorescente sobre el ADN peinado, la posición de la sonda genómica se puede visualizar directamente, proporcionando un medio para construir mapas físicos y, por ejemplo, detectar microrreordenamientos. La replicación del ADN de una sola molécula también se puede monitorizar mediante la detección fluorescente en moléculas de ADN combado de análogos nucleotídicos incorporados.

FISH (hibridación in situ fluorescente) es una técnica citogenética que se puede usar para detectar y localizar secuencias de ADN en los cromosomas. Usa sondas fluorescentes que se unen únicamente a aquellas partes del cromosoma con las que muestran un alto grado de similitud de secuencia. Se puede usar microscopía de fluorescencia para encontrar dónde la sonda fluorescente se une al cromosoma.

En el procedimiento de FISH, primero se construye una sonda. La sonda tiene que ser de una longitud suficiente para hibridarse específicamente a su diana (y no a secuencias similares en el genoma), pero no demasiado grande para impedir el proceso de hibridación, y se debería de etiquetar directamente con fluoróforos, con dianas para anticuerpos o con biotina. Esto se puede hacer de diversas maneras, por ejemplo mediante traducción por muescas y PCR usando nucleótidos etiquetados. Después, se produce una preparación cromosómica. Los cromosomas se unen firmemente a un sustrato, habitualmente vidrio. Tras la preparación, la sonda se aplica al ADN cromosómico y comienza a hibridarse. En varias etapas de lavado, todas las sondas no hibridadas o parcialmente hibridadas se eliminan por lavado. Si es necesaria la amplificación de la señal para superar el umbral de detección del microscopio (que depende de muchos factores tales como la eficiencia del marcaje de la sonda, del tipo de sonda y del colorante fluorescente), los anticuerpos etiquetados fluorescentes o la estreptavidina se unen a las moléculas de la etiqueta, amplificando así la fluorescencia. Finalmente, la muestra se embebe en un agente contra la lixiviación y se observa en un microscopio de fluorescencia.

En FISH de fibras, los cromosomas de la interfase se unen a un portaobjetos de tal manera que se estiran en una línea recta, en lugar de enrollarse fuertemente, como en FISH convencional, o en lugar de adoptar una conformación aleatoria, como en FISH de interfase. Esto se logra aplicando cizallamiento mecánico a lo largo de la longitud del portaobjetos; bien a las células que se han fijado al portaobjetos y después se han lisado, o bien a una disolución de ADN purificado. La conformación extendida de los cromosomas permite una resolución drásticamente mayor – incluso hasta unas pocas kilobases. Sin embargo, la preparación de las muestras de FISH de fibras, aunque conceptualmente simple, es una técnica más bien experta, queriendo decir que solamente los laboratorios especializados son capaces de usarla de forma habitual.

Anteriormente se describe un protocolo del método de Fish sobre fibras:

Equipo y reactivos:

- 5
 - cultivo de células linfoblastoides
 - PBS
 - Hemocitómetro
- 10
 - Disolución de lisis
 - 5 partes de NaOH 70 mM, 2 partes de etanol absoluto (Fidlerova et al. 1994). Esta disolución se puede almacenar a RT durante varios meses.

Método

- 15
 - Tomar 1-2 ml de suspensión celular procedente de un cultivo sano.
- 20
 - Lavar dos veces en 5 ml de PBS.
 - Resuspender en 1 ml de PBS.
 - Recontar una alícuota de células usando el hemocitómetro.
- 25
 - Diluir las células con PBS adicional para dar una concentración final de aproximadamente 2×10^6 /ml.
 - Extender 10 μ l de suspensión celular sobre un área de 1 cm en la parte superior de un portaobjetos de microscopio limpio.
- 30
 - Secar al aire.
 - Ajustar un portaobjetos en una cámara de plástico Cadenza (Shandon Southern) y fíjese en una posición casi vertical.
- 35
 - Aplicar 150 μ l de disolución de lisis en la parte superior del Cadenza.
 - A medida que el nivel cae por debajo del borde congelado del portaobjetos, añadir 200 μ l de etanol.
- 40
 - Permitir drenar brevemente.
 - Agarrar los bordes, levántese con cuidado el portaobjetos y la unidad de Cadenza fuera de la pinza.
 - Tirar de la parte superior del portaobjetos desde el Cadenza, permitiendo que el menisco se mueva hacia abajo del portaobjetos.
- 45
 - Secar al aire en un ángulo.
 - Fijar en acetona durante 10 minutos. Los portaobjetos se pueden almacenar satisfactoriamente a temperatura ambiente durante varios meses.
- 50

En una forma de realización, todas las sondas están marcadas con el mismo marcador.

- 55

En una forma de realización, las sondas están marcadas con por lo menos dos marcadores diferentes. En una forma de realización, las sondas están marcadas con tres marcadores.

- 60

En una forma de realización, las sondas unidas o hibridadas se detectan detectando uno o más marcadores unidos a las sondas. Los marcadores se pueden incorporar por cualquiera de un número de medios bien conocidos por los expertos en la materia. Sin embargo, en una forma de realización preferida, el marcador se incorpora simultáneamente durante la etapa de amplificación en la preparación de las sondas. Por ejemplo, la reacción en cadena de la polimerasa (PCR) con cebadores marcados o con oligonucleótidos marcados proporcionará un producto de amplificación marcado. La sonda (por ejemplo, ADN) se amplifica en presencia de trifosfatos de desoxinucleótidos marcados (dNTPs).

En una forma de realización preferida, la amplificación de la transcripción, como se describe anteriormente, usando un nucleótido marcado (por ejemplo, UTP y/o CTP marcado con fluoresceína) incorpora un marcador en los ácidos nucleicos transcritos.

Alternativamente, se puede añadir directamente un marcador a la sonda original (por ejemplo, ARNm, ARNm de poliA, ADNc, etc.) o al producto de amplificación después de que la amplificación esté terminada. Tal marcaje puede dar como resultado un mayor rendimiento de los productos de amplificación y reducir el tiempo requerido para la reacción de amplificación. Los medios para unir marcadores a las sondas incluyen, por ejemplo, traducción por muescas o marcaje de los extremos (por ejemplo con un ARN marcado) mediante el tratamiento con cinasas del ácido nucleico y la unión (ligación) subsiguiente de un ligador de ácido nucleico que une la sonda a un marcador (por ejemplo, un fluoróforo).

Preferentemente, el nucleótido marcado según la presente invención es clorodesoxiuridina (CldU), bromodesoxiuridina (BrdU) y o yododesoxiuridina (IdU).

Los marcadores detectables adecuados para uso en la presente invención incluyen cualquier composición detectable por medios espectroscópicos, fotoquímicos, bioquímicos, inmunoquímicos, eléctricos, ópticos o químicos. Los marcadores útiles en la presente invención incluyen biotina para la tinción con un conjugado de estreptavidina marcado, perlas magnéticas (por ejemplo, Dynabeads.TM.), colorantes fluorescentes (por ejemplo, fluoresceína, Rojo Tejas, rodamina, proteína fluorescente verde, y similares; véase, por ejemplo, Molecular Probes, Eugene, Oreg., USA), radiomarcadores (por ejemplo, ^3H , ^{125}I , ^{35}S , ^{14}C , o ^{32}P), enzimas (por ejemplo, peroxidasa de rábano picante, fosfatasa alcalina y otras usadas habitualmente en un ELISA), y marcadores colorimétricos tales como oro coloidal (por ejemplo, las partículas de oro en el intervalo de tamaños de 40-80 nm de diámetro dispersan la luz verde con una eficiencia elevada) o perlas de vidrio o de plástico (por ejemplo, poliestireno, polipropileno, látex, etc.) coloreadas. Las patentes que enseñan el uso de tales marcadores incluyen las patentes U.S. n^{os} 3.817.837; 3.850.752; 3.939.350; 3.996.345; 4.277.437; 4.275.149; y 4.366.241.

Se prefiere un marcador fluorescente debido a que proporciona una señal muy fuerte con un fondo muy bajo. También es ópticamente detectable a una resolución y sensibilidad elevadas a través de un procedimiento de barrido rápido.

La sonda se puede marcar toda ella con un único marcador, por ejemplo un marcador fluorescente único. Alternativamente, en otra forma de realización, se pueden hibridar simultáneamente diferentes sondas, en el que cada sonda tiene un marcador diferente. Por ejemplo, una diana podría tener un marcador fluorescente verde, y una segunda diana podría tener un marcador fluorescente rojo. La etapa de barrido distinguirá sitios de unión del marcador rojo de aquellos que se unen al marcador fluorescente verde. Cada sonda (ácido nucleico diana) se puede analizar independientemente entre sí.

Los cromógenos adecuados que se pueden emplear incluyen aquellas moléculas y compuestos que absorben luz en un intervalo característico de longitudes de onda de manera que se puede observar un color, o, como alternativa, que emiten luz cuando se irradian con radiación de una longitud de onda particular o intervalo de longitudes de onda, por ejemplo fluorescentes.

Existe una amplia variedad de colorantes adecuados, seleccionándose principalmente para proporcionar un color intenso con absorción mínima por sus entornos. Los tipos de colorante ilustrativos incluyen colorantes de quinolina, colorantes de triarilmetano, colorantes de acridina, colorantes de alizarina, ftaleínas, colorantes de insectos, azocolorantes, colorantes antraquinoides, colorantes de cianina, colorantes de fenazationio, y colorantes de fenazoxonio.

Se puede utilizar una amplia variedad de fluorescentes, ya sea solos o, como alternativa, junto con moléculas apagadoras. Los fluorescentes de interés caen en una variedad de categorías que tienen ciertas funcionalidades principales. Estas funcionalidades principales incluyen 1- y 2-aminonaftaleno, p,p'-diaminoestilbenos, pirenos, sales de fenantridina cuaternaria, 9-aminoacridinas, iminas de p,p'-diaminobenzofenona, antracenos, oxacarbocianina, marocianina, 3-aminoequilenina, perileno, bisbenzoxazol, bis-p-oxazolil benceno, 1,2-benzofenazina, retinol, sales de bis-3-aminopiridinio, helebrigenina, tetraciclina, esterofenol, bencimidazolilfenilamina, 2-oxo-3-cromeno, indol, xanteno, 7-hidroxycumarina, fenoxazina, salicilato, estrofantidina, porfirinas, triarilmetanos y flavina.

