

19



OFICINA ESPAÑOLA DE  
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 634 266**

51 Int. Cl.:

**C12Q 1/68**

(2006.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

86 Fecha de presentación y número de la solicitud internacional: **10.01.2008 PCT/IB2008/000625**

87 Fecha y número de publicación internacional: **17.07.2008 WO08084405**

96 Fecha de presentación y número de la solicitud europea: **10.01.2008 E 08719319 (9)**

97 Fecha y número de publicación de la concesión europea: **21.06.2017 EP 2121977**

54 Título: **Captura de conformación cromosómica circular (4C)**

30 Prioridad:

**11.01.2007 US 999750 P**  
**05.10.2007 US 977900 P**

45 Fecha de publicación y mención en BOPI de la traducción de la patente:  
**27.09.2017**

73 Titular/es:

**ERASMUS UNIVERSITY MEDICAL CENTER**  
**(100.0%)**  
**FACULTY OF MEDICINE DEPT. OF CELL**  
**BIOLOGY AND GENETICS P.O. BOX 1738**  
**3000 DR ROTTERDAM, NL**

72 Inventor/es:

**DE LAAT, WOUTER y**  
**GROSVELD, FRANK**

74 Agente/Representante:

**ISERN JARA, Jorge**

**ES 2 634 266 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

Captura de conformación cromosómica circular (4C)

5 Campo de la invención

La presente invención se refiere a los análisis de la frecuencia de interacción de dos o más secuencias de nucleótidos en el espacio nuclear. Como una herramienta para detectar reordenamientos del genoma para el diagnóstico y pronóstico se utilizan cambios en interacciones.

10

Antecedentes de la invención

El objetivo de los estudios realizados sobre arquitectura nuclear en mamíferos es entender cómo se pliegan 2 metros de ADN en un núcleo de 10 µm de ancho, permitiendo al mismo tiempo la expresión exacta de los genes que especifican el tipo de célula, y cómo esto se propaga fielmente durante cada ciclo celular. Los progresos realizados en este campo provienen mayormente de estudios realizados con microscopía, que revelan que los genomas se ordenan de un modo no aleatorio en el espacio nuclear. Por ejemplo, la heterocromatina densamente empaquetada se separa de la eucromatina más abierta y los cromosomas ocupan distintos territorios en el espacio nuclear. Existe una relación compleja entre el posicionamiento nuclear y la actividad transcripcional. Aunque la transcripción se produce por todo el interior nuclear, los genes activos, que se agrupan en cromosomas, se localizan preferentemente en el borde o en el exterior de su territorio cromosómico. Los genes individuales pueden migrar sobre cambios en su estado de transcripción, como se mide frente a referencias nucleares relativamente grandes tales como territorios de cromosomas, centrómeros o la periferia nuclear. Además, genes activamente transcritos, decenas de megabases separados del cromosoma, pueden reunirse en el núcleo, como se demuestra recientemente por hibridación *in situ* con fluorescencia (FISH, *fluorescence in situ hybridization*) para el locus de la β globina y algunos otros genes seleccionados. Además de la transcripción, la organización genómica está asociada con la coordinación de la replicación, recombinación y con la probabilidad de que los locus se transloquen (lo que conduce a neoplasias) y con el establecimiento y restablecimiento de programas epigenéticos. Basándose en estas observaciones se piensa que la organización arquitectónica del ADN en el núcleo celular es un contribuidor clave de la función genómica.

30

Se han desarrollado diferentes ensayos para permitir una visión de la organización espacial de los locus genómicos *in vivo*. Se ha desarrollado un ensayo denominado ARN-TRAP (Carter *et al.* (2002) Nat. Genet. 32, 623) que implica el direccionamiento de la peroxidasa de rábano picante (HRP, *horseradish peroxidase*) a transcritos de ARN nacientes, seguido por la cuantificación de la deposición de biotina catalizada por HRP en la cromatina cercana.

35

Otro ensayo que se ha desarrollado se denomina tecnología de captura de conformación cromosómica (3C), que proporciona una herramienta para estudiar la organización estructural de una región genómica. La tecnología 3C implica el análisis de PCR cuantitativa de frecuencias de entrecruzamiento entre dos fragmentos de restricción de ADN determinados, que proporciona una medida de su proximidad en el espacio nuclear (véase la Figura 1). Originalmente desarrollada para analizar la conformación de los cromosomas en las levaduras (Dekker *et al.*, 2002), esta tecnología se ha adaptado para investigar la relación entre la expresión génica y el plegamiento de la cromatina en grupos complejos de genes de mamíferos (véase, por ejemplo, Tolhuis *et al.* 2002, Palstra *et al.*, 2003 y Drissen *et al.*, 2004). En resumen, la tecnología 3C implica el entrecruzamiento de células *in vivo* con formaldehído y la digestión nuclear de la cromatina con una enzima de restricción, seguido de ligamiento de fragmentos de ADN que se entrecruzaron en un complejo. Los productos de ligamiento se cuantifican después por PCR. La etapa de amplificación por PCR requiere el conocimiento de la información de la secuencia de cada uno de los fragmentos de ADN que se va a amplificar. Por tanto, la tecnología 3C proporciona una medida de las frecuencias de interacción entre los fragmentos de ADN seleccionados.

50

La tecnología 3C se ha desarrollado para identificar elementos que interactúan entre partes seleccionadas del genoma y ambas técnicas requieren el diseño de cebadores para todos los fragmentos de restricción analizados. Recientemente se han desarrollado nuevas estrategias que permiten explorar todo el genoma, de una manera imparcial, de segmentos de ADN que interactúan físicamente con un fragmento de ADN de elección. Dichas estrategias se basan en tecnología 3C y conjuntamente reciben el nombre de "tecnología 4C". La tecnología 4C permite explorar todo el genoma, de una manera imparcial, de segmentos de ADN que físicamente interactúan con un fragmento de ADN de elección. La tecnología 4C depende del ligamiento selectivo de fragmentos de ADN entrecruzados con un fragmento de restricción de elección (el "cebo"). En la tecnología 4C, todos los fragmentos de ADN capturados por el cebo en la población de células se amplifican simultáneamente por PCR inversa, utilizando dos cebadores específicos de cebo que amplifican a partir de productos de ligamiento circularizados.

60

Esencialmente, se pueden seguir dos estrategias para obtener estos círculos de ADN. Una estrategia se basa en la formación de círculos durante la etapa de ligamiento de 3C convencional, es decir, mientras que el ADN está aún entrecruzado (Zhao *et al.* (2006) Nat Genet 38, 1341-7). En este caso, la formación del círculo requiere que ambos extremos del fragmento de cebo se ligen a ambos extremos de un fragmento de restricción capturado. Si varios fragmentos de restricción están entrecruzados entre sí, todavía pueden formarse círculos pero pueden contener más

65

de un fragmento capturado y por tanto será más grande. Después del desentrecruzamiento, los fragmentos de ADN capturados se amplifican directamente por PCR inversa, utilizando cebadores específicos de cebo orientados hacia el exterior. En esta configuración puede utilizarse enzimas de restricción que reconocen cuatro o seis pares de bases. No obstante, en este método se prefieren cuatro cortadores ya que producen fragmentos de restricción más pequeños (de un tamaño promedio de 256 pb, frente a ~ 4 kb para seis cortador) y la amplificación por PCR lineal de los fragmentos de ADN capturados requiere que el tamaño promedio del producto sea pequeño. Esencialmente, este método comprende por tanto las etapas de: (a) proporcionar una muestra de ADN entrecruzado; (b) digerir el ADN entrecruzado con una enzima de restricción primaria - tal como una enzima cortadora de 4 pb o de 5 pb; (c) ligar las secuencias de nucleótidos entrecruzadas; (d) invertir el entrecruzamiento y (e) amplificar la una o más secuencias de nucleótidos de interés utilizando al menos dos cebadores oligonucleotídicos, en el que cada cebador se hibrida con las secuencias de ADN que flanquean las secuencias de nucleótidos de interés. La secuencia o secuencias amplificadas pueden hibridarse con una matriz para ayudar a determinar la frecuencia de interacción entre las secuencias de ADN.

La segunda estrategia se basa ventajosamente en la formación de círculos de ADN después de que la cromatina se haya desentrecruzado como se describe en el presente documento y en la solicitud relacionada WO2007/004057. Como se describe en esta, la tecnología 4C permite una investigación ologénica imparcial de fragmentos de ADN que interactúan con un locus de elección. En resumen, el análisis 3C se realiza como de costumbre, pero omitiendo la etapa de la PCR. El molde 3C contiene una secuencia diana o "cebo" (por ejemplo un fragmento de restricción de elección que incluye un gen seleccionado) ligado a muchas secuencias de nucleótidos de interés diferentes (representando este entorno genómico de genes). El molde se escinde con otra enzima de restricción secundaria y posteriormente vuelve a ligarse para formar círculos de ADN pequeños. Ventajosamente, la una o más secuencias de nucleótidos de interés que se ligan con la secuencia de nucleótidos diana se amplifican utilizando al menos dos cebadores oligonucleotídicos, donde al menos un cebador se hibrida con la secuencia diana. El segundo cebador también se hibrida preferentemente con la secuencia diana, de tal manera que ambos cebadores flanquean el nucleótido de interés. Como alternativa, el segundo cebador se hibrida con una secuencia adaptadora que está ligada al sitio de restricción secundario, de tal manera que los dos cebadores flanquean el nucleótido de interés. Típicamente, esto produce un patrón de fragmentos de PCR que es muy reproducible entre reacciones de amplificación, independientes y específicas, para un tejido determinado. Como enzimas de restricción primaria y secundaria puede utilizarse *HindIII* y *DpnII*. A continuación, los fragmentos amplificados pueden marcarse e hibridarse opcionalmente con una matriz, típicamente contra una muestra control que contiene ADN genómico digerido con la misma combinación de enzimas de restricción. Por lo tanto, la tecnología 3C se ha modificado de tal manera, que todas las secuencias de nucleótidos de interés que interactúan con una secuencia de nucleótidos diana se amplifican. Prácticamente esto significa que, en lugar de realizar una reacción de amplificación con cebadores que son específicos para los fragmentos que se desea analizar, se realiza una amplificación utilizando uno o más cebadores oligonucleotídicos que se hibridan con una secuencia de ADN que flanquea las secuencias de nucleótidos de interés. Ventajosamente, la tecnología 4C no está sesgada hacia el diseño de cebadores de PCR que se incluyen en la etapa de amplificación de PCR y por lo tanto puede utilizarse para investigar el genoma completo para la interacción de elementos de ADN.

Existe una necesidad importante de tecnología de alto rendimiento que pueda explorar sistemáticamente todo el genoma de una manera imparcial para los locus de ADN que entran en contacto entre sí en el espacio nuclear.

Además, existe una necesidad de mejorar dichas tecnologías que permitan realizar el análisis simultáneo de interacciones múltiples que tienen lugar con secuencias múltiples en el genoma y para analizar el genoma para inserciones, deleciones, translocaciones, inversiones y reordenamientos que tienen lugar en lugares desconocidos y que pueden asociarse con una enfermedad.

La presente invención pretende proporcionar mejoras en la tecnología 3C y 4C y en técnicas relacionadas con las mismas.

#### Sumario de la invención

La tecnología 3C, tal como se aplica actualmente, solo permite el análisis de un número limitado de interacciones ADN-ADN seleccionadas debido a las limitaciones de la etapa de amplificación por PCR, que requiere el conocimiento de la información de secuencia específica para cada fragmento que se va a analizar. Además, la selección de fragmentos de restricción como candidatos para interacciones de ADN de largo alcance requiere una cantidad sustancial de conocimiento previo (por ejemplo, la localización de sitios hipersensibles) del locus de interés, que normalmente no está disponible. Dada la importancia funcional de muchas interacciones ADN-ADN de largo alcance descritas hasta ahora, la capacidad de explorar al azar elementos de ADN que se enlazan a una secuencia de interés -tal como un promotor génico, un potenciador, un aislante, un silenciador, un origen de replicación o MAR/SAR - o a una región genómica de interés - tal como una región densa en genes, una región pobre en genes o elemento repetitivo - puede facilitar enormemente el mapeo de secuencias implicadas en una red reguladora.

La presente invención se refiere a tecnología 4C (es decir capturar y caracterizar cromatina colocalizada) y a mejoras en la misma, que proporciona el análisis de alto rendimiento de la frecuencia de interacción de dos o más

secuencias de nucleótidos en el espacio nuclear. La tecnología 4C puede utilizarse para identificar interacciones ADN-ADN de largo alcance (por ejemplo estudiar el plegamiento de los cromosomas), aunque también para identificar reordenamientos genómicos equilibrados y no equilibrados - tales como translocaciones, inversiones, delecciones, amplificaciones, etc.- que pueden ser la causa de un rasgo o enfermedad en sujetos humanos.

Habitualmente, la tecnología 4C implica el uso de micromatrices para analizar los fragmentos de ADN capturados por una sola secuencia diana seleccionada ("cebo") (Simonis *et al.*, Nature Genetics 2006). Las micromatrices tienen la desventaja de que ofrecen un intervalo dinámico limitado dado que las sondas presentes en las matrices pueden saturarse, lo que dificulta el análisis cuantitativo de intensidades de señal. La secuenciación de alto rendimiento franquea este problema ya que ofrece un intervalo dinámico ilimitado. También es cuantitativa ya que proporciona números absolutos de secuencias.

Adicionalmente, se prefiere analizar interacciones de ADN con varias secuencias diana simultáneamente. Esto es cierto para todas las aplicaciones basadas en 4C y, en particular, para los análisis de reordenamientos genómicos basados en 4C. La tecnología 4C puede utilizarse como una herramienta de diagnóstico que permite escanear todo el genoma de una manera imparcial con respecto a la presencia de reordenamientos genómicos. Puede utilizarse una serie de secuencias diana a lo largo de cada cromosoma que capture conjuntamente todas las secuencias (es decir fragmentos de restricción). Posteriormente, para identificar el reordenamiento genómico, se identifica el "cebo" mediante el cual se capturó la secuencia.

Los fragmentos capturados pueden secuenciarse individualmente para cada secuencia diana ("cebo"). Sin embargo, preferentemente, todos los productos de ligamiento formados con todas las secuencias diana se analizan simultáneamente. Para ello, cada lectura debe dirigirse a la unión de ligamiento y proporcionar suficiente información de secuencia para identificar inequívocamente tanto la secuencia diana como la secuencia capturada.

#### Aspectos sumario de la presente invención

Los aspectos de la presente invención se exponen en las reivindicaciones acompañantes.

En un primer aspecto, se proporciona un método para analizar la frecuencia de interacción de una secuencia de nucleótidos diana con una o más secuencias de nucleótidos de interés (por ejemplo uno o más locus genómicos) que comprende las etapas de: (a) proporcionar una muestra de ADN entrecruzado; (b) digerir el ADN entrecruzado con una enzima de restricción primaria; (c) ligar las secuencias de nucleótidos entrecruzadas; (d) invertir el entrecruzamiento; (e) opcionalmente digerir las secuencias de nucleótidos con una enzima de restricción secundaria; (f) opcionalmente ligar una o más secuencias de ADN, de composición de nucleótidos conocida, con el sitio o sitios de digestión con enzimas de restricción secundarias disponibles que flanquean la una o más secuencias de nucleótidos de interés; (g) amplificar la una o más secuencias de nucleótidos de interés utilizando al menos dos cebadores oligonucleotídicos, en el que cada cebador se hibrida con las secuencias de ADN que flanquean las secuencias de nucleótidos de interés; (h) hibridar la secuencia o secuencias amplificadas con una matriz; e (i) determinar la frecuencia de interacción entre las secuencias de ADN.

En un segundo aspecto, se proporciona un método para analizar la frecuencia de interacción de una secuencia de nucleótidos diana con una o más secuencias de nucleótidos (por ejemplo uno o más locus genómicos) que comprende las etapas de: (a) proporcionar una muestra de ADN entrecruzado; (b) digerir el ADN entrecruzado con una enzima de restricción primaria; (c) ligar las secuencias de nucleótidos entrecruzadas; (d) invertir el entrecruzamiento; (e) opcionalmente digerir las secuencias de nucleótidos con una enzima de restricción secundaria; (f) circularizar las secuencias de nucleótidos; (g) amplificar la una o más secuencias de nucleótidos que están ligadas con la secuencia de nucleótidos diana; (h) opcionalmente hibridar las secuencias amplificadas con una matriz o analizar las secuencias amplificadas mediante secuenciación de alto rendimiento; e (i) determinar la frecuencia de interacción entre las secuencias de ADN.

En un tercer aspecto, se proporciona un método para identificar una o más interacciones ADN-ADN que son indicativas de una patología particular que comprende la etapa de realizar las etapas (a)-(i) del método de acuerdo con el primer aspecto, en el que en la etapa (a) se proporciona una muestra de ADN entrecruzado de una célula enferma y de una célula no enferma, y en el que una diferencia entre la frecuencia de interacción entre las secuencias de ADN de las células enfermas y no enfermas, indica una diferencia en la organización lineal de los moldes cromosómicos (por ejemplo, un reordenamiento genómico), que es indicativa de un rasgo o patología particular.

En un cuarto aspecto, se proporciona un método de diagnóstico o pronóstico de una enfermedad o síndrome causado por, o asociado a, un cambio en una interacción ADN-ADN que comprende la etapa de realizar las etapas (a)-(i) del método de acuerdo con el primer aspecto, en el que la etapa (a) comprende proporcionar una muestra de ADN entrecruzado de un sujeto; y en el que la etapa (i) comprende comparar la frecuencia de interacción entre las secuencias de ADN con la de un control no afectado; en el que una diferencia entre el valor obtenido del control y el valor obtenido del sujeto es indicativa de que el sujeto padece la enfermedad o síndrome o es indicativa de que el sujeto padecerá la enfermedad o síndrome.

5 En un quinto aspecto, se proporciona método de diagnóstico o pronóstico de una enfermedad o síndrome causado por, o asociado a, un cambio en una interacción ADN-ADN que comprende la etapa de: realizar las etapas (a)-(i) del método de acuerdo con el primer aspecto, en el que la etapa (a) comprende proporcionar una muestra de ADN entrecruzado de un sujeto; y en el que dicho método comprende la etapa adicional de: (j) identificar uno o más locus que hayan experimentado un reordenamiento genómico que está asociado a la enfermedad.

10 En un sexto aspecto, se proporciona un método de ensayo para identificar uno o más agentes que modulan una interacción ADN-ADN que comprende las etapas de: (a) poner en contacto una muestra con uno o más agentes; y (b) realizar las etapas (a) a (i) del método de acuerdo con el primer aspecto, en el que la etapa (a) comprende proporcionar ADN entrecruzado de la muestra; en el que una diferencia entre (i) la frecuencia de interacción entre las secuencias de ADN en presencia del agente y (ii) la frecuencia de interacción entre las secuencias de ADN en ausencia del agente, es indicativa de un agente que modula la interacción ADN-ADN.

15 En un séptimo aspecto, se proporciona un método para detectar la localización de un reordenamiento equilibrado y/o no equilibrado (por ejemplo, una translocación) que comprende la etapa de: (a) realizar las etapas (a) a (i) del método de acuerdo con el primer aspecto; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con las de un control; en el que una transición de baja a alta frecuencia de interacción ADN-ADN en la muestra, en comparación con el control, es indicativa de la localización de un punto de corte.

20 En un octavo aspecto, se proporciona un método para detectar la localización de una inversión equilibrada y/o no equilibrada, que comprende las etapas de: (a) realizar las etapas (a) a (i) del método de acuerdo con el primer aspecto; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; en el que un patrón invertido de frecuencias de interacción ADN-ADN para la muestra en comparación con el control es indicativo de una inversión.

25 En un noveno aspecto, se proporciona un método para detectar la localización de una delección que comprende las etapas de: (a) realizar las etapas (a) a (i) del método de acuerdo con el primer aspecto; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; en el que una reducción en la frecuencia de interacción ADN-ADN para la muestra en comparación con el control es indicativa de delección.

30 En un décimo aspecto, se proporciona un método para la detección de la localización de una duplicación que comprende las etapas de: (a) realizar las etapas (a) a (i) del método de acuerdo con el primer aspecto; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; en el que un aumento o una disminución en la frecuencia de interacción ADN-ADN para la muestra del sujeto en comparación con el control es indicativo de una duplicación o inserción.

35 En un decimoprimer aspecto, se proporciona un agente obtenido, o que puede obtenerse, mediante el método de ensayo descrito en el presente documento.

40 En un duodécimo aspecto, se proporciona un método para analizar la frecuencia de interacción de una o más secuencias de nucleótidos diana con una o más secuencias de nucleótidos de interés (por ejemplo, uno o más locus genómicos) que comprende las etapas de: (a) proporcionar una muestra de ADN entrecruzado; (b) digerir el ADN entrecruzado con una enzima de restricción primaria; (c) ligar las secuencias de nucleótidos entrecruzadas; (d) invertir el entrecruzamiento; y (e) secuenciar las secuencias de nucleótidos ligadas.

45 En un decimotercer aspecto, se proporciona un método para determinar la presencia de un reordenamiento genómico en una muestra que comprende las etapas de: (a) proporcionar una muestra de ácido nucleico (por ejemplo ADN genómico), en el que dicho ácido nucleico comprende una secuencia de nucleótidos de secuencia conocida adyacente a la localización del reordenamiento genómico sospechado; (b) digerir el ADN con una enzima de restricción primaria para formar una pluralidad de fragmentos de restricción; (c) opcionalmente, purificar los fragmentos de restricción; (d) ligar los fragmentos de restricción para formar ADN circularizado; (e) opcionalmente, purificar el ADN circularizado; (f) digerir el ADN circularizado con una enzima de restricción secundaria para formar una pluralidad de fragmentos de restricción; (g) ligar los fragmentos de restricción para formar ADN circularizado; (h) amplificar el reordenamiento genómico sospechado utilizando uno o más cebadores que se hibridan con la secuencia de nucleótidos de secuencia conocida; e (i) secuenciar el reordenamiento genómico sospechado.

50 En un decimocuarto aspecto, se proporciona una base de datos de secuencias de ácido nucleico de aproximadamente 6-50 pares de bases que flanquean directamente, y opcionalmente incluyen, el sitio de reconocimiento de la enzima de restricción primaria o el sitio de reconocimiento de la enzima de restricción secundaria de cada secuencia diana.

60 En un decimoquinto aspecto se proporciona una base de datos de secuencias de ácido nucleico de aproximadamente 12-50 pares de bases que flanquean directamente todos los sitios de reconocimiento de enzimas de restricción primaria y secundaria en el genoma.

65

En un decimosexto aspecto, se proporciona el uso de la base de datos de secuencias de ácido nucleico para determinar la posición genómica de cada una de las secuencias capturadas identificadas.

5 En un decimoséptimo aspecto, se proporciona un método o un agente o una base de datos o un uso como se describe esencialmente en el presente documento y con referencia a cualquiera de los Ejemplos o Figuras.

Realizaciones de la invención

10 Adecuadamente, la reacción de ligamiento en la etapa (c) o (f) da como resultado la formación de círculos de ADN.

Adecuadamente, la etapa (h) comprende el análisis de los productos de ligamiento entre secuencias diana y secuencias de interés entrecruzadas por medio de secuenciación (por ejemplo, secuenciación a alto rendimiento).

15 Adecuadamente, el método es para analizar la frecuencia de interacción de dos o más secuencias de nucleótidos diana con una o más secuencias de nucleótidos de interés, que comprende el uso de PCR múltiple en la etapa (g).

20 Adecuadamente, el método es para analizar la frecuencia de interacción de dos o más secuencias de nucleótidos diana con una o más secuencias de nucleótidos de interés, que comprende el agrupamiento de algunos o de todos los productos de PCR obtenidos para cada una de las secuencias diana en la etapa (g) y posterior análisis simultáneo de sus interacciones de ADN.

Adecuadamente, dos o más secuencias amplificadas se marcan diferencialmente antes del agrupamiento y análisis por hibridación con una matriz.

25 Adecuadamente, dos o más secuencias amplificadas se marcan idénticamente y se analizan por hibridación con una matriz cuando las secuencias residen en diferentes cromosomas.

30 Adecuadamente, dos o más secuencias amplificadas se marcan idénticamente cuando las secuencias residen en el mismo cromosoma a una distancia que está lo suficientemente lejos para que se produzca un solapamiento mínimo entre señales de interacción ADN-ADN.

Adecuadamente, se utiliza secuenciación a alto rendimiento para analizar las uniones de ligamiento formadas entre secuencias diana y secuencias de interés capturadas.

35 Adecuadamente, la secuenciación se dirige a las uniones de ligamiento formadas entre las secuencias diana y las secuencias de interés capturadas por la adición de secuencias adaptadoras necesarias para la secuenciación en los extremos de las secuencias amplificadas.

40 Adecuadamente, la secuenciación se dirige a las uniones de ligamiento formadas entre las secuencias diana y las secuencias de interés capturadas por la adición de las secuencias adaptadoras completas, o partes de las mismas, necesarias para la secuenciación como salientes 5' con los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés.

45 Adecuadamente, la secuenciación se dirige a las uniones del ligamiento formadas entre las secuencias diana y las secuencias de interés capturadas por la conjugación de una sustancia de biotina u otra fracción con los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés, seguido de purificación mediada por estreptavidina o de cualquier otra manera, del material amplificado por PCR.

50 Adecuadamente, la secuenciación se dirige a las uniones de ligamiento entre las secuencias diana y las secuencias de interés capturadas diseñando los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés entre 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 nucleótidos desde el sitio o sitios de reconocimiento de las enzimas de restricción primaria y/o secundaria analizadas.

55 Adecuadamente, la secuenciación se dirige a las uniones de ligamiento entre las secuencias diana y las secuencias de interés capturadas diseñando los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés de tal manera que se solapan parcial o completamente con el sitio de reconocimiento de las enzimas de restricción primaria y/o secundaria analizadas.

60 Adecuadamente, las secuencias se leen a través de la unión de ligamiento de tal manera que cuando analizan muestras de PCR múltiples o agrupadas, se obtiene suficiente información de secuencia (por ejemplo 12 nucleótidos o más) en cualquier lado de la unión de ligamiento para identificar con claridad cada secuencia diana y cada secuencia de interés capturada.