Los compuestos fluorescentes individuales que tienen funcionalidades para el enlazamiento o que se pueden modificar para incorporar tales funcionalidades incluyen, por ejemplo, cloruro de dansilo, fluoresceínas tales como 3,6-dihidroxi-9-fenilxantidrol; isotiocianato de rodamina; N-fenil-1-amino-8-sulfonatonaftaleno; N-fenil-2-amino-6-sulfonatonaftaleno; ácido 4-acetamido-4-isotiocianato-estilben-2,2'-disulfónico; ácido piren-3-sulfónico; 2-toluidinonaftalen-6-sulfonato; N-fenil,N-metil 2-aminonaftalen-6-sulfonato; bromuro de etidio; estebrina; auromina-0,2-(9'-antroi)palmitato; dansil fosfatidiletanolamina; N,N'-dioctadecil oxacarbocianina; N,N'-dihexil oxacarbocianina; merocianina, 4(3'pirenil)butirato; d-3-aminodesoxi-equilenina; 12-(9'-antroi)estearato; 2-metilantraceno; 9-vinilantraceno; 2,2'-(vinilen-p-fenil)bisbenzoxazol; p-bis[2-(4-metil-5-fenil-oxazolil)]benceno; 6-dimetilamino-1,2-benzofenazina; retinol; diyoduro de bis(3'-aminopiridinio) 1,10-decandiilo; sulfonaftilhidrazona de helibrienina;

clorotetraciclina; N(7-dimetilamino-4-metil-2-oxo-3-cromenil)maleimida; N-[p-(2-bencimidazolil)-fenil]maleimida; N-(4-fluorantil)maleimida; bis(ácido homovanílico); resazarina; 4-cloro-7-nitro-2,1,3-benzooxadiazol; merocianina 540; resorrufina; rosa de bengala; y 2,4-difenil-3(2H)-furanona.

5 En particular, los marcadores fluorescentes según la presente invención son 1-cloro-9,10-bis(feniletinil)antraceno, 5,12-bis(feniletinil)naftaceno, 9,10-bis(feniletinil)antraceno, naranja de acridina, auramina O, benzantrona, cumarina, 4',6-diamidino-2-fenilindol (DAPI), bromuro de etidio, fluoresceína, fluorescente proteína verde, tinción de Hoechst, amarillo indio, luciferina, ficobilina, ficoeritrina, rodamina, rubreno, estilbeno, TSQ, Rojo Tejas, y umbeliferona.

10 De forma deseable, los fluorescentes deberían absorber luz por encima de aproximadamente 300 nm, preferentemente de aproximadamente 350 nm, y más preferentemente por encima de aproximadamente 400 nm, emitiendo habitualmente a longitudes de onda mayores que aproximadamente 10 nm mayores que la longitud de onda de la luz absorbida. Se debería observar que las características de absorción y emisión del colorante unido pueden diferir del colorante no unido. Por lo tanto, cuando se hace referencia a los diversos intervalos y
15 características de las longitudes de onda de los colorantes, se pretende indicar los colorantes como se utilizan, y no el colorante que no está conjugado y que está caracterizado en un disolvente arbitrario.

Los fluorescentes son preferidos generalmente debido a que al irradiar un fluorescente con luz se puede obtener una pluralidad de emisiones. De este modo, un único marcador puede proporcionar una pluralidad de sucesos medibles.

20 Según la presente invención, cuando el marcaje se realiza con un marcador fluorescente, la lectura de las señales se realiza mediante detección fluorescente: la sonda marcada fluorescentemente es excitada por la luz y entonces la emisión de la excitación se detecta mediante un fotosensor tal como una cámara CCD equipada con filtros de emisión apropiados que captura una imagen digital y permite el análisis posterior de los datos.

25 La señal detectable también se puede proporcionar mediante fuentes quimioluminiscentes y bioluminiscentes. Las fuentes quimioluminiscentes incluyen un compuesto que se excita electrónicamente mediante una reacción química y entonces puede emitir luz que sirve como la señal detectable, o dona energía a un aceptor fluorescente. Se ha encontrado un número diverso de familias de compuestos que proporcionan quimioluminiscencia bajo una variedad de condiciones. Una familia de compuestos es 2,3-dihidro-1,4-ftalazindiona. El compuesto más popular es luminol, que es el compuesto 5-amino. Otros miembros de la familia incluyen el análogo 5-amino-6,7,8-trimetoxi- y el dimetilamino[ca]benzo. Se puede hacer que estos compuestos produzcan luminiscencia con peróxido de hidrógeno alcalino o hipoclorito de calcio y una base. Otra familia de compuestos son los 2,4,5-trifenilimidazoles, con lofina como el nombre común para el producto progenitor. Los análogos quimioluminiscentes incluyen sustituyentes para-
30 dimetilamino y -metoxi. La quimioluminiscencia también se puede obtener con oxalatos, habitualmente ésteres activos de oxalilo, por ejemplo p-nitrofenilo, y un peróxido, por ejemplo peróxido de hidrógeno, en condiciones básicas. Alternativamente, las luciferinas se pueden usar junto con luciferasa o lucigeninas para proporcionar bioluminiscencia.

40 Los marcadores de espín se proporcionan por moléculas informadoras con un espín electrónico desapareado que se puede detectar mediante espectroscopía de resonancia de espín electrónico (ESR). Los marcadores de espín ejemplificativos incluyen radicales libres orgánicos, complejos de metales de transición, particularmente vanadio, cobre, hierro, y manganeso, y similares. Los marcadores de espín ejemplificativos incluyen radicales libres de nitróxido.

45 El marcador se puede añadir a la sonda (o diana, que es en particular ácido o ácidos nucleicos) antes o después de la hibridación. Los denominados "marcadores directos" son marcadores detectables que están unidos directamente o se incorporan a la sonda antes de la hibridación. Por el contrario, los denominados "marcadores indirectos" se unen al dúplex híbrido tras la hibridación. A menudo, el marcador indirecto se une a un resto de unión que se ha unido a la sonda antes de la hibridación. De este modo, por ejemplo, la sonda puede estar biotinilada antes de la hibridación. Tras la hibridación, un fluoróforo conjugado a avidina se unirá a los dúplex híbridos que portan la biotina, proporcionando un marcador que es fácilmente detectado. Para una revisión detallada de los métodos de marcaje de ácidos nucleicos y de detección de ácidos nucleicos hibridados marcados, véase Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 24: Hybridization With Nucleic Acid Probes, P. Tijssen, ed. Elsevier, N.Y.,
50 (1993)).

Los marcadores se pueden unir directamente o a través de un resto enlazador. En general, el sitio de la unión del marcador o del ligador-marcador no está limitado a ninguna posición específica. Por ejemplo, un marcador se puede unir a un nucleósido, nucleótido, o análogo del mismo en cualquier posición que no interfiera con la detección o
60 hibridación según se desee. Por ejemplo, ciertos reactivos Label-ON de Clontech (Palo Alto, Calif.) proporcionan el marcaje intercalado a lo largo de la cadena principal de fosfato de un oligonucleótido, y el marcaje terminal en los extremos 3' y 5'. Como se muestra para el ejemplo en la presente memoria, los marcadores se pueden unir en las posiciones en el anillo de ribosa, o la ribosa se puede modificar o incluso eliminar según se desee. Los restos básicos de reactivos de marcaje útiles pueden incluir aquellos que son de origen natural o aquellos que se modifican de manera que no interfieran con el fin para el que se proponen. Las bases modificadas incluyen, pero no se limitan a, 7-desaza A y G, 7-desaza-8-aza A y G, y otros restos heterocíclicos.

Con respecto a sondas de marcaje de los extremos, en muchas aplicaciones es útil marcar directamente sondas sin tener que pasar a través de la amplificación, transcripción u otra etapa de conversión. En general, los métodos de marcaje de los extremos permiten la optimización del tamaño del ácido nucleico a marcar. Los métodos del marcaje de los extremos también disminuye la tendencia de la secuencia asociada algunas veces con métodos de marcaje ayudados por polimerasas. El marcaje de los extremos se puede llevar a cabo usando transferasa terminal (TdT).

El marcaje de los extremos también se puede lograr ligando un oligonucleótido marcado o análogo del mismo al extremo de una sonda. Otros métodos de marcaje de los extremos incluyen la creación de una "cola" marcada o no marcada para el ácido nucleico usando ligasa o transferasa terminal, por ejemplo. El ácido nucleico con la cola se expone entonces a un resto marcado que se asociará preferentemente a la cola. La cola y el resto que se asocia preferentemente a la cola pueden ser un polímero tal como un ácido nucleico, péptido, o hidrato de carbono. La cola y su resto de reconocimiento pueden ser aquello que permita el reconocimiento entre los dos, e incluyen moléculas que tienen relaciones de ligando-sustrato tales como haptenos, epítomos, anticuerpos, enzimas y sus sustratos, y ácidos nucleicos complementarios y sus análogos.

Los marcadores asociados con la cola o el resto de reconocimiento de la cola incluyen restos detectables. Cuando la cola y su resto de reconocimiento están ambos marcados, los marcadores respectivos asociados con cada uno pueden tener ellos mismos una relación de ligando-sustrato. Los marcadores respectivos también pueden comprender reactivos de transferencia de energía, tales como colorantes que tienen diferentes características espectroscópicas. El par de transferencia de energía se puede escoger para obtener las características espectrales combinadas deseadas. Por ejemplo, un primer colorante que absorbe a una longitud de onda más corta que la absorbida por el segundo colorante puede, con la absorción a esa longitud de onda más corta, transferir energía al segundo colorante. El segundo colorante emite entonces radiación electromagnética a una longitud de onda más larga que la que se habría emitido por el primer colorante solo. Los reactivos de transferencia de energía pueden ser particularmente útiles en esquema de marcaje de dos colores, tales como los expuestos en la solicitud de patente US en trámite junto con la presente, presentada el 23 de diciembre de 1996, y que es una continuación en parte de la Serie US nº 08/529.115, presentada el 15 de septiembre de 1995, y la Solicitud Internacional nº WO 96/14839, presentada el 13 de septiembre de 1996, que es también una continuación en parte de la Serie US nº 08/670.118, presentada el 25 de junio de 1996, que es una división de la Serie US nº 08/168.904, presentada el 15 de diciembre de 1993, que es una continuación de la Serie US nº 07/624.114, presentada el 6 de diciembre de 1990. La Serie US nº 07/624.114 es una CIP de la Serie US nº 07/362.901, presentada el 7 de junio de 1990.

En una forma de realización, cuando el marcaje se realiza con un marcador radioactivo, la lectura de las señales se realiza mediante detección radioactiva.

La detección radioactiva se puede realizar con película de rayos X o con un formador de fosfoimágenes.

Los ejemplos de marcadores radioactivos según la presente invención son ^3H , ^{125}I , ^{35}S , ^{14}C , o ^{32}P .

En una forma de realización preferida, las sondas están marcadas con marcadores fluorescentes.

En una forma de realización preferida, las sondas están marcadas con un marcador radioactivo.

Según la presente invención, en el caso de que las sondas estén marcadas con por lo menos dos marcadores diferentes, la firma de un dominio de interés resulta de la sucesión de marcadores.

En una forma de realización, la unión de por lo menos tres sondas que corresponden a un dominio de interés en la macromolécula forma una secuencia de por lo menos dos espacios escogidos entre un grupo de por lo menos dos espacios diferentes (por ejemplo, "corto" y "largo"), siendo dicho grupo idéntico para cada dominio de interés.

En una forma de realización, el conjunto de sondas comprende además dos sondas (sonda 1 o sonda 2), siendo cada sonda capaz de unirse en un extremo diferente del dominio de interés, estando asociada la lectura de la señal de una de dicha sonda 1 o sonda 2 con su sonda consecutiva en el dominio de interés, denominado "pareja de sondas de extremo de comienzo o final" que permite obtener una información de comienzo o final de la lectura.

En una forma de realización, la información del comienzo de la lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas de la pareja de sondas de extremo de comienzo.