65 Adecuadamente, la secuencia de nucleótidos diana se selecciona del grupo que consiste en un reordenamiento genómico, un promotor, un potenciador, un silenciador, un aislante, una región de unión a matriz, una región de

control de locus, una unidad de transcripción, un origen de replicación, un punto caliente de recombinación, un punto de corte de translocación, un centrómero, un telómero, una región densa en genes, una región pobre en genes, un elemento repetitivo y un sitio de integración (vírico).

5 Adecuadamente, la secuencia de nucleótidos diana es una secuencia de nucleótidos que está asociada a una enfermedad o que la causa, o que está localizada a una distancia de hasta 15 Mb, o mayor, en un molde de ADN lineal de un locus que está asociado a una enfermedad o que la causa.

10 Adecuadamente, la secuencia de nucleótidos diana se selecciona del grupo que consiste en: *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, *IGH*, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCR $\alpha$  $\delta$* , *TCR $\beta$*  y *HOX* u otros locus asociados a la enfermedad, como se describe en "Catalogue of Unbalanced Chromosome Aberrations in Man" 2ª edición. Albert Schinzel. Berlín: Walter de Gruyter, 2001. ISBN 3-11-011607-3.

15 Adecuadamente, las secuencias diana se distribuyen a lo largo del molde del genoma lineal, de tal manera que las secuencias que interactúan cubren un cromosoma completo o el genoma.

Adecuadamente, la enzima de restricción primaria es una enzima de restricción que reconoce un sitio de reconocimiento de 6-8 pb.

20 Adecuadamente, la enzima de restricción primaria se selecciona del grupo que consiste en *BglII*, *HindIII*, *EcoRI*, *BamHI*, *SpeI*, *PstI* y *NdeI*.

Adecuadamente, la enzima de restricción primaria se selecciona basándose en su ausencia de, o bajo representación en secuencias repetitivas.

25 Adecuadamente, la enzima de restricción secundaria es una enzima de restricción que reconoce un sitio de reconocimiento de secuencias de nucleótidos de 4 o 5 pb.

30 Adecuadamente, el sitio de reconocimiento de la enzima de restricción secundaria se localiza a más de aproximadamente 350 pb del sitio de restricción primario en la secuencia de nucleótidos diana.

Adecuadamente, una transición desde frecuencias de interacción bajas a altas es indicativa de la localización de un reordenamiento genético equilibrado y/o no equilibrado.

35 Adecuadamente, un patrón invertido de frecuencias de interacción ADN-ADN para la muestra del sujeto en comparación con el control, es indicativo de una inversión equilibrada y/o no equilibrada.

Adecuadamente, una reducción en la frecuencia de interacción ADN-ADN para la muestra del sujeto en comparación con el control, en combinación con un aumento en la frecuencia de interacción ADN-ADN para regiones más distantes, es indicativa de una deleción equilibrada y/o no equilibrada.

40 Adecuadamente, un aumento o una disminución en la frecuencia de interacción ADN-ADN para la muestra del sujeto en comparación con el control es indicativa de una duplicación o inserción equilibrada y/o desequilibrada.

45 Adecuadamente, el cariotipado espectral y/o FISH se utiliza antes de realizar dicho método.

Adecuadamente, la enfermedad es una enfermedad genética.

Adecuadamente, la enfermedad es cáncer.

50 Adecuadamente, secuencias de nucleótidos que interactúan con dos o más secuencias diana se amplifican.

Adecuadamente, las secuencias diana se posicionan en el locus genómico, o cerca del mismo, que se sabe que está asociado a una patología.

55 Adecuadamente, las secuencias diana se seleccionan sin conocimiento previo sobre la localización de un reordenamiento y están separadas de tal manera que las secuencias que interactúan un cromosoma completo o el genoma, y en el que las secuencias de interacción identificadas permiten la reconstrucción de mapas de cromosomas lineales y reordenamientos genómicos que se producen en los cromosomas y entre los mismos.

60 Adecuadamente, las secuencias amplificadas están marcadas.

Adecuadamente, las secuencias amplificadas están diferencialmente marcadas de acuerdo con su posición en el genoma.

65

Adecuadamente, el método es para la detección de un reordenamiento, una traslocación, una inversión, una delección, una duplicación o una inserción, equilibrado y/o desequilibrado.

Adecuadamente, la etapa de hibridación con la matriz se reemplaza por una etapa de secuenciación.

Adecuadamente, tanto la secuencia de nucleótidos diana como la secuencia de nucleótidos de interés se identifican por secuenciación.

Adecuadamente, las secuencias adaptadoras se ligan con los productos de la PCR.

Adecuadamente, cada una de las secuencias que interactúan con dos o más secuencias diana se amplifican en reacciones de PCR distintas.

Adecuadamente, cada una de las secuencias que interactúan con dos o más secuencias diana se amplifican en reacciones de PCR distintas y posteriormente se agrupan para análisis simultáneo.

Adecuadamente, las secuencias que interactúan con dos o más secuencias diana se amplifican por PCR múltiple.

### VENTAJAS

La presente invención tiene diversas ventajas. Estas ventajas serán obvias a partir de la siguiente descripción.

Como ejemplo, la tecnología 4C puede multiplexarse, de tal manera que, en un solo experimento, por ejemplo, en una sola matriz, puedan analizarse interacciones con dos o más secuencias diana.

Como ejemplo adicional, la tecnología 4C multiplexada puede utilizarse para explorar reordenamientos en el ADN genómico en todo el genoma, en posiciones desconocidas.

Como ejemplo adicional, para analizar los fragmentos de ADN capturados, puede utilizarse secuenciación a alto rendimiento en lugar de micromatrices. Las mejoras del multiplexado y de la secuenciación pueden incluso combinarse.

Como ejemplo adicional, en lugar de multiplexarse, las secuencias de interés capturadas por diferentes secuencias diana pueden amplificarse para cada secuencia diana por separado y después posteriormente agruparse para analizarse simultáneamente en una micromatriz.

Como ejemplo adicional, en lugar de multiplexarse, las secuencias de interés capturadas por diferentes secuencias diana pueden amplificarse por separado para cada secuencia diana y posteriormente agruparse para analizarse simultáneamente por secuenciación a alto rendimiento.

Como ejemplo, la presente invención es ventajosa ya que proporciona, entre otras cosas, secuencias de nucleótidos, procesos, sondas y matrices, útiles desde el punto de vista comercial.

Como ejemplo adicional, la presente invención es ventajosa ya que proporciona análisis de alto rendimiento de la frecuencia de interacción de dos o más secuencias de nucleótidos en el espacio nuclear.

Como ejemplo adicional, la presente invención es ventajosa ya que utiliza tecnología 3C convencional, cada interacción ADN-ADN simple debe analizarse mediante una sola reacción de PCR que contiene un solo par de cebadores. Por lo tanto, los análisis de alto rendimiento son solo posibles si la PCR se automatiza, aunque los costes de tantos cebadores serán demasiados altos. Por consiguiente, el análisis de alto rendimiento (hologenómico) de interacciones ADN-ADN no es viable con tecnología 3C convencional. En cambio, la presente invención permite ahora la exploración simultánea de miles de interacciones ADN-ADN. El análisis de alto rendimiento de interacciones ADN-ADN de acuerdo con la presente invención aumentará enormemente la escala y la resolución de los análisis.

Como ejemplo adicional, la presente invención es ventajosa ya que utiliza tecnología 3C convencional, la exploración está sesgada hacia aquellas secuencias de ADN para las que se diseñaron, encargaron e incluyeron cebadores oligonucleotídicos en el análisis.

La elección de dichos cebadores oligonucleotídicos se basa típicamente en el conocimiento con respecto a la posición de, por ejemplo, potenciadores (distantes) y/u otros elementos reguladores/sitios hipersensibles que se cree que se entrecruzarán con la secuencia de nucleótidos que se está investigando. Por tanto, la tecnología 3C convencional está sesgada hacia el diseño de cebadores de PCR que están incluidos en la etapa de amplificación de PCR, mientras que la tecnología 4C es imparcial y puede utilizarse para investigar el genoma completo para la interacción de elementos de ADN. Esto se debe a que la amplificación de secuencias entrecruzadas en 4C no se basa en el conocimiento predicho de secuencias que se entrecruzan con la secuencia de nucleótidos que se



investiga. Más bien, en una realización de 4C, las secuencias que se entrecruzan con la primera secuencia de nucleótidos (diana) pueden amplificarse utilizando cebadores de PCR que hibridan con esa secuencia de nucleótidos. Por lo tanto, la presente invención permite una exploración hologenómica imparcial para interacciones ADN-ADN.

Como ejemplo adicional, la presente invención es ventajosa ya que utiliza tecnología 3C convencional que solo permite la amplificación selectiva de una interacción ADN-ADN sencilla. Esto no es informativo cuando se hibrida con una matriz. La tecnología se ha mejorado de tal manera que todos los fragmentos que interaccionan con una primera secuencia de nucleótidos (diana) están ahora amplificados, por ejemplo, amplificados de manera selectiva.

Como ejemplo adicional, la presente invención es ventajosa ya que la tecnología 4C se puede utilizar para detectar aberraciones genéticas equilibradas o no equilibradas - tales como todos los tipos de translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos - en el ácido nucleico, por ejemplo, en los cromosomas. La tecnología 4C (que mide proximidad de fragmentos de ADN) puede incluso determinar una predisposición del sujeto para adquirir determinadas translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos (por ejemplo translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos, equilibrados o no equilibrados). Una ventaja sobre las estrategias actuales es que no es preciso conocer la posición exacta del cambio ya que la resolución de la tecnología 4C es tal que esta puede utilizarse para detectar reordenamientos incluso cuando el "cebo 4C" (como se define por los sitios de reconocimiento de enzima de restricción primaria y secundaria que se analizan) se localiza lejos (por ejemplo hasta una megabase o incluso más) del cambio. Otra ventaja sobre las estrategias actuales es que permite realizar una investigación hologenómica simultánea, imparcial, de reordenamientos genómicos tanto equilibrados como no equilibrados. Otra ventaja es que la tecnología 4C permite realizar un mapeo preciso de cambios dado que puede utilizarse para definir los dos sitios de restricción (primarios) entre los que se producen cambios. Otra ventaja es que las células no requieren cultivarse antes de fijarse. Por tanto, también pueden analizarse, por ejemplo, tumores sólidos con respecto a reordenamientos genómicos.

Como ejemplo adicional, la presente invención es ventajosa porque la tecnología 4C también puede detectar cambios (por ejemplo, reordenamientos) en un estado pre-maligno, es decir, antes de que todas las células contengan estos cambios. Por tanto, la tecnología puede utilizarse no solo en el diagnóstico de la enfermedad sino también en el pronóstico de la misma.

Como ejemplo adicional, el diseño de matrices de acuerdo con la presente invención es particularmente ventajoso en comparación con el de las matrices en mosaico (*tiling*) genómicas existentes - tales como las matrices en mosaico genómicas de Nimblegen - dado que el diseño permite la representación de una parte mucho más grande del genoma en una sola matriz. Como ejemplo, para una enzima de restricción que reconoce una secuencia hexanucleotídica, aproximadamente 3 matrices con aproximadamente 385.000 sondas cada una de ellas, será suficiente para cubrir, por ejemplo, el genoma completo de ser humano o de ratón. Para una enzima de restricción que reconoce más de 6 pb, puede utilizarse una sola matriz de aproximadamente 385.000 sondas para cubrir, por ejemplo, el genoma completo de ser humano o de ratón. Las ventajas del diseño de la matriz son que: (1) cada sonda es informativa dado que cada una de ellas analiza un evento de ligamiento independiente, lo que facilita enormemente la interpretación de los resultados; y (2) puede detectarse una gran representación del genoma en una sola matriz, lo cual es rentable.

La tecnología 4C puede utilizarse ventajosamente para el mapeo refinado de reordenamientos mal caracterizados originalmente detectados por estrategias citogenéticas (microscopía óptica, FISH, SKY, etc.).

La tecnología 4C puede utilizarse ventajosamente para la exploración simultánea en una sola matriz para combinaciones de reordenamientos que se han producido cerca de locus múltiples.

#### Breve descripción de las figuras

##### *Figura 1*

El principio de la tecnología 3C

##### *Figura 2*

(a) El principio de una realización de tecnología 4C. El análisis 3C se realiza de manera habitual, por ejemplo, con *HindIII* (H) como una enzima de restricción. Después de invertir los entrecruzamientos, la mezcla de ADN contendrá una primera secuencia de nucleótidos (diana) ligada a muchos fragmentos diferentes. Estos fragmentos se amplificarán y marcarán utilizando métodos de amplificación - tales como PCR inversa - sobre, por ejemplo, círculos de *DpnII*, utilizando primero cebadores específicos de secuencias de nucleótidos (diana). Los productos de amplificación marcados pueden hibridarse con las matrices, como se describe en el presente documento. *HindIII* y *DpnII* se proporcionan como ejemplos, pero también pueden utilizarse otras combinaciones de enzimas de restricción, tales como enzimas cortadoras de 6 u 8 y 4 o 5 pb. (b) Resultados de la PCR separados por electroforesis en gel de dos muestras independientes de hígado (*Liver, L*) (L1, L2) y de cerebro (*Brain, B*) (B1, B2) fetal. (c) Representación esquemática de la localización de las sondas de la micromatriz. Se

diseñaron sondas dentro de 100 pb de sitios *HindIII*. Por tanto, cada sonda analiza un posible compañero de ligamiento.

**Figura 3**

La tecnología 4C detecta el entorno genómico de Rad23A (cromosoma 8). Se muestran proporciones no procesadas (señales 4C para Rad23A divididas entre la señal obtenida para la muestra de control) para sondas localizadas en ~ 15 Mb o más regiones cromosómicas en el cromosoma 10, 11, 12, 14, 15, 7 y 8 de ratón (de arriba a abajo); se muestran regiones que están a la misma distancia de cada centrómero correspondiente). Obsérvese el grupo grande de señales fuertes alrededor del cebo (Rad23A) en el cromosoma 8 (fila 7), lo que demuestra que la tecnología 4C detecta fragmentos genómicos cercanos al molde cromosómico lineal (acorde con el hecho de que las frecuencias de interacción son inversamente proporcionales a la separación de sitios genómicos). Obsérvese que la región unida en *cis* alrededor del cebo que muestra intensidades de alta señal es grande (> 5 Mb), lo que implica que, por ejemplo, pueden detectarse translocaciones incluso con cebos a una distancia mayor de 1 MB del punto de corte.

**Figura 4**

Interacciones 4C de la  $\beta$  globina en el cromosoma 7 (~ 135 Mb) para un tejido en transcripción (hígado fetal) y un tejido no en transcripción (cerebro fetal) (analizado mediante una estrategia de media móvil (*running mean*)). Obsérvese que las interacciones de largo alcance con la  $\beta$  globina difieren entre tejidos (probablemente dependiendo del estado de transcripción del gen). Independientemente del tejido, las señales 4C fuertes delimitan una región grande (> 5 Mb) alrededor del cebo.

**Figura 5**

*Uros* y *Eraf* interactúan con la  $\beta$  globina en células del hígado fetal. La estrategia 4C revela que dos genes, *Eraf* y *Uros*, interactúan sobre > 30 Mb con el locus de la  $\beta$  globina localizado a una distancia de ~ 30 Mb. Estas dos interacciones se descubrieron previamente a través de una tecnología diferente (Hibridación In Situ con Fluorescencia, FISH) como describen Osborne *et al.*, en Nature Genetics 36, 1065 (2004). Este ejemplo muestra que las interacciones de largo alcance detectadas por la tecnología 4C pueden verificarse por FISH y que reflejan realmente proximidad nuclear.

**Figura 6**

La tecnología 4C identifica con precisión transiciones entre regiones genómicas no relacionadas que están unidas en *cis*. Para estos experimentos se utilizaron ratones transgénicos que contenían un casete de la región de control del locus (LCR, *Locus Control Region*) de la  $\beta$  globina (~ 20 kb) insertado (por recombinación homóloga) en el locus Rad23A en el cromosoma 8 de ratón. La tecnología 4C se realizó en hígados fetales E14.5 de ratones transgénicos que eran homocigotos para esta inserción. Como "cebo 4C", se utilizó un fragmento de *HindIII* dentro del casete de integración (HS2). Los datos muestran que la tecnología 4C define con precisión ambos extremos del casete transgénico (fila inferior: solamente sondas en la LCR humana (~ 20 kb) proporciona señales 4C y sin sondas en el resto de la secuencia de la  $\beta$  globina humana de ~ 380 kb) y claramente revela la posición de integración en el cromosoma 8 de ratón (panel superior: se comparan señales en el cromosoma 8 (para la posición de integración, véase la fila) con señales en 6 cromosomas distintos de ratón) (se representan cromosomas completos). Este experimento demuestra que la tecnología 4C puede utilizarse para detectar la posición genómica de fragmentos de ADN ectópicamente integrados (virus, transgenes, etc.). Muestra que las transiciones entre regiones genómicas no relacionadas que están unidas en *cis* pueden identificarse con precisión, que pueden utilizarse para identificar puntos de corte genómicos y compañeros de translocación.

**Figura 7**

La tecnología 4C produce datos reproducibles dado que los perfiles de HS2 y de la  $\beta$  globina son muy similares. Se realizaron cuatro experimentos 4C biológicamente independientes en hígados fetales E14.5, utilizando bien el gen  $\beta$ -mayor de la  $\beta$  globina (2 filas superiores) o el H2S de la  $\beta$  globina (dos filas inferiores) como cebo. Estos cebos estaban a una distancia de ~ 40 kb en el molde cromosómico lineal aunque previamente se demostró que estaban cerca del espacio nuclear (Tolhuis *et al.*, Molecular Cell 10, 1453 (2002)) Se representa una región de ~ 5 Mb en el cromosoma 7 de ratón que está a una distancia de 20-20 Mb del locus de la  $\beta$  globina. Los datos muestran una alta reproducibilidad entre experimentos independientes y demuestran que dos fragmentos cercanos en el espacio nuclear comparten compañeros de interacción localizados en cualquier parte en el genoma.

**Figura 8**

La tecnología 4C se aplica para medir frecuencias de interacción ADN-ADN con secuencia X (en el cromosoma A) en células de una persona sana (parte superior) y de un paciente con translocación (A; B) (parte inferior). Las intensidades de señal que representan las frecuencias de interacción ADN-ADN (eje Y) se representan gráficamente para sondas ordenadas en moldes de cromosoma lineal (eje X). En células normales, las interacciones frecuentes ADN-ADN se detectan en el cromosoma A alrededor de la secuencia X. En las células de pacientes, se observa una reducción en cuanto a las frecuencias de interacción del 50 % para sondas en el cromosoma A localizadas en el otro lado del punto de corte (PR) (compárese la curva gris (paciente) con línea negra (persona sana)). Además, la translocación trae parte del cromosoma B en estrecha proximidad física a la secuencia X, y ahora se observan interacciones ADN-ADN frecuentes para esta región en el cromosoma B. La transición brusca de baja a alta frecuencia de interacción en este cromosoma marca la localización de este punto de corte.

*Figura 9*

A través de la tecnología 4C pueden detectarse una o más inversiones (equilibradas). Se observan patrones inversos de frecuencias de interacción ADN-ADN (medidas por tecnología 4C como intensidades de señales de hibridación) en sujetos enfermos (curva continua) en comparación con sujetos no enfermos (curva discontinua), que revela la presencia y el tamaño de la inversión.

*Figura 10*

Detección de una o más deleciones heterocigotas por tecnología 4C. Las sondas con frecuencias de interacción ADN-ADN reducidas (medidas por tecnología 4C como intensidades de señales de hibridación) en sujetos enfermos (curva gris) en comparación con sujetos no enfermos (curva negra), revelan la posición y el tamaño de la región delecionada. Las señales de hibridación residuales en la región delecionada del sujeto enfermo proceden de un alelo intacto (deleción heterocigota). La deleción viene típicamente acompañada de un aumento de intensidades de señal para sondas localizadas directamente más allá de la región delecionada (obsérvese que la curva gris está por encima de la curva negra a la derecha de la deleción), ya que estas regiones se acercan más estrechamente a la secuencia 4C (cebo).

*Figura 11*

Duplicación detectada por la tecnología 4C. Las sondas con aumento de las señales de hibridación en un paciente (curva continua) en comparación con un sujeto normal (curva discontinua) indican la posición y el tamaño de la duplicación. La duplicación, detectada por tecnología 4C, viene típicamente acompañada por una disminución de las señales de hibridación en sujetos enfermos en comparación con los no enfermos para sondas más allá de la región duplicada (la duplicación aumenta su separación en el sitio genómico desde la secuencia 4C).

*Figura 12*

Interacciones de largo alcance con la  $\beta$  globina reveladas por tecnología 4C. **a**, Proporciones no procesadas de 4C sobre señales de hibridación control, revelan interacciones de HS2 de la  $\beta$  globina con el cromosoma 7 y dos cromosomas no relacionados (8 y 14). **b-c**, Datos no procesados de dos muestras independientes de hígado fetal (parte superior, en rojo) y de cerebro fetal (parte inferior, en azul) representadas gráficamente a lo largo de dos regiones 1-2 Mb diferentes en el cromosoma 7. Se observan grupos de interacciones muy reproducibles bien en las dos muestras de hígado fetal (b) o en las dos muestras de cerebro (c). **d-e**, Datos de media móvil para las mismas regiones. La proporción de descubrimientos falsos se estableció al 5% (línea discontinua). **f**, Representación esquemática de regiones de interacción con  $\beta$  globina en el cromosoma 7 activa (hígado fetal, parte superior) e inactiva (cerebro fetal, parte inferior).

*Figura 13*

La  $\beta$ -globina activa e inactiva interacciona con regiones cromosómicas activas e inactivas, respectivamente. **a**, Comparación entre interacciones de largo alcance en hígado fetal (media móvil 4C, parte superior), análisis de expresión en micromatriz en hígado fetal (escala logarítmica, centro) y la localización de genes (parte inferior) representada a lo largo de una región de 4 Mb que contiene el gen Uros (a una distancia de ~ 30 Mb de la  $\beta$  globina), mostrando que la  $\beta$  globina activa interacciona preferentemente con otros genes transcritos de manera activa. **b**, La misma comparación en cerebro fetal alrededor de un grupo de genes OR localizado a una distancia de ~ 38 Mb de la globina, mostrando que la  $\beta$  globina inactiva interacciona preferencialmente con regiones inactivas. **c**, Caracterización de regiones que interaccionan con la  $\beta$  globina en hígado fetal (izquierda) y cerebro fetal (derecha) en cuanto a contenido y actividad génicos.

*Figura 14*

Rad23A expresado de manera ubicua interacciona con regiones activas, muy similares, en hígado y cerebro fetal. **a**, Representación esquemática de regiones en el cromosoma 8 que interaccionan con Rad23A activo en hígado fetal (parte superior, color rojo) y cerebro fetal (parte inferior, color azul). Comparación entre interacciones de largo alcance de Rad23A (media móvil 4C) y análisis de expresión de micromatriz (escala logarítmica) en hígado fetal (dos paneles superiores), interacciones de largo alcance de Rad23A (media móvil 4C) y análisis de expresión de micromatriz (escala logarítmica) en cerebro fetal (paneles 3 y 4) y la localización de genes (panel inferior) representada gráficamente a lo largo de una región de 3 Mb del cromosoma 8. **c**, Caracterización de regiones que interaccionan con Rad23A en hígado fetal (izquierda) y cerebro (derecha) en cuanto a contenido y actividad génicos.

*Figura 15*

La Crio-FISH (*hibridación in situ con fluorescencia (FISH) en criosecciones finas*) confirma que la tecnología 4C identifica realmente regiones de interacción. **a**, ejemplo de parte de una criosección (200 nm) que muestra más de 10 núcleos, algunos de los cuales contienen el locus de la  $\beta$  globina (en verde) y/o Uros (en rojo). Debido al seccionamiento, muchos núcleos no contienen señales para estos dos locus. **b-d**, ejemplos de señales completamente (b) y parcialmente (c) solapantes y señales de contacto (d), todas ellas puntuadas como positivas para la interacción. **e-g**, ejemplos de núcleos que contienen alelos que no están en contacto (e-f) y de un núcleo que solo contiene  $\beta$  globina (g), todos ellos puntuados como negativos para la interacción. **h-i**, representación esquemática de resultados crio-FISH. Se indican porcentajes de interacción con  $\beta$  globina (h) y Rad23A (i) encima de cromosomas para regiones identificadas positivamente (cabeza de flecha roja) e identificadas negativamente (cabeza de flecha azul) por tecnología 4C. Para los dos tejidos se utilizaron los mismos BAC. Las frecuencias de interacción medidas por crio-FISH entre dos grupos de genes OR distantes en hígado y cerebro fetal se indican debajo de los cromosomas.

*Figura 16*

Los análisis 4C de HS2 y  $\beta$ -major dan resultados muy similares. (a) Datos 4C no procesados de cuatro muestras independientes de hígado E14.5 muestran un patrón de interacción muy similar con HS2 (parte superior) y  $\beta$ -major (parte inferior). (b) Existe un solapamiento grande entre sondas con puntuación positiva para la interacción en el experimento con HS -2 y sondas con puntuación positiva para la interacción en el experimento con  $\beta$ -major.

*Figura 17*

Las regiones que interaccionan con la  $\beta$  globina, frecuentemente también se ponen en contacto entre sí. Dos regiones (a una distancia de casi 60 Mb), que contenían genes transcritos de manera activa e identificados por tecnología 4C para interaccionar con la  $\beta$  globina en hígado fetal, mostraron frecuencias de colocalización por crio-FISH del 5,5 %, que era significativamente mayor que las frecuencias de colocalización de fondo.

*Figura 18*

Ejemplo de una deleción heterocigota revelada por 4C múltiple utilizando un solo colorante para el marcaje de fragmentos de ADN que interaccionan con secuencias diana múltiples. La proporción de las frecuencias de interacción observadas en un paciente (muestra) en relación con una persona sana (control) se representa a la derecha.