En una forma de realización, la información del extremo de la lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas de la pareja de sondas de extremo de final.

En forma de realización, la información del comienzo de la lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas de la pareja de sondas de extremo de comienzo y la información de final de la lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas de la pareja de sondas de extremo de final, siendo dicho espaciamiento diferente para la pareja de sondas de extremo de comienzo y la pareja de sondas

de extremo de final, a fin de diferenciar la información del comienzo y del final.

Otro aspecto de la invención se refiere a un método para la determinación de la presencia de por lo menos un dominio de interés en una macromolécula, diseñándose un conjunto de por lo menos tres sondas capaces de unirse específicamente a dicho dominio, en el que la unión de un conjunto de por lo menos tres sondas en la macromolécula forma una secuencia de por lo menos dos códigos seleccionados entre un grupo de por lo menos dos códigos diferentes (por ejemplo "puntos" y "guiones"), siendo dicho grupo idéntico para cada dominio, y siendo dicha secuencia de códigos específica del dominio y siendo la firma del dominio, y comprendiendo dicho método las siguientes etapas:

- a) poner en contacto dichos por lo menos dos conjuntos de por lo menos tres sondas con la macromolécula, siendo cada conjunto específico de cada dominio de interés;
- b) leer las señales, correspondiendo una señal a una sonda;
- c) transcribir dichas señales en códigos;
- d) detectar la presencia o no de una secuencia de códigos específica del dominio de interés, correspondiendo la presencia de dicha secuencia de códigos a la presencia e indicando la posición del dominio en la macromolécula;

y opcionalmente una etapa de linealización de la macromolécula antes de la etapa a) o entre la etapa a) y b).

En una forma de realización preferida, un método es para la determinación de la presencia de por lo menos dos dominios de interés en una macromolécula.

En una forma de realización preferida, el código obtenido resulta de la lectura de la característica de dos señales consecutivas.

En una forma de realización, la característica de dos señales consecutivas se da mediante el marcaje de las sondas.

En una forma de realización, el marcaje se realiza con un marcador fluorescente, correspondiendo un marcador fluorescente específico a un código específico.

En una forma de realización, el marcaje se realiza con marcador radioactivo, correspondiendo un marcador radioactivo específico a un código específico.

En una forma de realización, el código obtenido resulta de la lectura del espaciamiento entre dos señales consecutivas.

En una forma de realización, un espaciamiento particular entre dos señales corresponde a un código particular.

Por ejemplo, según la invención, los espacios pequeños corresponden a un código denominado puntos, y los espacios grandes a un código denominado guiones. Se pueden diseñar códigos pequeños para medir, por ejemplo, 25-30 kb, y se pueden diseñar códigos grandes para medir, por ejemplo, 55-70 kb.

En una forma de realización, la característica de dos señales consecutivas se da mediante la longitud de las sondas, correspondiendo cada longitud a un código específico.

Es obvio que la invención no se reduce a dos códigos, sino que puede comprender tres, cuatro o más de cuatro códigos, escogiéndose cualquier tamaño de espacio para definir un código, y siendo arbitrarios dichos tamaños.

Según la presente invención, el conjunto de sondas comprende además dos sondas (sonda 1 o sonda 2), cada sonda es capaz de unirse a un extremo diferente del dominio de interés, la lectura de la señal de una de dicha sonda 1 o sonda 2 está asociada con su sonda consecutiva en el dominio de interés, denominada "pareja de sondas de extremo de comienzo o final" que permite obtener una información del comienzo o final de la lectura.

En una forma de realización, la firma según la presente invención comprende un espacio entre la primera y la segunda sonda en un conjunto de sondas, siendo el espacio diferente de todos los otros espacios en la firma, y el espacio se puede usar para obtener información sobre el comienzo de la firma.

En una forma de realización, la firma según la presente invención comprende un espacio entre la siguiente a la última y la última sonda en un conjunto de sondas, siendo el espacio diferente de todos los otros espacios en la firma, y el espacio se puede usar para obtener información sobre el final de la firma.

En una forma de realización adicional, la invención describe un kit para la detección de la presencia de por lo menos

dos dominios de interés en una macromolécula, que comprende por lo menos dos conjuntos de tres sondas, siendo capaz cada conjunto de sondas de unirse específicamente a un dominio, en el que la unión de un conjunto de por lo menos tres sondas en la macromolécula forma una secuencia de por lo menos dos códigos escogidos entre un grupo de por lo menos dos códigos diferentes (por ejemplo "puntos" y "guiones"), siendo dicho grupo idéntico para todos los dominios, siendo dicha secuencia de códigos específica de un dominio y siendo la firma de dicho dominio.

La presente invención se pondrá más claramente de manifiesto a partir de la descripción de los estudios experimentales realizados en el contexto de la investigación llevada a cabo por el solicitante, que no se debe de interpretar de manera limitativa.

Ejemplo

Los orígenes de la replicación del ADN de mamífero se localizan en sitios que oscilan desde pb a decenas de kb. Una distribución normal de las iniciaciones en ciclos celulares individuales sugiere que solamente un número limitado de estos numerosos sitios de comienzo potenciales se convierten en orígenes activados. La interferencia del origen puede silenciar orígenes redundantes. Sin embargo, se desconoce actualmente si la interferencia participa en el espaciamiento de sucesos de iniciación humanos funcionales. Al usar una nueva estrategia de hibridación, el Código Morse Genómico, en moléculas de ADN peinado individuales procedentes de queratinocitos primarios, los inventores dan a conocer los sitios de iniciación presentes en 1,5 Mb del cromosoma humano 14q11.2. Los inventores confirman que las zonas de iniciación están ampliamente extendidas en las células humanas. Los orígenes usados en ciclos celulares individuales son menos abundantes que los sitios potenciales de iniciación, y su uso limitado produce distancias de accionamiento entre orígenes normales.

La interferencia entre zonas disminuye en proporción a la distancia del origen activo, mientras que la interferencia en las zonas es 100% eficiente. Estos resultados identifican una organización jerárquica de actividad del origen en células humanas. Los orígenes funcionales gobiernan la probabilidad de que los orígenes cercanos se activarán, que está mediado por la interferencia entre zonas y en las zonas. Esto asegura una distribución uniforme de los sucesos de iniciación a lo largo de la longitud de los cromosomas humanos en el contexto de sitios de comienzo potenciales múltiples y complejos de la replicación del ADN.

Introducción

Las células eucariotas tienen una cantidad limitada de tiempo, definido por la duración de la fase S, para duplicar sus genomas. Esto se logra sintetizando ADN en horquillas de replicación, que se extienden a partir de múltiples sitios de iniciación. Puesto que la velocidad de la horquilla no aumenta en escala según la duración de la fase S, es necesario regular la frecuencia de iniciación a lo largo de cada cromosoma respectivo para evitar espacios sin replicar antes del comienzo de la mitosis (Hand y Tamm, 1973; Edenberg y Huberman, 1975). Aunque hay excepciones, la visión normal es que los orígenes mamíferos somáticos se encienden a intervalos de 50-300 kb (Edenberg y Huberman, 1975; Berezney et al., 2000). Esto sugiere que los metazoos poseen un mecanismo para distribuir uniformemente sucesos de iniciación. La colocación de secuencias replicadoras fuertes en distancias regulares es uno de tales mecanismos que se emplean por la levadura emergente, *Saccharomyces cerevisiae*. (Newlon et al., 1991; Shirahige et al., 1993). En eucariotas superiores, los elementos genéticos desempeñan un papel en la activación del origen; sin embargo, no son suficientes por sí mismos para conducir la iniciación (Gilbert, 2004). Además, mientras que algunos orígenes de metazoos se localizan en sitios bien circunscritos de unos pocos pb, un gran número se localiza en zonas de iniciación más dispersas que oscilan hasta decenas de kb (DePamphilis, 1999). Esto crea el problema de cómo lograr una distribución normal de orígenes activados a partir de un intervalo de sitios potenciales que poseen una eficiencia intrínseca baja.

Un método para regular la actividad del origen es cambiar la probabilidad con la que se replicará pasivamente. Puesto que una horquilla de elongación de un origen vecino media esta supresión, esta forma de desactivación del origen se ha denominado "interferencia del origen" (Brewer y Fangman, 1993). La mayoría de nuestra comprensión con respecto a la interferencia del origen ha sido proporcionada por el trabajo en *S. cerevisiae*. En levadura emergente, hay muchos más precomplejos replicativos (pre-RCs) ensamblados que aquellos que se necesitan o que se usan para completar la replicación (Dershowitz y Newlon, 1993; Raghuraman et al., 2001; Wyrick et al., 2001; Pasero et al., 2002). El análisis de la eficiencia del origen en los cromosomas III y VI de levadura reveló que los orígenes se usan entre 5-90% de los ciclos celulares (Friedman et al., 1997; Yamashita et al., 1997; Poloumienko et al., 2001). Los orígenes autorizados son ineficientes debido a su tiempo programado tarde en la fase S o relativamente tarde en comparación con otros orígenes en la vecindad (Santocanale y Diffley, 1996; Vujcic et al., 1999). Como consecuencia, estos orígenes competentes se replican de forma pasiva por las horquillas que se alargan desde los sitios de iniciación de flanco (Santocanale et al., 1999).

Según datos de levadura, si los orígenes interfieren entre sí, los orígenes vecinos se deben de preprogramar en G1 para comenzar en tiempos diferentes durante la fase S (Raghuraman et al., 1997). Sin embargo, este requisito puede no ser satisfecho en eucariotas superiores, en las que 1) el control del tiempo se ejerce a lo largo de regiones extendidas de ~100 kb (MacAlpine et al., 2004; Norioet al., 2005), y 2) los orígenes situados próximos entre sí comienzan simultáneamente en agrupamientos (Berezney et al., 2000).

De hecho, en el locus AMPD2 amplificado de células CHO, no se observó diferencias de tiempo predefinidas significativas entre orígenes cercanos (Anglana et al., 2003). No obstante, la coactivación de orígenes adyacentes en localizaciones de pb bien definidas se bloqueó (Anglana et al., 2003). Está por determinarse si esto se aplica a zonas de iniciación amplias.

Para entender cómo se logran intervalos de iniciación regulares en células humanas, y si la interferencia del origen contribuye a este proceso, los inventores consultaron una región de 1,5 Mb del cromosoma humano 14q11.2 de queratinocitos primarios en busca de la actividad del origen. Se escogió un enfoque de una sola molécula, que explota la tecnología del peinado molecular, por las siguientes razones. En primer lugar, se pueden obtener suficientes sucesos de disparo del origen para posicionar todos los sitios de comienzo potenciales de la replicación del ADN en un tipo celular particular. En segundo lugar, los inventores pudieron determinar qué orígenes usan las células individuales en las fases S individuales y su tiempo de activación unas con respecto a las otras. Esto es necesario para averiguar la distribución espacio-temporal de los sucesos de iniciación. Estos datos se combinaron para evaluar si los orígenes que se habían ya accionado regulan el uso de los sitios de iniciación potenciales aguas abajo. Los inventores encontraron que los orígenes se autorregulan entre sí según una jerarquía establecida por el origen activo, que se selecciona estocásticamente sin preferencias de tiempo predefinidas. Además, la interferencia del origen produce un espaciamiento de sucesos de iniciación conservado. Se discuten las razones y los mecanismos usados para implementar la interferencia del origen humano.