*Figura 19*

La presencia de una deleción presente en un paciente con leucemia, como se revela por 4C, utilizando una secuencia de nucleótidos diana que está a 2 Mb (A) o a 1,3 Mb (B) cadena arriba ("a la izquierda") del primer punto de corte. Obsérvese que las deleciones causan una reducción de las señales de interacción de ADN en la región delecionada, pero también causan un aumento en las frecuencias de interacción ADN:ADN para secuencias directamente cadena abajo ("a la derecha") del último punto de corte. Esto es particularmente obvio cuando se examinan estrechamente interacciones con una secuencia B de nucleótidos diana (véanse los dos gráficos de inferiores). Basándose en datos 4C, se diseñaron cebadores a cada lado de la región delecionada y se identificó un punto de corte por secuenciación: el texto sin formato es la secuencia de deleción cadena arriba, en negrita se indica un nucleótido insertado, en subrayado se indica la secuencia de la deleción cadena abajo.

*Figura 20*

Una inversión heterocigota revelada por 4C múltiple utilizando un solo colorante para el marcaje de fragmentos de ADN que interaccionan con secuencias diana múltiples. La proporción de las frecuencias de interacción observadas en un paciente (muestra) en relación con una persona sana (control) se representa a la izquierda. Obsérvese que las proporciones cercanas a los puntos de corte pueden ser diferentes cuando la posición de los puntos de corte con respecto a las secuencias de nucleótidos diana es diferente.

*Figura 21*

Los colores alternan entre secuencias de nucleótidos diana vecinas, lo que permite la detección (en rojo) de una deleción próxima a una secuencia de nucleótidos diana (en azul) que no detecta la deleción debida a señales de hibridación saturadas. En el caso de que la cantidad de sonda en la matriz no esté saturando, la señal azul también disminuiría en la deleción.

*Figura 22*

Los colores alternan entre secuencias de nucleótidos diana vecinas, lo que permite la detección (en rojo) de una inversión. La proporción de frecuencias de interacción observada en el paciente (muestra) en relación con una persona sana (control) se representa a la izquierda. Obsérvese que comparado con un solo experimento de tinción (véase la Figura 2) el uso de colorantes alternos facilitó la detección de reordenamientos tales como inversiones. Obsérvese también que las proporciones próximas a los puntos de corte pueden ser diferentes cuando la posición de los puntos de corte con respecto a las secuencias de nucleótidos diana es diferente.

*Figura 23*

Ejemplo de una inversión heterocigota, revelada por 4C múltiple, utilizando diferentes colorantes para el marcaje de fragmentos de ADN que interaccionan con diferentes secuencias diana. Se indican los puntos de corte de la inversión por la posición de señales roja y verde en la muestra del paciente que no están presentes en la muestra de control. Obsérvese que la introducción de más colores facilita la detección de reordenamientos (comparar, por ejemplo, las Figuras 4 y 5).

*Figura 24*

Detección de translocaciones equilibradas. Cada cromosoma está marcado con dos colorantes únicos que se utilizan de tal manera que los colorantes alternan entre secuencias diana colindantes en el molde de cromosoma lineal. Si las translocaciones están equilibradas, cada uno de los dos colorantes específicos del cromosoma deben dar señales de hibridación fuertes en un conjunto de sondas mutuamente exclusivo directamente colindantes entre sí en el molde lineal del cromosoma no relacionado. El punto de corte en este cromosoma no relacionado se localiza entre los dos conjuntos de sondas que muestran señales de hibridación positiva. Las señales cromosómicas originales en y más allá del punto de corte serían la mitad de la intensidad del control (no mostrado en la figura).

*Figura 25*

Estudio de viabilidad (*Proof of principle*) para la detección de translocaciones equilibradas. La detección de la translocación t(1;7) como se describe en (R. Burnett *et al.*, Blood, Vol 84, No 4 (15 de agosto), 1994: págs. 1232-1236). Las secuencias de nucleótidos diana flanquean el locus de TCR $\beta$  en el cromosoma 7, representando las señales en rojo interacciones ADN:ADN con la secuencia diana que se localiza cadena arriba del locus de TCR $\beta$ , y representando las señales en azul interacciones ADN:ADN con la secuencia diana que se localiza cadena abajo del locus de TCR $\beta$ . Se ilustran las señales de interacción de ADN encontradas en el cromosoma 1. El panel superior muestra la distribución teórica de señales. Los paneles central e inferior, muestran la distribución

real de las señales. El panel inferior muestra señales a una resolución de sondas individuales yuxtapuestas en el molde del cromosoma. Obsérvese que, en el caso de una translocación equilibrada, las secuencias de nucleótidos diana que flanquean el punto de corte mostrarán un conjunto mutuamente exclusivo de señales de interacción de ADN intercromosómico que directamente limitan entre sí el molde de cromosoma lineal del cromosoma de translocación compañero. La posición del punto de corte secuenciado (descrito en Burnett *et al.*, 1994) se indica con una flecha en el panel inferior.

**Figura 26**

Ejemplo teórico de la detección de translocaciones no equilibradas. Cada cromosoma está marcado con dos colorantes exclusivos que se utilizan de tal manera que los colorantes alternan entre secuencias diana que son vecinas en el molde de cromosoma lineal. Si las translocaciones se produjeron con pérdida de ADN en los puntos de corte (es decir translocaciones no equilibradas), cada uno de los dos colorantes específicos de cromosoma dará señales fuertes de hibridación en un conjunto de sondas mutuamente exclusivo en el cromosoma no relacionado que no son directamente vecinas entre sí en el molde lineal del cromosoma no relacionado. Se indica la región delecionada.

**Figura 27**

Detección de translocaciones no equilibradas. Detección de la translocación t(4;7) como se describe en (RJ Galjaard *et al.*, Am J Med Genet A. 30 de agosto de 2003; 121(2): 168-73). Secuencias de nucleótidos diana localizadas en el cromosoma 7; las señales de interacción de ADN representadas se localizan en el cromosoma 4. Se utilizaron dos secuencias diana localizadas cadena arriba (5') y cadena abajo (3') del punto de corte del cromosoma 7. Se indican las señales de interacción de ADN localizadas en el cromosoma 4 (en azul para ambas secuencias diana). La región entre los grupos de fragmentos de interacción de ADN en el cromosoma 4 se ha delecionado en este paciente. Parte superior: señales del cromosoma 4 completo. Panel inferior, datos 4C: señales en una región de 11,5 MB alrededor de los puntos de corte en el cromosoma 4. Basándose en estos datos 4C, se identificó el fragmento de restricción *HindIII* en el cromosoma 4 que contenía el punto de corte de translocación y se utilizó para mapear el punto de corte por secuenciación. La secuencia se proporciona en la parte inferior de la figura, donde la secuencia subrayada es del cromosoma 4, la negrita se encuentra tanto en 7 como en 4 y la secuencia sin formato es del cromosoma 7.

**Figura 28**

Marcaje específico de cromosoma de interacciones de ADN. La señal azul que aparece en el cromosoma 3 y las señales naranjas que aparecen en el cromosoma 1 revelan la translocación de cromosomas compañeros y la posición aproximada de los puntos de corte.

**Figura 29**

Secuenciación de secuencias de nucleótidos amplificadas por PCR de interés (azul: dirección de viaje) ligadas a una secuencia de nucleótidos diana (rojo: dirección de domicilio). La amplificación se realizó utilizando cebadores (rojos), siendo al menos uno de los cuatro complementario a la secuencia de nucleótidos diana. Opcionalmente, pueden introducirse adaptámeros (gris) de diversas maneras en el extremo de los productos de la PCR, según se indica

**Figura 30**

4C detecta con precisión una translocación e inversión equilibradas (A-B). La tecnología 4C detecta una translocación equilibrada t(1;7). (A) En una muestra de control sana, los fragmentos diana a (rojo) y b (azul) que se localizan en sitios opuestos del locus *TCRB* en el cromosoma 7 no capturan regiones en el cromosoma 1. (B) En la línea celular HSB-2 que contiene una translocación equilibrada t(1;7)(p35; q35), cada fragmento diana *TCRB* captura una región en el cromosoma 1. Las regiones capturadas tienen diversas megabases de tamaño (aumento 1), son directamente adyacentes entre sí (aumento 2) y flanquean el punto de corte clonado anteriormente (flecha). Véase la Fig. S1 para los resultados de otros cromosomas. (C-D) 4C detecta una inversión equilibrada. (C) En una muestra de control sana, los fragmentos diana a (rojo) y b (azul) que están localizados en sitios opuestos del locus *TCRB* en el cromosoma 7 no capturan grandes regiones en ninguna parte del cromosoma 7. (D) En una muestra de un paciente T-ALL, cada fragmento diana *TCRB* captura una región adicional en el otro extremo del cromosoma 7. El fragmento diana más cercano a 5' (a; rojo) captura una región 5' del grupo *HOXA*, el fragmento diana más cercano a 3' (b; azul) captura una región 3' del grupo *HOXA*, demostrando una inversión. Las regiones capturadas son de diversas megabases de tamaño (aumento 1) y son directamente adyacentes entre sí (aumento 2), lo que muestra que la inversión está equilibrada. Ambos fragmentos diana identifican el punto de corte (flecha) cerca de *HOXA9* dentro de una región de 6 kb. Los datos de media móvil se representaron, utilizando un tamaño de ventana de ~60 kb. Los aumentos muestran intensidades de señal no procesadas.

**Figura 31**

4C detecta con precisión reordenamientos no equilibrados. (A) 4C detecta con precisión una translocación t(4;7) (p15.2; q35) en combinación con una microdeleción (es decir reordenamiento no equilibrado) en una línea celular de un recién nacido con malformación congénita. Los dos fragmentos diana a (rojo) y b (azul) que se localizaron en lados opuestos de los puntos de corte en el cromosoma 7 capturan fragmentos que abarcan diversas megabases en el cromosoma 4 (para señales en otros cromosomas, véase la Figura 35). (B) Las dos regiones capturadas con altas señales no son directamente adyacentes entre sí, mostrando que la translocación está acompañada por una deleción del cromosoma 4. (B) La secuencia de uno de los puntos de corte (flecha), con secuencias del cromosoma 7 y 4 en letras pequeñas y mayúsculas, respectivamente. (C) 4C identifica con precisión una deleción homocigota en una muestra de un paciente T-ALL. Un fragmento diana localizado en el cromosoma 9 en 19,3 Mb identifica una región (entre flechas) que carece de señales altas en la muestra del

paciente (inferior) en comparación con una muestra de control (superior) que muestra una delección de ~2 Mb en 9p21. Señales 3' de la delección son superiores en pacientes frente a control, ya que esta región es más cercana en proximidad al fragmento diana debido a la delección. **(D)** La secuencia a través de los puntos de corte indicada por las flechas en **(C)**, que confirma la delección. En esta figura se representan intensidades de señales no procesadas. Las señales altas poco frecuentes en las regiones delecionadas indican que estas sondas muestran hibridación no específica.

**Figura 32**

La exploración con 4C identifica *LMO3* como un nuevo compañero de translocación de *TCRB*. Cinco muestras de paciente T-ALL no caracterizadas se exploraron con 4C, utilizando un fragmento diana cerca del locus *TCRB* en el cromosoma 7. **(A)** En una muestra del paciente apareció una alta señal específicamente en el cromosoma 12, revelando una translocación t(7;12)(q35;p12.3). Para señales en otros cromosomas, véase la figura S4. Una delección está presente diversas megabases desde el sitio de translocación (flecha) en el cromosoma 12 (aumento 1). El sitio de translocación está presente en una región de 6 kb cerca del gen *LMO3* (aumento 2). **(B)** Secuencias de ambos puntos de corte de t(7;12)(q35; p12.3); los nucleótidos en letra mayúscula son de 12, en letra minúscula de 7 y en cursiva de origen desconocido. **(C)** Representación esquemática del sitio de translocación de t(7;12)(q35;p12.3). El potenciador de *TCRB* se posiciona 70 kb cadena abajo del gen *LMO3*. Se representaron datos de media móvil, utilizando un tamaño de ventana de ~60 kb. Los aumentos muestran intensidades de señal no procesadas.

**Figura 33**

4C señala a través de todos los cromosomas en un control sano y en una muestra que lleva t(1;7)(p35;q35). Las cabezas de flecha negras indican la posición de las secuencias diana. Las cabezas de flecha rojas indican la posición del sitio de translocación. Se representaron gráficamente datos de media móvil, utilizando un tamaño de ventana de ~60 kb. La escala en el eje Y (unidades arbitrarias) es idéntica en todos los cromosomas.

**Figura 34**

Secuenciación de extremo emparejado con fragmento de restricción **(A)** Representación esquemática de secuenciación de extremos emparejados de fragmentos de restricción. **(B)** Secuencias de punto de corte de una inversión entre *TCRB* y *HOXA* en el cromosoma 7 (véase la figura 30). La secuencia en negrita se localiza en la posición en *HOXA* encontrada con tecnología 4C entre las sondas que marcan la transición de fragmentos capturados a no capturados. La secuencia roja se encontró con secuenciación de extremo emparejado de fragmento de restricción de la secuencia en negrita y se localiza en el locus *TCRB*.

**Figura 35**

4C señala a través de todos los cromosomas obtenidos con dos fragmentos diana de 7 cromosomas diferentes en una muestra que lleva t(4; 7)(p15.2; q35). Las cabezas de flecha negras indican la posición de las secuencias diana. Las cabezas de flecha rojas indican la posición de los sitios de translocación. Se representaron gráficamente datos de media móvil, utilizando un tamaño de ventana de ~60 kb. La escala en el eje Y (unidades arbitrarias) es idéntica en todos los cromosomas.

**Figura 36**

4C señala a través de todos los cromosomas obtenidos con una secuencia diana cerca del locus *TCRB* en el cromosoma 7 en dos muestras de pacientes T-ALL, una de las cuales porta una translocación t(7;12). Las cabezas de flecha negras indican la posición de las secuencias diana. La cabeza de flecha en rojo indica la posición del sitio de translocación. Se representaron gráficamente datos de media móvil utilizando un tamaño de ventana de ~60 kb. La escala en el eje Y (unidades arbitrarias) es idéntica para todos los cromosomas.

**Figura 37**

Expresión de *LMO3* en muestras de pacientes T-ALL. La expresión génica se midió en matrices de expresión génica affymetrix. *LMO3* se expresó en el paciente que llevaba t(7;12)(q35; p12.3), pero no en los otros pacientes.

**Figura 38**

Cebadores de PCR 4C y productos a analizar por secuenciación Solexa. Secuencias producidas por Solexa (flechas) primero leyeron el "cebador *DpnII*" (18 nucleótidos, incluyendo GATC (es decir sitio de reconocimiento *DpnII*), seguido por la secuencia capturada.

**Figura 39**

Resultados de PCR utilizando cebadores con salientes 5' que contienen secuencias adaptadoras Solexa. Para comparación, también se muestran los resultados obtenidos con cebadores convencionales (carriles 1,5,9 de izquierda a derecha; sin salientes).

**Figura 40**

Resultados de secuenciación 4C.

**Figura 41**

Conjunto de cebadores 3 (139 Mb) captura las secuencias de cromosoma 1 a través del punto de corte en la línea celular HSB-2 T-ALL (en el cromosoma 7 a ~142 Mb, es decir a 3 Mb del cebo). Para comparación, se muestran resultados de micromatriz. Obsérvese que la secuencia diana (cebo) utilizada para el experimento en micromatriz estaba más cerca (<1 Mb) del punto de corte, lo que explica por qué mapea el punto de corte en el cromosoma 1 mejor.

Descripción detallada de la invenciónTECNOLOGÍA 3C

5 El método 3C se ha descrito con detalle en Dekker *et al.* (2002), Tolhuis *et al.* (2002), Palstra *et al.* (2003), Splinter *et al.* (2004) y Drissen *et al.* (2004). En resumen, el 3C se realiza por digestión de ADN entrecruzado con una enzima de restricción primaria seguido de ligamiento a concentraciones de ADN muy bajas. En estas condiciones, el ligamiento de fragmentos entrecruzados, que es intramolecular, se favorece enormemente sobre el ligamiento de fragmentos al azar, que es intermolecular. El entrecruzamiento se invierte después y se detectan productos de ligamiento individuales y se cuantifican mediante la reacción en cadena de la polimerasa (PCR) utilizando cebadores específicos de locus. La frecuencia de entrecruzamiento (X) de dos locus específicos se determina por reacciones de PCR cuantitativas utilizando moldes de control y entrecruzados, y X se expresa como la proporción de la cantidad del producto obtenido con el molde entrecruzado y con el molde control.

15 De acuerdo con la presente invención, se prepara un molde 3C utilizando los métodos descritos por Splinter *et al.*, (2004) *Methods Enzymol.* 375, 493-507 (es decir, fijación con formaldehído, digestión con enzimas de restricción (primarias), religamiento de fragmentos de ADN entrecruzados y purificación de ADN). En resumen, una muestra - tal como de células, tejidos o núcleos - se fija utilizando un agente de entrecruzamiento - tal como formaldehído. Después se realiza la digestión con las enzimas de restricción primarias de manera que el ADN se digiera en el contexto del núcleo entrecruzado. Después el ligamiento intramolecular se realiza a concentraciones de ADN bajas (por ejemplo, de aproximadamente 3,7 ng/ $\mu$ l), lo cual favorece el ligamiento entre fragmentos de ADN entrecruzados (es decir ligamiento intramolecular) sobre el ligamiento entre fragmentos de ADN no entrecruzados (es decir ligamiento intermolecular o al azar). A continuación, los entrecruzamientos se invierten y el ADN puede purificarse. El molde 3C que se produce contiene fragmentos de restricción que se ligan debido a que están originalmente cerrados en el espacio nuclear.

Dado que se utiliza una enzima de restricción primaria para digerir el ADN antes de la etapa de ligamiento intramolecular, un sitio de reconocimiento enzimático para la enzima de restricción primaria separará la primera secuencia de nucleótidos (diana) y la secuencia de nucleótidos que se ha ligado. Por consiguiente, el sitio de reconocimiento primario se localiza entre la primera secuencia de nucleótidos (diana) y la secuencia de nucleótidos ligada (es decir, la segunda secuencia ligada).

SECUENCIA DE NUCLEÓTIDOS

35 La presente invención implica el uso de secuencias de nucleótidos (por ejemplo, moldes 3C, moldes 4C, moldes de ADN, moldes de amplificación, fragmentos de ADN y ADN genómico), que pueden estar disponible en bases de datos.

40 La secuencia de nucleótidos puede ser ADN o ARN de origen genómico, sintético o recombinante, por ejemplo ADNc. Por ejemplo, pueden prepararse secuencias de nucleótidos recombinantes utilizando técnicas de clonación con PCR. Esto implicará la preparación de un par de cebadores que flanquean una región de la secuencia que se desea clonar, poner los cebadores en contacto con un ARNm o ADNc obtenido por ejemplo de un mamífero (por ejemplo, una célula animal o de ser humano) o una célula que no es de mamífero, realizar una reacción en cadena de la polimerasa (PCR) en condiciones que lleven a la amplificación de la región deseada, aislar el fragmento amplificado (por ejemplo purificando la mezcla de reacción en un gel de agarosa) y recuperar el ADN amplificado. Los cebadores pueden diseñarse para que contengan sitios de reconocimiento de enzimas de restricción adecuadas de manera que el ADN amplificado pueda clonarse en un vector de clonación adecuado.

50 La secuencia de nucleótidos puede ser mono o bicatenaria bien representando la cadena con sentido o antisentido o combinaciones de las mismas.

Para algunos aspectos, se prefiere que la secuencia de nucleótidos sea ADN monocatenario - tal como cebadores y sondas monocatenarios.

55 Para algunos aspectos, se prefiere que la secuencia de nucleótidos sea ADN bicatenario - tal como moldes 3C y 4C bicatenarios.

Para algunos aspectos, se prefiere que la secuencia de nucleótidos sea ADN genómico - tal como uno o más locus genómicos.

60 Para algunos aspectos, se prefiere que la secuencia de nucleótidos sea ADN cromosómico.

La secuencia de nucleótidos puede comprender una primera secuencia de nucleótidos (diana) y/o una segunda secuencia de nucleótidos.

65

Los sitios de reconocimiento de enzimas de restricción primarias y secundarias diferirán entre sí y típicamente se producirán solo una vez en la secuencia de nucleótidos

5 En un aspecto, se proporciona una secuencia de nucleótidos circularizada que comprende una primera secuencia de nucleótidos y (por ejemplo ligada a) una segunda secuencia de nucleótidos separada (por ejemplo, dividida o apartada) mediante un sitio de reconocimiento de enzimas de restricción primarias y secundarias, en el que dicha primera secuencia de nucleótidos es una secuencia de nucleótidos diana y dicha segunda secuencia de nucleótidos puede obtenerse por entrecruzamiento de ADN genómico (por ejemplo, *in vivo* o *in vitro*). Los sitios de reconocimiento de enzimas de restricción primarias y secundarias serán diferentes entre sí y típicamente se producirán una sola vez en la secuencia de nucleótidos.

15 En un aspecto adicional, se proporciona una secuencia de nucleótidos circularizada que comprende una primera secuencia de nucleótidos y (por ejemplo ligada a) una segunda secuencia de nucleótidos separada (por ejemplo dividida o apartada) por un sitio de reconocimiento de enzimas de restricción primarias y secundarias, en el que dicha primera secuencia de nucleótidos es una secuencia de nucleótidos diana y en el que dicha primera y segunda secuencias de nucleótidos pueden obtenerse mediante un proceso que comprende las etapas de: (a) entrecruzar ADN genómico (por ejemplo *in vivo* o *in vitro*); (b) digerir el ADN entrecruzado con una enzima de restricción primaria; (c) ligar las secuencias de nucleótidos entrecruzadas; (d) invertir el entrecruzamiento y (e) digerir las secuencias de nucleótidos con una enzima de restricción secundaria para circularizar las secuencias de nucleótidos.

20 Preferentemente, la segunda secuencia de nucleótidos corta (por ejemplo biseca) la primera secuencia de nucleótidos (diana). Por consiguiente, la secuencia de nucleótidos comprende la segunda secuencia de nucleótidos, que separa la primera secuencia de nucleótidos (diana) en dos partes o fragmentos - tal como aproximadamente dos partes o fragmentos del mismo tamaño. Típicamente, las partes o fragmentos tendrán al menos una longitud de aproximadamente 16 nucleótidos.

25 En un aspecto adicional, se proporciona una base de datos de secuencias de 6-50 pares de bases que flanquean directamente, y opcionalmente incluyen, el sitio de reconocimiento de la enzima de restricción primaria de cada secuencia diana incluida, y que puede utilizarse en los métodos descritos en el presente documento para identificar cada secuencia diana.

30 En otro aspecto, se proporciona una base de datos de secuencias de 12-50 pares de bases que flanquean directamente todos los sitios de reconocimiento de las enzimas de restricción primarias en el genoma y que pueden utilizarse en los métodos descritos en el presente documento para determinar la posición genómica de cada una de las secuencias capturadas identificadas.

35 En otro aspecto, se proporciona una base de datos de secuencias de 6-50 pares de bases que flanquean directamente, y opcionalmente incluyen, el sitio de reconocimiento de enzima de restricción secundario relevante de cada una de las secuencias diana incluidas, y que puede utilizarse en los métodos descritos en el presente documento para identificar cada secuencia diana.

40 En otro aspecto, se proporciona una base de datos de secuencias de 12-50 pares de bases que flanquean directamente todos los sitios de reconocimiento de enzimas de restricción secundarias relevantes en el genoma y que pueden utilizarse en los métodos descritos en el presente documento para determinar la posición genómica de cada una de las secuencias capturadas identificadas.

#### PRIMERA SECUENCIA DE NUCLEÓTIDOS

45 La primera secuencia de nucleótidos es una secuencia de nucleótidos diana.

50 Tal y como se utiliza en el presente documento, la expresión "secuencia de nucleótidos diana" se refiere a la secuencia que se utiliza como una secuencia cebo para identificar la una o más secuencias con las que se entrecruza (por ejemplo una o más secuencias de nucleótidos de interés o una o más secuencias de la composición de secuencias de nucleótidos desconocida).

55 La secuencia de nucleótidos diana es de secuencia conocida.

60 El entrecruzamiento es indicativo de que la secuencia de nucleótidos diana y la secuencia entrecruzada con la misma están originalmente cercanas en el espacio nuclear. Determinando la frecuencia a la cual las secuencias están próximas entre sí, es posible entender, por ejemplo, la conformación de los cromosomas y de las regiones cromosómicas en el contexto espacial del núcleo (por ejemplo *in vivo* o *in vitro*). Así mismo, es posible entender las organizaciones estructurales complejas dentro del genoma, por ejemplo, cuando los potenciadores u otros elementos reguladores de la transcripción comunican con promotores distantes localizados en *cis* o incluso en *trans*.



Adicionalmente, es incluso posible entender el posicionamiento de una región genómica determinada con respecto a las secuencias de nucleótidos presentes en el mismo cromosoma (en *cis*) así como a las secuencias de nucleótidos en otros cromosomas (en *trans*). Por tanto, es posible mapear secuencias de nucleótidos en diferentes cromosomas que frecuentemente comparten sitios en el espacio nuclear. Así mismo, es también posible detectar aberraciones genéticas equilibradas y/o desequilibradas - tales como translocaciones, deleciones, inversiones, duplicaciones u otros reordenamientos genómicos equilibrados y/o no equilibrados (por ejemplo, deleciones o translocaciones en uno o más cromosomas). En este sentido, las aberraciones genéticas resultan en cambios de interacciones ADN-ADN en la posición en la que el cambio se ha producido, que pueden detectarse.

La primera secuencia de nucleótidos (diana) de acuerdo con la presente invención puede ser cualquier secuencia en la que se desee determinar la frecuencia de interacción en el espacio nuclear con una o más secuencias distintas.

En una realización, la primera secuencia de nucleótidos (diana) será mayor que aproximadamente 350 pb de longitud dado que se selecciona una enzima de restricción secundaria que corta la primera secuencia de nucleótidos (diana) a aproximadamente 350 pb o más del sitio de restricción primario. Esto puede minimizar un sesgo en la formación de círculo debido a limitaciones topológicas (Rippe *et al.* (2001) Trends in Biochem. Sciences 26, 733-40).

Adecuadamente, la primera secuencia de nucleótidos (diana) después de la amplificación comprende al menos aproximadamente 32 pb en virtud del hecho de que la longitud mínima de los al menos dos cebadores de amplificación utilizados para amplificar la segunda secuencia de nucleótidos tiene aproximadamente 16 bases cada una.