Materiales y métodos

Preparación del ADN

Los queratinocitos primarios humanos normales derivaron de biopsias de piel. Las células se cultivaron en SFM de queratinocitos estándar. Después, el ADN naciente se marcó con IdU y CldU durante 20 min. cada uno como se describió (Lebofsky y Bensimon, 2005). La extracción y peinado del ADN se realizaron según protocolos estándar.

Hibridación y revelación fluorescente

Las 25 sondas de GMC se produjeron mediante PCR de largo alcance usando los pares de cebadores enumerados en la Tabla S1. Para ayudar a amplificar sondas de 5-7 kb, se usó TaKaRa LA Taq (Takara Bio). Los BACs que sirvieron como moldes para los productos de la PCR se encuentran en la Tabla S2. Las sondas se reunieron a una concentración final de ~20 ng/μl según su símbolo (A, B, C, y D). La biotilación de las sondas se logró cebando al azar (Invitrogen) los cuatro símbolos separadamente. Para los ensayos de portaobjetos individuales, se combinaron ~250 ng de cada sonda biotinilada con 10 μg de ADN cot-1 humano (Gibco BRL). Después de eso, la hibridación en el ADN peinado se ajustó a los métodos publicados (Lebofsky y Bensimon, 2005). Las etapas inmunofluorescentes para detectar las sondas, IdU y CldU, fueron como siguen: (i) estreptavidina conjugada con Alexa 488 (Molecular Probes), anti-bromodesoxiuridina de ratón (Becton Dickinson), y anti-bromodesoxiuridina de rata (Harlan Seralab); (ii) antiestreptavidina de conejo conjugado con biotina (Rockland), anti-ratón de cabra conjugado con Alexa 350 (Molecular Probes), y anti-rata de burro conjugado con Rojo Texas (Jackson); (iii) estreptavidina conjugada con Alexa 488 (Molecular Probes) y anti-cabra de burro conjugado con Alexa 350 (Molecular Probes). Las incubaciones de los anticuerpos, los lavados, y el montaje de los portaobjetos se realizaron como se dio a conocer previamente (Lebofsky y Bensimon, 2005).

Adquisición de imágenes

La mitad de las imágenes se capturaron con un microscopio Zeiss Axioplan 2 equipado con una cámara CCD HQ (Photometrics). La otra mitad se adquirió usando el dispositivo de barrido de alto rendimiento Cytoscout (Upper Austrian Research). Los puntos fluorescentes del fondo se eliminaron usando Photoshop (Adobe), para destacar la molécula de interés.

Resultados

Estrategia de hibridación del Código Morse Genómico

Como una primera etapa para analizar qué orígenes son activos y silenciosos en un ciclo celular dado, se obtuvieron todos los sitios de iniciación potenciales en una región. Para este fin, los inventores usaron un enfoque de una sola molécula basado en el peinado molecular (Bensimon et al., 1994). A queratinocitos primarios humanos asíncronos que presentan un cariotipo normal se les proporcionaron pulsos de 20 min. de yododesoxiuridina (IdU) seguido de clorodesoxiuridina (CldU). El ADN procedente de esta población celular se extrajo y se peinó. La incorporación de los análogos de BrdU a lo largo del ADN recientemente sintetizado se visualizó mediante detección inmunológica de IdU en azul y CldU en rojo. Este paradigma experimental da lugar a tres tipos de señales a partir de las cuales se puede inferir el sitio de comienzo de la replicación bidireccional (Figura 1A; para una descripción detallada, véanse Anglana et al., 2003; Lebofsky y Bensimon, 2005). Combinando la revelación mediante BrdU con FISH, la iniciación se puede atribuir a tramos de secuencias específicos, independientemente de qué tramos de la replicación solapen

con las señales de las sondas.

Convencionalmente, para visualizar una región genómica en el portaobjetos, se usan pares de sondas de diferentes colores (Figura 1Bi) o de diferentes tamaños (Figura 1Bii). Sin embargo, la ruptura del ADN limita la distancia que pueden cubrir. Como resultado, el paseo sobre un cromosoma usando esta estrategia de hibridación es un procedimiento que consume tiempo puesto que los experimentos incrementan la proporcionalidad con el número de pares de sondas. Idealmente, se podrían usar varios colores y/o sondas de longitudes variables para cubrir una región grande; sin embargo, estas estrategias no son adecuadas debido a solapamiento espectral y a hibridación no específica de secuencias repetitivas, respectivamente.

La primera parte de nuestra solución única llegó al darnos cuenta de que los espacios de diferente tamaño proporcionan la misma información que las sondas de diferente color o tamaño. En el ejemplo proporcionado en la Figura 1Biii, el espacio 1 se define por un conjunto de sondas, y el espacio 2 se define por otro conjunto de sondas. También, el tamaño del espacio entre los dos conjuntos de sondas es distinto de los espacios 1 y 2. El espacio 3 es útil durante la ruptura del ADN. Con esta ayuda, la molécula todavía se puede orientar incluso aunque no se visualice el conjunto completo de sondas (Figuras 1Biv y v). Puesto que los espacios proporcionan información posicional, sus números ya no están limitados, es decir, el solapamiento espectral y las secuencias repetitivas durante la hibridación ya no son un problema. Mediante el uso de espacios de diferentes tamaños, se generó un Código Morse Genómico (GMC) que cubre ~1,5 Mb en el cromosoma humano 14q11.2 (Figura 1C). Todo el GMC se hibridó en ensayos individuales, y todas las sondas se detectaron en verde. Antes del peinado molecular, la manipulación del ADN provoca que las fibras se rompan en localizaciones aleatorias. Sin embargo, el tamaño de las fibras fue suficiente para permitir la visualización de múltiples símbolos en moléculas individuales. Por el contrario, debido a la ruptura de las fibras, ocasionalmente solo fueron detectables unas pocas sondas a partir de un símbolo. Los orígenes se cartografiaron siempre que las pistas de replicación que denotan iniciación se colocalizaron con un conjunto descodificable de sondas de GMC (Figura 1D). De este modo, la nueva estrategia de hibridación, GMC, permitió el cartografiado de los orígenes a lo largo de una región grande en un número limitado de experimentos.

Cartografiado de iniciación en 1,5 Mb del cromosoma humano 14q11.2

Usando este paradigma experimental, se detectaron 307 sucesos de iniciación en 232 moléculas de ADN individuales en la región del GMC. Se llevó a cabo el agrupamiento de datos para establecer de forma objetiva zonas de iniciación preferente. En primer lugar, se creó un árbol de agrupamiento jerárquico. Para lograr la mejor partición, la extensión de los datos en los agrupamientos se debería de minimizar, y se debería de maximizar la separación entre agrupamientos. Estos dos rasgos se denominan varianza dentro (W) y entre varianza (B) (Figura 2A). Por tanto, el conjunto de agrupamientos deseado debe tener una W pequeña y B grande, o valores máximos de $B - W$. Cuando el dato se dividió en 9, 22 y 45 agrupamientos, se obtuvieron valores de $B - W$ relativamente altos (Figura 2A). La Figura 2B muestra cómo los sucesos de iniciación se reparten según estos conjuntos de agrupamientos. La división del conjunto de datos en 45 agrupamientos produce las regiones más estrechas de iniciación. Los inventores consideraron que estos 45 agrupamientos representan zonas de iniciación individuales (Figura 2C). Algunos de estos agrupamientos contenían muy pocos sucesos de iniciación, lo que puede haber sido debido a ruido de fondo. Por lo tanto, para los análisis subsiguientes, se usaron solamente agrupamientos con más de tres sucesos de iniciación. De esta manera, se identificaron 38 zonas de iniciación.

Los tamaños de las zonas de iniciación variaron entre 2,6 kb (mín.) y 21,6 kb (máx.) con una media de 13,5 +/- 5,2 kb. Estos valores caen en el intervalo de otras zonas de iniciación dadas a conocer para células de mamífero (DePamphilis, 1999). En las zonas, no se observaron sitios de iniciación específicos de 1-2 kb.

Esto puede haber sido debido a números insuficientes de datos para distinguir los picos de los orígenes. La resolución máxima en moléculas peinadas está entre 1-4 kb. Por lo tanto, estas preferencias de iniciación también se pueden haber suavizado como resultado de la desviación estándar inherente a cada origen cartografiado.

Alternativamente, las zonas de iniciación que contienen sitios de disparo dispersos pueden ser más predominantes que las que se pensaron previamente. Los datos en el contexto de la presente invención proporcionan regiones diana que se pueden sonar con otras técnicas de mayor resolución (Todorovic et al., 2005). En consecuencia, estas posibilidades se pueden discriminar.

Análisis espacio-temporal de vecinos de orígenes activados

A continuación, los inventores volvieron la atención hacia cómo se distribuyeron las zonas de iniciación unas con respecto a otras. La medida de las distancias entre centroides de las zonas reveló una media entre zonas de 40,6 +/- 20,7 kb (mín. = 14,3 kb, máx. = 93,1 kb). Esto fue sorprendente considerando que las distancias entre orígenes en células de mamíferos generalmente oscilan entre 100-150 kb (Berezney et al., 2000). La discrepancia se puede explicar si solamente se activa un subconjunto de zonas por ciclo celular. Para explorar esta posibilidad, se analizó el espaciamiento entre iniciaciones múltiples en fibras individuales (Figura 3A).

Debido al nivel de moléculas individuales de nuestro análisis, estos orígenes corresponden a aquellos que son

usados realmente por una célula en una fase S. La ruptura del ADN evitó la visualización de orígenes de flanco para 173 de los 307 sucesos de iniciación observados. Los casos restantes se observaron en presencia de un vecino de origen activo (134/307). Los dos orígenes funcionales más cercanos y los dos orígenes funcionales más alejados estaban separados por 31,4 kb y 390,8 kb, respectivamente. De forma interesante, la distancia media entre orígenes se calculó como 113 +/- 66,4 kb (Figura 3B). En comparación con la distancia entre zonas (~40 kb), este resultado sugiere que, de media, solamente un origen se activa de tres zonas potenciales en un ciclo celular dado (Figura 3C).

Para investigar si los orígenes de zonas específicas se dispararon reproduciblemente más pronto o más tarde unos con respecto a otros, se examinaron los tiempos de activación. Sobre la base del tipo de señales de replicación que indican un origen (Figura 1A), se pudo atribuir el tiempo de iniciación con respecto a los períodos de marcaje. Esto se aplica a orígenes que se dispararon durante los pulsos de IdU o CldU, que duraron cada uno 20 min. (Figuras 1Aii y iii). Esto se aplica igualmente a orígenes que se dispararon antes del período de marcaje, con la condición de que las horquillas salientes se pudiesen visualizar por su incorporación de los nucleótidos modificados (Figura 1Ai). Para que esto suceda, el tiempo de la activación de los orígenes no debería de preceder a los pulsos de IdU/CldU en más de 20 min. de media (por ejemplo, véase la 2ª molécula en la Figura 3Ci). Por lo tanto, la ventana de análisis cubre ~60 min. en total, que comprende 20 min. antes de los pulsos, 20 min. durante el pulso de IdU, y 20 min. durante el pulso de CldU. Los inventores no pudieron encontrar preferencias de tiempo para cualquiera de las 38 zonas de iniciación. Además, los orígenes adyacentes no se dispararon al mismo tiempo (Figura 3C). Se debería observar que fueron detectables múltiples iniciaciones en fibras individuales en los 60 min. dados por nuestro paradigma experimental. Por lo tanto, aunque no se observó sincronía precisa entre los sucesos de iniciación, las diferencias de tiempo entre dos orígenes activados cualesquiera están limitadas a aproximadamente una hora.

Puesto que los tiempos de activación entre orígenes adyacentes estaban ligeramente escalonados, los orígenes potenciales en las regiones no replicadas entre dos horquillas que se aproximan todavía pueden haber sido activados en algún tiempo posterior (para ejemplos, véanse las 3ª y 5ª moléculas en la Figura 3Cii). El disparo de los orígenes a partir de estas regiones producirían distancias entre orígenes más pequeñas. Se considera que la mayoría de los orígenes adyacentes se disparan en 30 min. entre sí (Berezney et al., 2000). Puesto que nuestra ventana de análisis es 60 min. (véase el párrafo anterior), se predice que la mayoría de todos los orígenes en un agrupamiento están activados.

Por lo tanto, es improbable que el disparo retardado de los orígenes alterase significativamente las distancias entre orígenes medidas.

Extensión de la horquilla a través de los sitios de iniciación potenciales

Hasta ahora, las pistas de replicación se han usado solamente con el fin de inferir su sitio de comienzo o su iniciación. Su extensión bidireccional en la región circundante, sin embargo, proporciona otra cantidad de datos importante. Las señales que se originan de una zona de iniciación que solapa una zona de flanco implican para esta última el paso previo de una horquilla de replicación y la retirada de una autorización del origen.

Esto hace a la zona replicada de forma pasiva refractaria a partir del disparo en algún tiempo posterior en la fase S.

En tanto que todos los orígenes potenciales en células humanas son autorizadas ya que están en levadura (Santocanale y Diffley, 1996), esta observación proporciona pruebas de interferencia del origen (Figura 4A). Para horquillas que se extienden parcialmente en una zona de iniciación, las zonas se consideraron solamente como suprimidas si se alcanzó el centroide. Los inventores usaron señales de horquillas de alargamiento para analizar cuán lejos desde un origen activo se produce la interferencia (Figura 4B).

En total, se encontraron que se suprimieron 528 zonas de iniciación. Su distancia desde el sitio de iniciación no difirió significativamente cuando se consideraron los marcadores que representan horquillas que se mueven centroméricas (56,5 +/- 37,7 kb, N = 266) frente a marcadores que representan horquillas que se mueven teloméricas (55,1 +/- 38,4 kb, N = 262; Figura 4C). En consecuencia, para los siguientes análisis, se combinó la interferencia del origen mediada por horquillas que se mueven en ambas direcciones. De media, las pistas de replicación desde orígenes activos solaparon con zonas 55,8 +/- 38 kb de distancia. Las zonas suprimidas más próximas y más alejadas se localizaron 7 kb y 284,6 kb desde el origen funcional, respectivamente. Estos datos se traducen en lo siguiente. La zona situada inmediatamente próxima a un sitio de iniciación se suprimió 314 veces. La segunda zona se suprimió 137 veces, y las zonas subsiguientes después de esa (tres o mayor) se suprimieron 77 veces. Juntos, estos datos basados en 528 zonas de iniciación suprimidas sugieren que la interferencia del origen se extiende en la mayor parte a lo largo de 1-2 zonas de iniciación de flanco.

Además de la interferencia entre zonas, también se analizó la interferencia dentro de la zona.

Las horquillas desde un origen activo se extendieron más allá de las fronteras de su propia zona de iniciación 100% de las veces (Por ejemplo, véase la Figura 4B). Si esta forma de interferencia es robusta, la probabilidad de más de un disparo del origen por zona de iniciación en cualquier ciclo celular dado debería de ser baja. Para llevar a cabo

este análisis, se depende de nuestra capacidad para discriminar tramos de replicación cortos representativos de orígenes próximamente espaciados en zonas de iniciación relativamente pequeñas. La resolución máxima de segmentos fluorescentes lineales en ADN peinado es 1-4 kb. Esto complica la visualización de iniciaciones múltiples en zonas más pequeñas que la media de 13,5 kb. Sin embargo, para las zonas de iniciación más grandes, la observación de varios orígenes no está limitada por la resolución del peinado molecular. Independientemente del tamaño de la zona de iniciación, nunca se observó que dos o más orígenes se disparasen dentro de la misma zona de iniciación en fases S individuales. Por lo tanto, en contraste con la interferencia entre zonas, que disminuye con la distancia desde el suceso de iniciación, la interferencia dentro de la zona es extremadamente eficiente y no depende de la distancia que las horquillas tengan que viajar.

Conclusión

En conclusión, los inventores han cartografiado todas las posibles zonas de iniciación a lo largo de una región de 1,5 Mb en el cromosoma humano 14q11.2 usando un enfoque de una sola molécula (Figuras 1 y 2). También han demostrado que solamente una fracción de ellas son usadas realmente en ciclos celulares individuales (Figura 3).

Por último, la activación limitada de los orígenes produce distancias de disparo entre orígenes regulares (Figura 3).

En consecuencia, los inventores demuestran por primera vez que el espaciamiento de sucesos de iniciación conservado se mantiene en el contexto de zonas de iniciación de mamíferos amplias.

Se ha descrito una relación elevada de origen potencial a origen activo en levadura y células CHO (Raghuraman et al., 2001; Wyrick et al., 2001; Pasero et al., 2002; Anglana et al., 2003). Nuestros datos sugieren que esta relación es un rasgo conservado en células humanas. Esto origina una cuestión importante: ¿Por qué la redundancia del origen es un tema recurrente en células eucariotas? La supresión de varios orígenes en un brazo de un cromosoma de levadura tiene efectos desdeñables sobre la estabilidad del genoma (Dershowitz y Newlon, 1993). Esto sugeriría que tantos orígenes no son necesarios.

Más recientemente, sin embargo, se demostró que la prevención del complemento completo de pre-RC ensamblado dio como resultado reordenamientos cromosómicos (Lengronne y Schwob, 2002; Tanaka y Diffley, 2002). Aunque se desconoce la razón para esto, varias propuestas convergen en la idea de que un exceso de orígenes potenciales proporciona una red segura en el suceso de replicación perturbada del ADN (Schwob, 2004). En primer lugar, si se bloquea una horquilla, se puede convertir en un sustrato para la recombinación (Rothstein et al., 2000). La activación de un origen "extra" aguas abajo da lugar a una horquilla que se aproxima, esta horquilla se funde con la horquilla bloqueada, rescatándola de ese modo de la recombinación.

En segundo lugar, si algunos orígenes fracasan a la hora de dispararse, las células pueden sufrir mitosis con ADN sin replicar. Este fragmento se romperá cuando los centrómeros se separen. Un suministro excesivo de orígenes potenciales reduce la probabilidad de que esto ocurra. Finalmente, la detención óptima del ciclo celular por el punto de control de la fase S requiere un número suficiente de horquillas (Shimada et al., 2002). Las horquillas se pierden cuando fracasa un intento de inicio. El disparo de un origen de respaldo genera dos horquillas adicionales para compensar, haciendo así operacional al punto de control. Claramente, es necesaria una horquilla adicional para evaluar cuál de estos modelos es aplicable.

La interferencia del origen se ha invocado como un mecanismo para explicar cómo se logra en eucariotas una relación elevada de origen potencial a origen activo. Implica la eliminación de pre-RCs, que representan orígenes autorizados, por horquillas que progresan a partir de orígenes activados más tempranos (Brewer y Fangman, 1993).

La interferencia del origen se ha observado en levadura, *Xenopus*, y células CHO (Brewer y Fangman, 1993; Lucas et al., 2000; Anglana et al., 2003). En la presente memoria, los inventores muestran por primera vez que la interferencia del origen desempeña un papel significativo en la modulación de la función del origen en células humanas, y, además, que esto ocurre en el contexto de zonas de iniciación (Figura 4). Antes del peinado molecular, el ADN se desproteína. Por lo tanto, no fue posible observar qué zonas de iniciación contenían orígenes autorizados. De hecho, las zonas replicadas de forma pasiva, que se interpretaron como suprimidas, simplemente pueden no haber sido autorizadas a comenzar. Un trabajo futuro evaluará el ensamblaje de pre-RC entre zonas de iniciación. Esto nos permitirá determinar si la interferencia del origen ocurre según la definición canónica del término.

Para explicar la interferencia del origen, se ha citado el dato de levadura, con lo que los tiempos de activación diferentes establecidos en G1 provocan que un origen tardío se replique pasivamente mediante una horquilla desde un origen temprano (Lucas y Raghuraman, 2003). De acuerdo con el trabajo realizado en células CHO (Anglana et al., 2003), los inventores no encontraron ninguna diferencia de tiempo preprogramado potente para orígenes adyacentes (Figura 3). Por lo tanto, nuestros datos sugieren que en células humanas opera un mecanismo diferente del presente en levadura.

Los inventores observaron que la interferencia entre zonas disminuye gradualmente con la distancia desde el origen activo (Figura 4). Si la probabilidad del disparo del origen es baja debido a factores de iniciación limitados (Walter y

Newport, 1997), la interferencia del origen dada a conocer en la presente memoria puede ser un resultado indirecto de esta baja probabilidad, y, en consecuencia, un fenómeno pasivo. Por otro lado, si las probabilidades del disparo del origen son altas, la interferencia del origen se debe de regular positivamente. Por ejemplo, las proteínas del punto de control que están presentes en horquillas de alargamiento sin perturbar pueden suprimir orígenes distantes desde el disparo.

Esto incrementaría activamente la posibilidad de que los orígenes retrasados se replicasen pasivamente, y, por lo tanto, se suprimiesen (Marheineke e Hyrien, 2004; Shechter et al., 2004; Sorensen et al., 2004; Syljuasen et al., 2005). La investigación futura revelará cuáles de estos modelos es responsable para la interferencia entre zonas.

Recientemente, un estudio matemático propuso que solamente los orígenes potenciales separados 11 kb se pueden secuestrar juntos en un foco de replicación, y, por lo tanto, se pueden activar simultáneamente (Jun et al., 2004). Esta restricción está determinada por la longitud de persistencia del ADN, que limita el doblado del ADN. La longitud de persistencia puede explicar la interferencia dentro de la zona: la rigidez del ADN evita que dos sitios de iniciación potenciales de una zona se concentren en un foco de replicación, evitando así su activación simultánea. La robustez de un mecanismo basado en las propiedades físicas del ADN podría producir la eficiencia elevada de la interferencia dentro de la zona dada a conocer en la presente memoria.