En una realización preferida, la primera secuencia de nucleótidos (diana) puede comprender completa o parcialmente (por ejemplo, un fragmento), o puede estar cercana a (por ejemplo en proximidad de), un promotor, un potenciador, un silenciador, un aislante, una región de unión a matriz, una región de control de locus, una unidad de transcripción, un origen de replicación, un punto caliente de recombinación, una rotura de translocación, un centrómero, un telómero, una región de densidad génica, una región pobre génica, un elemento repetitivo, un sitio de integración (vírico), una secuencia de nucleótidos en la que las deleciones y/o mutaciones están relacionadas con un efecto (por ejemplo enfermedad, efecto fisiológico funcional o estructural - tal como un polimorfismo mononucleotídico (SNP) o una o más secuencias de nucleótidos que contienen dichas deleciones y/o mutaciones, o cualquier secuencia en la que se desee determinar la frecuencia de interacción en el espacio nuclear con otras secuencias.

Como se ha mencionado anteriormente, la primera secuencia de nucleótidos (diana) puede comprender completa o parcialmente (por ejemplo un fragmento), o estar cerca de (por ejemplo próxima a) una secuencia de nucleótidos en la que aberraciones genéticas - tales como deleciones y/o mutaciones - están relacionadas con un efecto (por ejemplo una enfermedad). De acuerdo con esta realización de la invención la primera secuencia de nucleótidos (diana) puede por lo tanto ser una secuencia de nucleótidos (por ejemplo un gen o un locus), adyacente a (en el molde de ADN físico), o en la región genómica en la que los cambios se han asociado con o correlacionados con una enfermedad - tal como una enfermedad genética o congénita. En otras palabras, la primera secuencia de nucleótidos (diana) puede ser o puede seleccionarse basándose en su asociación con un fenotipo clínico. En una realización preferida, los cambios son cambios en uno o más cromosomas y la enfermedad puede ser como consecuencia de, por ejemplo, una o más deleciones, una o más translocaciones, una o más duplicaciones y/o una o más inversiones etc. en su interior. Como ejemplos no limitantes de dichos genes/locus se encuentran *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, locus de la inmunoglobulina, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCRα/δ*, *TCRβHOX* y otros locus en varias leucemias linfoblásticas

Otros ejemplos se describen en bases de datos electrónicas, tales como:

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cancerchromosomes>  
<http://cgap.nci.nih.gov/Chromosomes/Mitelman>  
<http://www.progenetix.net/progenetix/P14603437/ideogram.html>  
<http://www.changbioscience.com/cytogenetics/cyto1.pl?query=47,xy>  
<http://www.possum.net.au/>  
<http://www.lmddatabases.com/>  
<http://www.wiley.com/legacy/products/subject/life/borgaonkar/index.html>  
<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>  
<http://www.sanger.ac.uk/PostGenomics/decipher/>  
<http://agserver01.azn.nl:8080/ecaruca/ecaruca.jsp>

Otros ejemplos se describen en "Catalogue of Unbalanced Chromosome Aberrations in Man" 2ª edición. Albert Schinzel. Berlín: Walter de Gruyter, 2001. ISBN 3-11-011607-3.

En una realización, el término "adyacente" significa "directamente adyacente" de tal manera que no hay nucleótidos de intervención entre dos secuencias adyacentes.

En otra realización, el término “adyacente” en el contexto de la secuencia de ácido nucleico y el sitio de reconocimiento de enzima de restricción primaria significa “directamente adyacente” de tal manera que no hay secuencias intermedias entre la secuencia de ácidos nucleicos y el sitio de reconocimiento de la enzima de restricción primaria.

## 5 SEGUNDA SECUENCIA DE NUCLEÓTIDOS

La segunda secuencia de nucleótidos puede obtenerse, se obtiene, se identifica o puede identificarse mediante entrecruzamiento de ADN genómico (por ejemplo *in vivo* o *in vitro*).

10 La segunda secuencia de nucleótidos (por ejemplo secuencia de nucleótidos de interés) se liga con la primera secuencia de nucleótidos (diana) después del tratamiento de una muestra con un agente de entrecruzamiento y digestión/ligamiento de los fragmentos de ADN entrecruzados. Dichas secuencias se entrecruzan con la primera secuencia de nucleótidos (diana) debido a que estaban originalmente cerca en el espacio nuclear y están ligadas con la primera secuencia de nucleótidos (diana) debido a que las condiciones de ligamiento favorecen el ligamiento entre fragmentos de ADN entrecruzado (intramolecular) sobre eventos de ligamiento al azar.

15 Las enfermedades basadas en alteraciones - tales como translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos - están generalmente ocasionadas por interacciones entre ADN-ADN aberrantes. La tecnología 4C mide frecuencias de interacción ADN-ADN, que principalmente son una función de la preparación del sitio genómico, es decir, frecuencias de interacción ADN-ADN son inversamente proporcionales a la distancia lineal (en kilobases) entre dos locus de ADN presentes en el mismo molde de ADN físico (Dekker *et al.*, 2002). Por lo tanto, la alteración o alteraciones que crean moldes de ADN nuevos y/o físicamente diferentes, viene acompañada por interacciones alteradas de ADN-ADN y esto puede medirse por tecnología 4C.

20 Adecuadamente, la segunda secuencia de nucleótidos tiene al menos 40 pares de bases.

25 Los agentes de entrecruzamiento - tales como formaldehído - pueden utilizarse para entrecruzar proteínas con otras proteínas vecinas y ácidos nucleicos. Por tanto, dos o más secuencias de nucleótidos pueden entrelazarse únicamente mediante proteínas unidas a (una de) estas secuencias de nucleótidos. También pueden utilizarse agentes de entrecruzamiento distintos de formaldehído de acuerdo con la presente invención, incluyendo aquellos agentes de entrecruzamiento que entrecruzan directamente secuencias de nucleótidos. Como ejemplos de agentes de entrecruzamiento de ADN se incluyen, pero sin limitación, luz UV, mitomicina C, mostaza de nitrógeno, melfalán, 1,3-butadieno diepóxido, *cis* diaminodicloroplatino (II) y ciclofosfamida.

30 Adecuadamente, el agente de entrecruzamiento formará entrecruzamientos que unen mediante puentes distancias relativamente cortas - tales como de aproximadamente 2 Å - seleccionando de este modo interacciones íntimas que pueden invertirse.

35 El entrecruzamiento puede realizarse, por ejemplo, incubando las células en formaldehído al 2 % a temperatura ambiente - tal como incubando  $1 \times 10^7$  células en 10 ml de DMEM-FCS 10 % complementado con formaldehído al 2 % durante 10 minutos a temperatura ambiente.

## 40 ENZIMA DE RESTRICCIÓN PRIMARIA

45 Tal y como se utiliza en el presente documento, la expresión “enzima de restricción primaria” se refiere a una primera enzima de restricción que se utiliza para digerir el ADN entrecruzado.

50 La enzima de restricción primaria se seleccionará dependiendo del tipo de secuencia diana (por ejemplo locus) a analizar. Es deseable que se realicen experimentos preliminares para optimizar las condiciones de la digestión.

55 La enzima de restricción primaria puede seleccionarse entre enzimas de restricción que reconocen al menos secuencias de 8 pb o más de ADN. La enzima de restricción primaria puede seleccionarse entre enzimas de restricción que reconocen al menos secuencias de 7 pb o más de ADN. La enzima de restricción primaria puede seleccionarse entre enzimas de restricción que reconocen al menos secuencias de 6 pb o más de ADN. En algunas realizaciones, la enzima de restricción primaria puede seleccionarse entre enzimas de restricción que reconocen una secuencia de 4 pb y/o 5 pb de ADN.

60 El uso de cortadores menos frecuentes aumentará la distancia genómica capturada (cubierta) por cada cebo.

Las enzimas de restricción que reconocen secuencias de ADN de 6 pb incluyen, sin limitación, AclI, HindIII, SspI, BspLU11I, AgeI, MluI, SpeI, BglII, Eco47III, StuI, ScaI, ClaI, AvaII, VspI, MfeI, PmaCI, PvuII, NdeI, NcoI, SmaI, SacII, AvrII, PvuI, XmaII, SphI, XhoI, PstI, AflII, EcoRI, AatII, SacI, EcoRV, SphI, NaeI, BsePI, NheI, BamHI, NarI, ApaI, KpnI, SnaI, Sall, ApaLI, HpaI, SnaBI, BspHI, BspMII, NruI, XbaI, BclI, MstI, Ball, Bsp1407I, PstI, AsuII y AhaIII.

Las enzimas de restricción que reconocen más de una secuencia de ADN de 6 pb incluyen, pero sin limitación BbvC, I, AscI, AsiS I, Fse I, Not I, Pac I, Pme I, Sbf I, SgrA I, Swa I, Sap I, Cci NI, FspA I, Mss I, Sgf I, Smi I, Srf I y Sse8387 I.

- 5 En algunos aspectos de la presente invención, en el caso de enzimas de restricción que reconocen secuencias de 6 pb, se prefieren *BglII*, *HindIII* o *EcoRI*.

10 Las enzimas de restricción que reconocen secuencias de ADN de 4 o 5 pb incluyen, pero sin limitación TspEI, MaeII, AluI, NlaIII, HpaII, FnuDII, MaeI, DpnI, MboI, HhaI, HaeIII, RsaI, TaqI, CviRI, MseI; Sth132I, AclI, DpnII, Sau3AI y MnlI. En una realización la enzima de restricción secundaria es NlaIII y/o DpnII.

La expresión "sitio de reconocimiento de enzima de restricción primaria" se refiere al sitio en una secuencia de nucleótidos que reconoce y escinde la enzima de restricción primaria.

- 15 En algunas realizaciones, la enzima de restricción no digiere ADN repetitivo o ADN que está relativamente infra representado en ADN repetitivo. Esto puede aumentar el número de lecturas interpretables.

#### ENZIMA DE RESTRICCIÓN SECUNDARIA

20 Tal y como se usa en el presente documento, la expresión "enzima de restricción secundaria" se refiere a una segunda enzima de restricción que se utiliza opcionalmente después de una digestión con enzima de restricción primaria, ligamiento de ADN entrecruzado, desentrecruzamiento y (opcional) purificación de ADN. En una realización, la enzima de restricción secundaria se utiliza para proporcionar extremos de ADN definidos en las secuencias de nucleótidos de interés, lo que permite el ligamiento de secuencias de composición nucleotídica conocida en los sitios de reconocimiento de enzimas de restricción secundarias que flanquean las secuencias de nucleótidos de interés.

30 En una realización, el ligamiento de secuencias de composición nucleotídica conocida en los sitios de reconocimiento de enzima de restricción secundaria que flanquean (por ejemplo, que están a cada lado o en el extremo) las secuencias de nucleótidos de interés implica el ligamiento en condiciones diluidas para favorecer el ligamiento intramolecular entre los sitios de reconocimiento de enzimas de restricción secundarias que flanquean las secuencias de nucleótidos diana y las secuencias de nucleótidos de interés unidas. Esto produce de un modo eficaz la formación de círculos de ADN en los que secuencias de nucleótidos diana conocidas flanquean secuencias de interés desconocidas.

35 En otra realización, el ligamiento de secuencias de composición nucleotídica conocida en los sitios de reconocimiento de la enzima de restricción secundaria que flanquean (por ejemplo, están a cada lado o en el extremo) de las secuencias de nucleótidos de interés implica la adición de secuencias de ADN únicas de composición nucleotídica conocida, seguido de ligamiento en condiciones que favorecen el ligamiento intermolecular entre los sitios de reconocimiento de la enzima de restricción secundaria que flanquean las secuencias de nucleótidos de interés y secuencias de ADN únicas introducidas de composición nucleotídica conocida.

45 En una realización, la enzima de restricción secundaria se selecciona de tal manera que no haya sitios de enzimas de restricción secundarias entre aproximadamente 350 pb (por ejemplo 350-400 pb) del sitio de restricción primario.

50 En otra realización, la enzima de restricción secundaria se selecciona de tal manera que el mismo sitio de enzima de restricción secundaria está probablemente localizado en la secuencia de nucleótidos ligada (es decir, la secuencia entrecruzada ligada). Dado que los extremos de la primera secuencia de nucleótidos (diana) y la secuencia de nucleótidos ligada pueden ser extremos cohesivos (o romos) compatibles, las secuencias pueden incluso ligarse para circularizar el ADN. Por consiguiente, después de la etapa de digestión se realiza ligamiento en condiciones diluidas que favorecen las interacciones intramoleculares y la circularización opcional del ADN mediante extremos compatibles.

55 Preferentemente, el sitio de reconocimiento de la enzima de restricción secundaria es un sitio de reconocimiento de secuencia de nucleótidos de 4 o 5 pb. Las enzimas que reconocen secuencias de ADN de 4 o 5 pb incluyen, pero sin limitación, TspEI, MaeII, AluI, NlaIII, HpaII, FnuDII, MaeI, DpnI, MboI, HhaI, HaeIII, RsaI, TaqI, CviRI, MseI, Sth132I, AclI, DpnII, Sau3AI y MnlI.

60 En una realización preferida, la enzima de restricción secundaria es NlaIII y/o DpnII.

La expresión "sitio de reconocimiento de enzima de restricción secundaria" se refiere al sitio en la secuencia de nucleótidos que reconoce y escinde la enzima de restricción secundaria.