El mecanismo de la interferencia del origen en y entre zonas de iniciación de mamíferos depende de los determinantes moleculares que subyacen a estas regiones. Durante la autorización, múltiples complejos de Mantenimiento de MiniCromosomas (MCM) se extienden desde pre-RCs (Ritzi et al., 1998; Edwards et al., 2002). Se ha sugerido que los orígenes que se disparan en uno de estos sitios de MCM explican la presencia de zonas de iniciación en células de mamíferos (Hyrien et al., 2003; Blow y Dutta, 2005; Cvetic y Walter, 2005). En consecuencia, las zonas de iniciación dadas a conocer en la presente memoria (Figura 2) pueden surgir debido a la carga reiterativa de MCM. La determinación de si los MCMs difusos u otros factores desconocidos son responsables de las zonas de iniciación de mamíferos ayudará a comprender cómo se ejecuta la interferencia del origen humana y cómo se regula el espaciamiento de los sucesos de iniciación.

Referencias

- Anglana, M., Apiou, F., Bensimon, A., y Debatisse, M. (2003). Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* 114, 385-394.
- Bensimon, A., Simon, A., Chiffaudel, A., Croquette, V., Heslot, F., y Bensimon, D. (1994). Alignment and sensitive detection of DNA by a moving interface. *Science* 265, 2096-2098.
- Berezney, R., Dubey, D.D., y Huberman, J.A. (2000). Heterogeneity of eukaryotic replicons, replicon clusters, and replication foci. *Chromosoma* 108, 471-484.
- Blow, J.J., y Dutta, A. (2005). Preventing re-replication of chromosomal DNA. *Nat. Rev. Mol. Cell Biol.* 6, 476-486.
- Brewer, B.J., y Fangman, W.L. (1993). Initiation at closely spaced replication origins in a yeast chromosome. *Science* 262, 1728-1731.
- Cvetic, C., y Walter, J.C. (2005). Eukaryotic origins of DNA replication: could you please be more specific? *Semin. Cell Dev. Biol.* 16, 343-353.
- DePamphilis, M.L. (1999). Replication origins in metazoan chromosomes: fact or fiction? *Bioessays* 21, 5-16.
- Dershowitz, A., y Newlon, C.S. (1993). The effect on chromosome stability of deleting replication origins. *Mol. Cell. Biol.* 13, 391-398.
- Edwards, M.C., Tutter, A.V., Cvetic, C., Gilbert, C.H., Prokhorova, T.A., y Walter, J.C. (2002). MCM2-7 complexes bind chromatin in a distributed pattern surrounding the origin recognition complex in *Xenopus* egg extracts. *J. Biol. Chem.* 277, 33049-33057.
- Friedman, K.L., Brewer, B.J., y Fangman, W.L. (1997). Replication profile of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2, 667-678.
- Gilbert, D.M. (2004). In search of the holy replicator. *Nat. Rev. Mol. Cell Biol.* 5, 848-855.
- Hand, R., y Tamm, I. (1973). DNA replication: direction and rate of chain growth in mammalian cells. *J. Cell Biol.* 58, 410-418.
- Hyrien, O., Marheineke, K., y Goldar, A. (2003). Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *Bioessays* 25, 116-125.

- Jun, S., Herrick, J., Bensimon, A., y Bechhoefer, J. (2004). Persistence length of chromatin determines origin spacing in *Xenopus* early-embryo DNA replication: quantitative comparisons between theory and experiment. *Cell Cycle* 3, 223-229.
- 5 Lebofsky, R., y Bensimon, A. (2005). DNA replication origin plasticity and perturbed fork progression in human inverted repeats. *Mol. Cell. Biol.* 25, 6789-6797.
- 10 Lengronne, A., y Schwob, E. (2002). The yeast CDK inhibitor Sic1 prevents genomic instability by promoting replication origin licensing in late G(1). *Mol. Cell* 9, 1067-1078.
- Lucas, I., Chevrier-Miller, M., Sogo, J.M., y Hyrien, O. (2000). Mechanisms ensuring rapid and complete DNA replication despite random initiation in *Xenopus* early embryos. *J. Mol. Biol.* 296, 769-786.
- 15 Lucas, LA., y Raghuraman, M.K. (2003). The dynamics of chromosome replication in yeast. *Curr. Top. Dev. Biol.* 55, 1-73.
- MacAlpine, D.M., Rodriguez, H.K., y Bell, S.P. (2004). Coordination of replication and transcription along a *Drosophila* chromosome. *Genes Dev.* 18, 3094-3105.
- 20 Marheineke, K., y Hyrien, O. (2004). Control of replication origin density and firing time in *Xenopus* egg extracts: role of a caffeine-sensitive, ATR-dependent checkpoint. *J. Biol. Chem.* 279, 28071-28081.
- 25 Newlon, C.S., Lipchitz, L.R., Collins, I., Deshpande, A., Devenish, R.J., Green, R.P., Klein, H.L., Palzkill, T.G., Ren, R.B., Synn, S., y et al. (1991). Analysis of a circular derivative of *Saccharomyces cerevisiae* chromosome III: a physical map and identification and location of ARS elements. *Genetics* 129, 343-357.
- Norio, P., Kosiyatrakul, S., Yang, Q., Guan, Z., Brown, N.M., Thomas, S., Riblet, R., y Schildkraut, C.L. (2005). Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Mol. Cell* 20, 575-587.
- 30 Pasero, P., Bensimon, A., y Schwob, E. (2002). Single-molecule analysis reveals clustering and epigenetic regulation of replication origins at the yeast rDNA locus. *Genes Dev.* 16, 2479-2484.
- 35 Poloumienko, A., Dershowitz, A., De, J., y Newlon, C.S. (2001). Completion of replication map of *Saccharomyces cerevisiae* chromosome III. *Mol. Biol. Cell* 12, 3317-3327.
- Raghuraman, M.K., Brewer, B.J., y Fangman, W.L. (1997). Cell cycle-dependent establishment of a late replication program. *Science* 276, 806-809.
- 40 Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., y Fangman, W.L. (2001). Replication dynamics of the yeast genome. *Science* 294, 115-121.
- 45 Ritzi, M., Baack, M., Musahl, C., Romanowski, P., Laskey, R.A., y Knippers, R. (1998). Human minichromosome maintenance proteins and human origin recognition complex 2 protein on chromatin. *J. Biol. Chem.* 273, 24543-24549.
- Rothstein, R., Michel, B., y Gangloff, S. (2000). Replication fork pausing and recombination or "gimme a break". *Genes Dev.* 14, 1-10.
- 50 Santocanale, C., y Diffley, J.F. (1996). ORC- and Cdc6-dependent complexes at active and inactive chromosomal replication origins in *Saccharomyces cerevisiae*. *EMBO J.* 15, 6671-6679.
- 55 Santocanale, C., Sharma, K., y Diffley, J.F.X. (1999). Activation of dormant origins of DNA replication in budding yeast. *Genes Dev.* 13, 2360-2364.
- Schwob, E. (2004). Flexibility and governance in eukaryotic DNA replication. *Curr. Opin. Microbiol.* 7, 680-690.
- 60 Shechter, D., Costanzo, V., y Gautier, J. (2004). ATR and ATM regulate the timing of DNA replication origin firing. *Nat. Cell Biol.* 6, 648-655.
- Shimada, K., Pasero, P., y Gasser, S.M. (2002). ORC and the intra-S-phase checkpoint: a threshold regulates Rad53p activation in S phase. *Genes Dev.* 16, 3236-3252.
- 65 Shirahige, K., Iwasaki, T., Rashid, M.B., Ogasawara, N., y Yoshikawa, H. (1993). Location and characterization of autonomously replicating sequences from chromosome VI of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 13, 5043-

5056.

- 5 Sorensen, C.S., Syljuasen, R.G., Lukas, J., y Bartek, J. (2004). ATR, Claspin and the Rad9- Rad1-Hus1 complex regulate Chk1 and Cdc25A in the absence of DNA damage. *Cell Cycle* 3, 941-945.
- 5 Syljuasen, R.G., Sorensen, C.S., Hansen, L.T., Fugger, K., Lundin, C., Johansson, F., Helleday, T., Sehested, M., Lukas, J., y Bartek, J. (2005). Inhibition of human Chk1 causes increased initiation of DNA replication, phosphorylation of ATR targets, and DNA breakage. *Mol. Cell. Biol.* 25, 3553-3562.
- 10 Tanaka, S., y Diffley, J.F. (2002). Deregulated G1-cyclin expression induces genomic instability by preventing efficient pre-RC formation. *Genes Dev.* 16, 2639-2649.
- 15 Todorovic, V., Giadrossi, S., Pelizon, C., Mendoza-Maldonado, R., Masai, H., y Giacca, M. (2005). Human origins of DNA replication selected from a library of nascent DNA. *Mol. Cell* 19, 567-575.
- 15 Vujcic, M., Miller, C.A., y Kowalski, D. (1999). Activation of silent replication origins at autonomously replicating sequence elements near the HML locus in budding yeast. *Mol. Cell. Biol.* 19, 6098-6109.
- 20 Walter, J., y Newport, J.W. (1997). Regulation of replicon size in *Xenopus* egg extracts. *Science* 275, 993-995.
- 20 Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., Bell, S.P., y Aparicio, O.M. (2001). Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* 294, 2357-2360.
- 25 Yamashita, M., Hori, Y., Shinomiya, T., Obuse, C., Tsurimoto, T., Yoshikawa, H., y Shirahige, K. (1997). The efficiency and timing of initiation of replication of multiple replicons of *Saccharomyces cerevisiae* chromosome VI. *Genes Cells* 2, 655-665.

REIVINDICACIONES

1. Método de detección de la presencia de por lo menos un dominio de interés sobre un ácido nucleico que se debe someter a prueba, en el que dicho método comprende las etapas siguientes:

- a) determinar previamente por lo menos tres regiones diana sobre el dominio de interés, diseñar y obtener unas sondas marcadas correspondientes de cada región diana, denominadas conjunto de sondas del dominio de interés, seleccionándose la posición de estas sondas, unas comparadas con las otras y formando una secuencia de por lo menos dos códigos seleccionados de entre un grupo de por lo menos dos códigos diferentes, siendo dicha secuencia de códigos específica de dicho dominio de interés sobre el ácido nucleico que se debe someter a prueba,
 - b) tras extender el ácido nucleico que se debe someter a prueba sobre el que se unen las sondas obtenidas en la etapa a), detectar las señales correspondientes a cada una de las sondas marcadas unidas sobre el ácido nucleico linealizado,
 - c) transcribir dichas señales en un código de secuencia, en el que dicho código resulta de la lectura del espaciamiento entre las señales consecutivas de las sondas,
- indicando la detección de la secuencia de códigos de un dominio de interés la presencia de dicho dominio de interés sobre el ácido nucleico que se va a someter a prueba, y al contrario indicando la ausencia de detección de la secuencia de códigos o parte de la secuencia de códigos de un dominio de interés la ausencia de dicho dominio o parte de dicho dominio de interés sobre el ácido nucleico que se debe someter a prueba.

2. Método según la reivindicación 1, en el que el método es para determinar la presencia de por lo menos dos dominios de interés.

3. Método según la reivindicación 1, en el que la posición del dominio de interés se utiliza como referencia para localizar una reacción química o bioquímica.

4. Método según la reivindicación 1, en el que algunas de las sondas de la región diana son asimismo parte de la secuencia de códigos de por lo menos otro dominio de interés situado próximo sobre el ácido nucleico.

5. Método según la reivindicación 1, en el que todas las sondas se marcan con el mismo marcador.

6. Método según la reivindicación 1, en el que las sondas se marcan con por lo menos dos marcadores diferentes.

7. Método según la reivindicación 1, en el que las sondas son oligonucleótidos de por lo menos 1 Kb.

8. Método según la reivindicación 1, en el que la linealización del ácido nucleico se realiza mediante peinado molecular o Fish sobre fibras.

9. Método según la reivindicación 1, en el que la unión de dichas por lo menos tres sondas forma una secuencia de por lo menos dos códigos seleccionados de entre: espacios pequeños que miden entre 25-30 kb, y espacios grandes que miden entre 55-70 kb.

10. Método según la reivindicación 1, en el que el conjunto de sondas comprende además dos sondas (sonda 1 o sonda 2), pudiendo cada sonda unirse sobre un extremo diferente del dominio de interés, permitiendo la lectura de la señal de una de dichas sonda 1 o sonda 2 asociada con su sonda consecutiva en el dominio de interés, denominadas "par de sondas de extremo de comienzo o final", obtener una información de comienzo o final de lectura.

11. Método según la reivindicación 10, en el que la información de comienzo de lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas del par de sondas de extremo de comienzo.

12. Método según la reivindicación 10, en el que la información de final de lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas del par de sondas de extremo de final.

13. Método según la reivindicación 10, en el que la información de comienzo de lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas del par de sondas de extremo de comienzo y la información de final de lectura resulta de la lectura del espaciamiento entre las dos sondas consecutivas del par de sondas de extremo de final, siendo dicho espaciamiento diferente para el par de sondas de extremo de comienzo y el par de sondas de extremo de final para diferenciar la información del comienzo y del final.

14. Método según la reivindicación 1, en el que las sondas se marcan con un marcador fluorescente o con un

marcador radioactivo.

5 15. Método según la reivindicación 1, en el que la secuencia de códigos comprende un espacio entre la primera y la segunda sonda en un conjunto de sondas, siendo el espacio diferente de todos los otros espacios en las secuencias de códigos y el espacio se puede utilizar para obtener información sobre el comienzo de la secuencia de códigos.

16. Método según la reivindicación 1, en el que la secuencia de códigos comprende un espacio entre la penúltima y la última sonda en un conjunto de sondas, siendo el espacio diferente de todos los otros espacios en la secuencia de códigos y el espacio se puede utilizar para obtener información sobre el final de la secuencia de códigos.

Figura 1

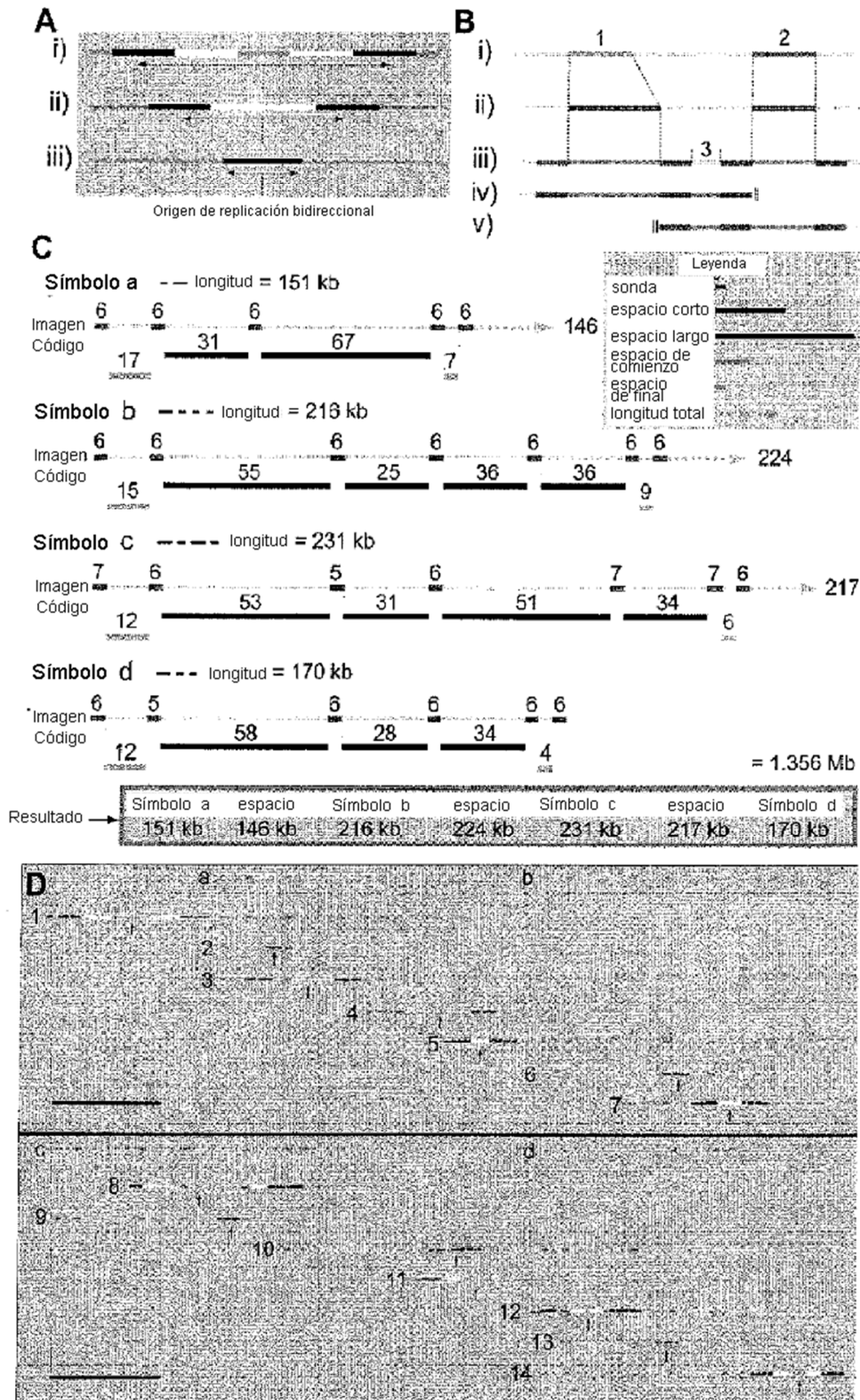


Figura 2

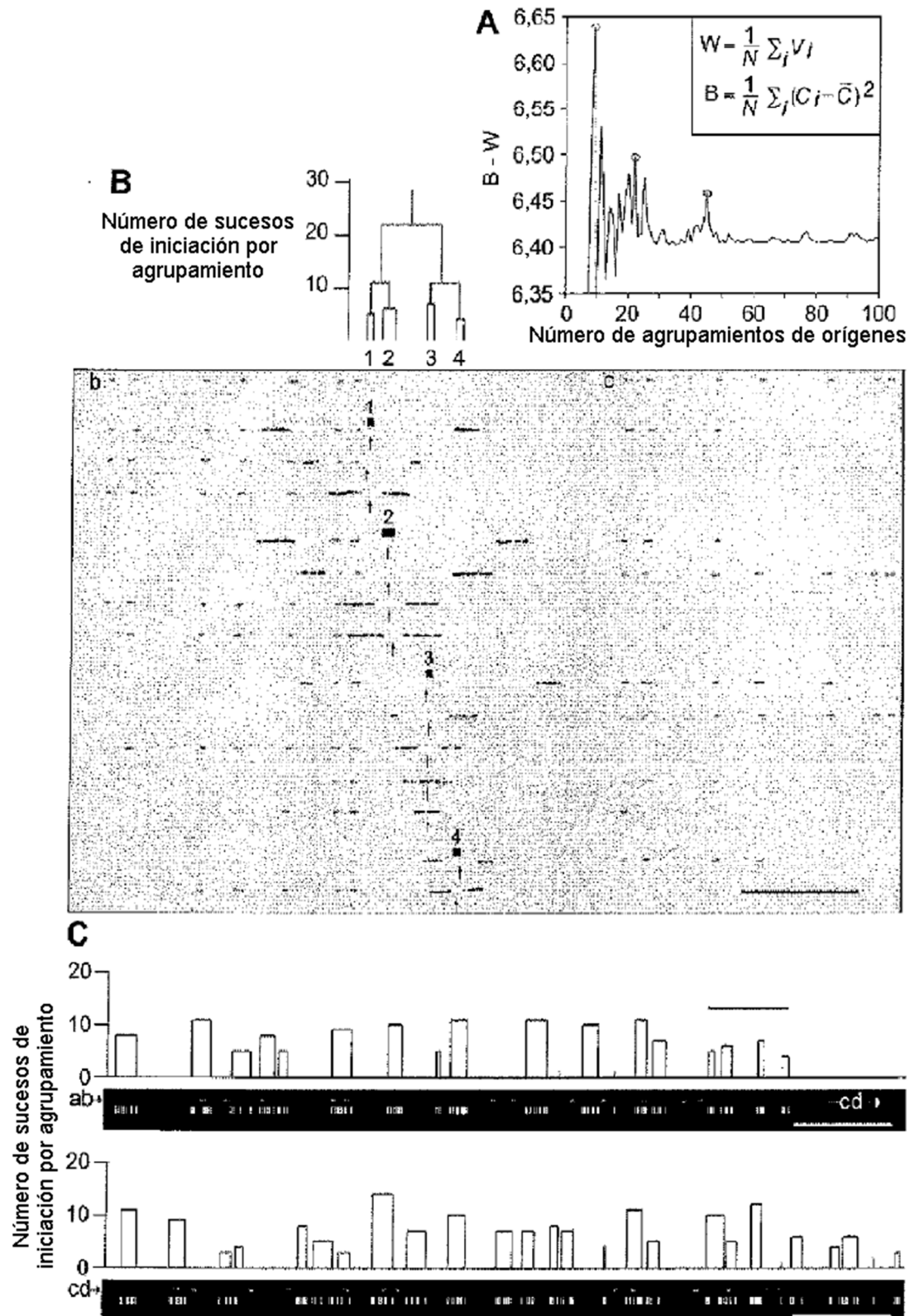


Figura 3

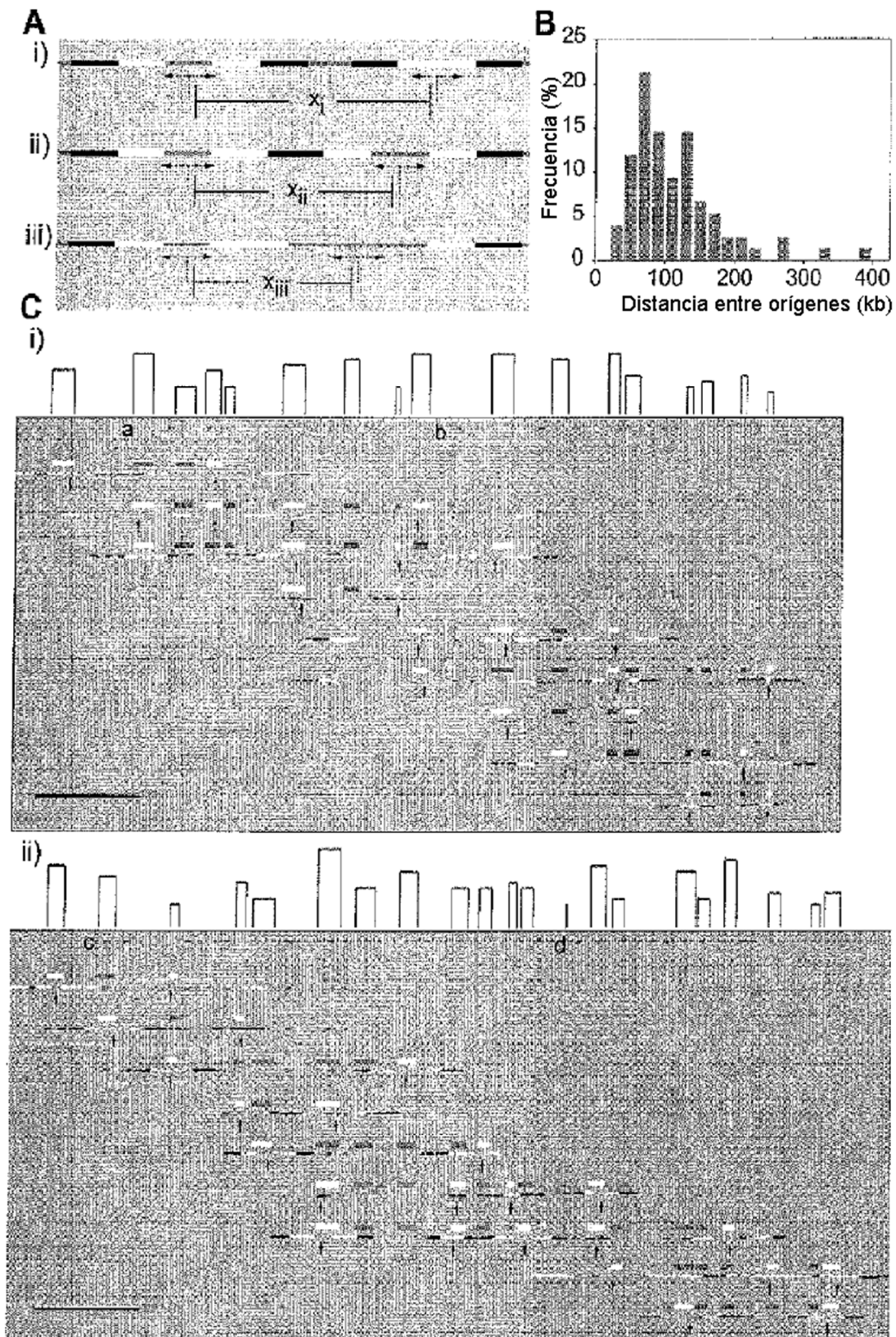


Figura 4

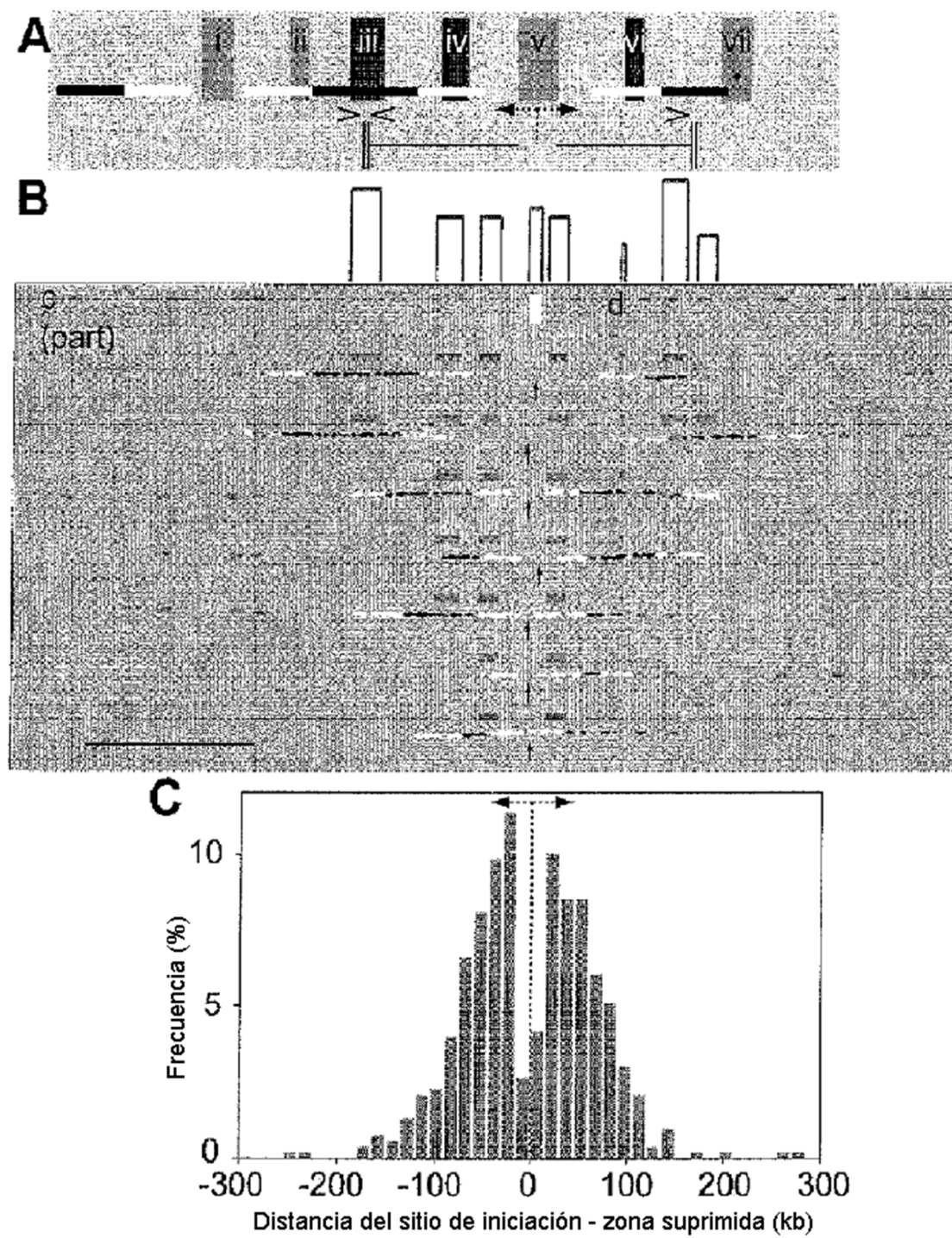
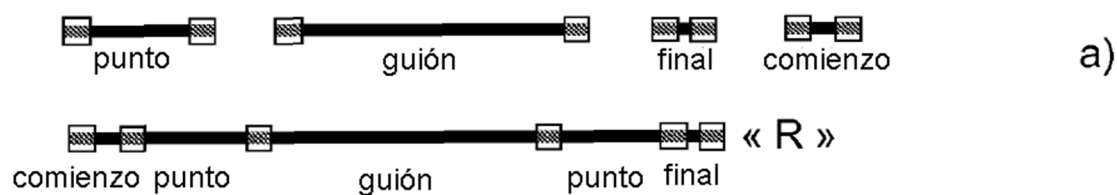


Figura 5

• ■ • « R » en Código Morse

Código basado en el espaciamento



Código basado en el color



Código redundante

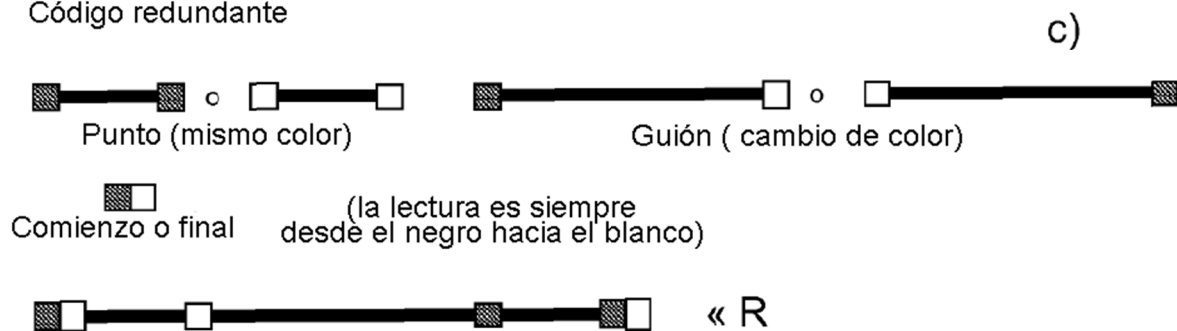


Figura 6

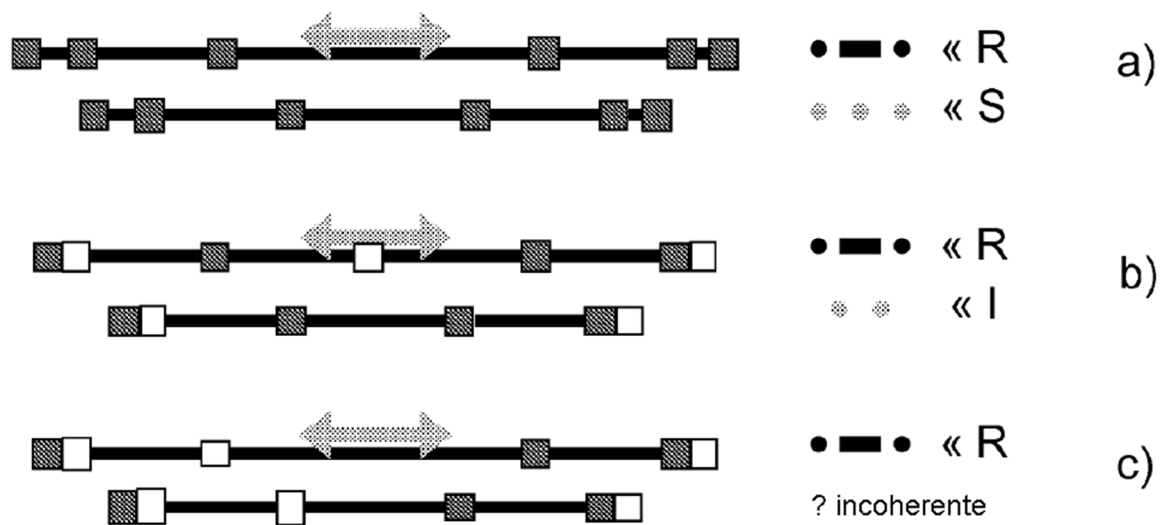


Figura 7

```
aacbbcb
 acbbcba
  cbbcbaa
   bbcbaac
    bcbaacb
     cbaacbb
      baacbbc
       aacbbcc
```

aacbbcbaacbbcc

[illegible]