65 Después de la digestión con la enzima de restricción secundaria, se realiza una reacción de ligamiento adicional. En una realización, esta reacción de ligamiento une las secuencias de ADN de composición de secuencia nucleotídica

conocida con el sitio de digestión de enzima de restricción secundaria de las una o más secuencias que están ligadas con la secuencia de nucleótidos diana.

5 En algunas realizaciones, el método excluye la etapa de digerir las secuencias de nucleótidos con una enzima de restricción secundaria.

En algunas realizaciones, el método excluye ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o con los sitios de digestión con enzima de restricción secundaria disponibles que flanquean las una o más secuencias de nucleótidos de interés.

10 ENZIMA DE RESTRICCIÓN TERCIARIA

15 Tal y como se utiliza en el presente documento, la expresión "enzima de restricción terciaria" se refiere a una tercera enzima de restricción que puede utilizarse opcionalmente después de la etapa de enzima de restricción secundaria para linealizar el ADN circularizado antes de la amplificación.

Preferentemente, la enzima de restricción terciaria es una enzima que reconoce un sitio de reconocimiento de nucleótidos de 6 pb o más.

20 Preferentemente, la enzima de restricción terciaria digiere la primera secuencia de nucleótidos (diana) entre los sitios de reconocimiento de enzima de restricción primaria y secundaria. Como entenderá un experto en la materia, es deseable que la enzima de restricción terciaria no digiera la primera secuencia de nucleótidos (diana) demasiado cercana a los sitios de reconocimiento de la enzima de restricción primaria y secundaria de tal manera que los cebadores de amplificación no pueden hibridarse. Por consiguiente, se prefiere que el sitio de reconocimiento de  
25 enzima de restricción terciaria se localice al menos a la misma distancia desde los sitios de reconocimiento de enzima de restricción primaria y secundaria que la longitud del cebador utilizado de tal manera que el cebador o cebadores de amplificación puedan aún hibridarse.

30 En una realización preferida, la enzima de restricción terciaria es una que reconoce una secuencia de ADN de 6 pb.

La expresión "sitio de reconocimiento de enzima de restricción terciaria" se refiere al sitio en la secuencia de nucleótidos que reconoce y escinde la enzima de restricción terciaria.

35 SITIO DE RECONOCIMIENTO

Las endonucleasas de restricción son enzimas que escinden el esqueleto de azúcar-fosfato del ADN. En entornos más prácticos, una enzima de restricción determinada corta ambas cadenas de un ADN dúplex en un tramo de tan solo algunas bases. Los sustratos para las enzimas de restricción son secuencias de ADN bicatenario denominados sitios/secuencias de reconocimiento.

40 La longitud de los sitios de reconocimiento de restricción varía, dependiendo de la enzima de restricción que se utilice. La longitud de la secuencia de reconocimiento dictamina con qué frecuencia la enzima cortará una secuencia de ADN.

45 Como ejemplo, diversas enzimas de restricción reconocen una secuencia de ADN de 4 pb. Las secuencias y la enzima que reconocen la secuencia de ADN de 4 pb incluyen, pero sin limitación, AATT (TspEI), ACGT (MaeII), AGCT (AclI), CATG (NlaIII), CCGG (HpaII), CGCG (FnuDII), CTAG (MaeI), GATC (DpnI, DpnII, Sau3AI & MboI), GCGC (HhaI), GGCC (HaeIII), GTAC (RsaI), TCGA (TaqI), TGCA (CviRI), TTAA (MseI), CCGG (Sth132I), CCGC (AclI) y CCTC (MnlI).

50 Como un ejemplo adicional, diversas enzimas de restricción reconocen una secuencia de ADN de 6 pb. Las secuencias y la enzima que reconocen la secuencia de ADN de 6 pares de bases pb, incluyen, pero sin limitación, AACGTT (AclI), AAGCTT (HindIII), AATATT (SspI), ACATGT (BspLU11I), ACCGGT (AgeI), ACGCGT (MluI), ACTAGT (SpeI), AGATCT (BglII), AGCGCT (Eco47III), AGGCCT (StuI), AGTACT (ScaI), ATCGAT (ClaI), ATGCAT (AvaIII), ATTAAT (VspI), CAATTG (MfeI), CACGTG (PmaCI), CAGCTG (PvuII), CATATG (NdeI), CCATGG (NcoI), CCCGGG (SmaI), CCGCGG (SacII), CCTAGG (AvrII), CGATCG (PvuI), CGGCCG (XmaIII), CGTACG (SphI), CTCGAG (XhoI), CTGCAG (PstI), CTTAAG (AflIII), GAATTC (EcoRI), GACGTC (AatII), GAGTC (SacI), GATATC (EcoRV), GCATGC (SphI), GCCGGC (NaeI), GCGCGC (BsePI), GCTAGC (NheI), GGATCC (BamHI), GGCGCC (NarI), GGGCCC (ApaI), GGTACC (KpnI), GTATAC (SnaI), GTCGAC (Sall), GTGCAC (ApaLI), GTTAAC (HpaI), TACGTA (SnaBI), TCATGA (BspHI), TCCGGA (BspMII), TCGCGA (NruI), TCTAGA (XbaI), TGATCA (BclI), TGCGCA (MstI), TGGCCA (BamI), TGTACA (Bsp1407I), TTATAA (PsiI), TTCGAA (ApuI) y TTTAAA (AhaIII).

60 Como otro ejemplo adicional, diversas enzimas de restricción reconocen una secuencia de ADN de 7 pb. Las secuencias y la enzima que reconocen la secuencia de ADN de 7 pb incluyen, pero sin limitación, CCTNAGG (SauI), GCTNAGC (EspI), GGTNACC BstEII y TCCNGGA PfoI.

Como otro ejemplo adicional, diversas enzimas de restricción reconocen una secuencia de ADN de 8 pb. Las secuencias y la enzima que reconocen la secuencia de ADN de 8 pb incluyen, pero sin limitación, ATTTAAAT (SwaI), CCTGCAGG (Sse8387I), CGCCGGCG (Sse232I), CGTCGACG (SgrDI), GCCCGGGC (SrfI), GCGATCGC (SgfI), GCGGCCGC (NotI), GGCCGGCC (FseI), GGCGGCC (AseI), GTTTAAAC (PmeI) y TTAATTAA (PacI).

Diversas de estas enzimas contienen la secuencia CG que puede metilarse *in vivo*. Diversas enzimas de restricción son sensibles a esta metilación y no escindirán la secuencia metilada, por ejemplo HpaII no escindirá la secuencia CC<sup>m</sup>GG, mientras que su isóquímico MspI es insensible a esta modificación y escindirá la secuencia metilada. Por consiguiente, en algunos casos, no se utilizan enzimas sensibles a metilación eucariota.

En una realización, un sitio de reconocimiento es un sitio de digestión.

En una realización, un sitio de reconocimiento de enzima de restricción es un sitio de digestión de enzima de restricción.

### CIRCULARIZACIÓN

De acuerdo con una realización de la presente invención, el material para 4C se prepara creando círculos de ADN digiriendo el molde 3C con una enzima de restricción secundaria, seguido de ligamiento.

Preferentemente, se selecciona una enzima de restricción secundaria que corta la primera secuencia de nucleótidos (diana) a más de aproximadamente 350 pb (por ejemplo 350-400 pb) desde el sitio de restricción primario. Ventajosamente, esto minimiza un sesgo en la formación de círculo debido a limitaciones topológicas (Rippe *et al* (2001) Trends in Biochem. Sciences 26, 733-40).

Preferentemente, la enzima de restricción secundaria es una enzima cortadora frecuente, que reconoce un sitio de reconocimiento de enzima de restricción de 4 o 5 pb. Por tanto es posible obtener los fragmentos de restricción más pequeños para eficiencias de amplificación iguales de todos los fragmentos ligados durante la amplificación.

Antes de la digestión con enzima de restricción secundaria y ligamiento, el molde de ADN comprenderá un sitio de reconocimiento de enzima secundaria en la primera secuencia de nucleótidos (diana) localizada a más de aproximadamente 350-400 pb desde el sitio de restricción primario y otro sitio de reconocimiento de enzima secundaria localizado en la secuencia de nucleótidos que se ha ligado (es decir en la segunda secuencia de nucleótidos).

Preferentemente, la etapa de digestión de la enzima de restricción secundaria se realiza durante más de 1 hora hasta toda la noche y después se realiza una termoinactivación de la enzima.

Preferentemente, el ADN en esta mezcla de reacción se purifica utilizando métodos/kits convencionales conocidos en la técnica.

Después de la etapa de digestión con la enzima de restricción secundaria, un sitio de enzima de restricción secundaria se localiza a más de 350-400 pb desde el sitio de restricción primario en la primera secuencia de nucleótidos (diana) y otro sitio de enzima de restricción secundaria se localizará en la secuencia de nucleótidos ligada (es decir la segunda secuencia de nucleótidos). Dado que los extremos de la primera secuencia de nucleótidos (diana) y la secuencia de nucleótidos ligada tienen extremos compatibles, las secuencias pueden ligarse para circularizar el ADN.

La etapa de digestión se realiza después por ligamiento en condiciones diluidas que favorecen interacciones intramoleculares y la circularización del ADN mediante los extremos compatibles.

Preferentemente, la reacción de ligamiento se realiza a una concentración de ADN de aproximadamente 1-5 ng/ $\mu$ l.

Preferentemente, la reacción de ligamiento se realiza durante más de 1 h (por ejemplo 2, 3, 4 o más horas) a aproximadamente 16-25 °C.

Por consiguiente, después de la reacción de ligamiento, puede prepararse ADN circularizado. El ADN circularizado comprenderá los sitios de reconocimiento para al menos la enzima de restricción secundaria o las enzimas de restricción primaria y secundaria. En ADN circularizado que contiene la primera secuencia de nucleótidos (diana), el sitio de reconocimiento de enzima de restricción primaria y los sitios de reconocimiento de enzima de restricción secundaria definirán los extremos de la primera secuencia de nucleótidos (diana) y la secuencia de nucleótidos ligada (es decir la segunda secuencia de nucleótidos). Por consiguiente la primera secuencia de nucleótidos (diana) y la secuencia de nucleótidos ligada se separan (por ejemplo se dividen) por el sitio de reconocimiento de enzima de restricción primaria y el sitio de reconocimiento de enzima de restricción secundaria.

AMPLIFICACIÓN

Para amplificar los moldes de ADN 4C pueden realizarse una o más reacciones de amplificación.

5 La amplificación de ADN puede realizarse utilizando diversos métodos diferentes conocidos en la técnica. Por ejemplo, el ADN puede amplificarse utilizando la reacción en cadena de la polimerasa (Saiki *et al.*, 1988); PCR mediada por ligamiento, amplificación Qb replicasa (Cahill, Foster y Mahan, 1991; Chetverin y Spirin, 1995; Katanaev, Kurnasov y Spirin, 1995); la reacción en cadena de la ligasa (LCR) (Landegren *et al.*, 1988; Barany, 1991); el sistema de replicación de secuencia autosostenido (Fahy, Kwoh y Gingeras, 1991) y amplificación con desplazamiento de cadena (Walker *et al.*, 1992).

Adecuadamente, el ADN se amplifica utilizando PCR. "PCR" se refiere al método de K. B. Mullis Patentes de Estados Unidos n.º 4.683.195, 4.683.202 y 4.965.188 que describen un método para aumentar la concentración de un segmento de una secuencia de nucleótidos en una mezcla de ADN genómico sin realizar clonación o purificación.

15 En una realización, se utiliza PCR inversa. La PCR inversa (IPCR) (descrita por Ochman *et al.* (1988) Genetics 120(3), 621-3) es un método para la rápida amplificación *in vitro* de secuencias de ADN que flanquean una región de secuencia conocida. El método utiliza la reacción en cadena de la polimerasa (PCR), pero tiene los cebadores orientados en dirección inversa de la orientación normal. El molde para los cebadores inversos es un fragmento de restricción que se ha ligado consigo mismo para formar un círculo. La PCR inversa tiene muchas aplicaciones en genética molecular, por ejemplo, la amplificación e identificación de secuencias que flanquean elementos transponibles. Para aumentar la eficacia y reproducibilidad de la amplificación se prefiere que los círculos de ADN se linealicen antes de la amplificación utilizando una enzima de restricción terciaria. Preferentemente, se utiliza una enzima de restricción terciaria que es una enzima cortadora de 6 pb o más. Preferentemente, la enzima de restricción terciaria corta la primera secuencia de nucleótidos (diana) entre los sitios de enzima de restricción primaria y secundaria.

La digestión del molde 3C con la enzima de restricción secundaria, opcional circularización, ligamiento (por ejemplo ligamiento en condiciones diluidas) y opcional linealización de la primera secuencia de nucleótidos (diana) que contiene círculos produce un molde de ADN para la amplificación ("molde de ADN 4C").

Para la etapa de amplificación, se utilizan al menos dos cebadores oligonucleotídicos en los que cada cebador se hibrida con una secuencia de ADN que flanquea las secuencias de nucleótidos de interés. En una realización preferida, se utilizan al menos dos cebadores oligonucleotídicos en los que cada cebador se hibrida con la secuencia diana que flanquea las secuencias de nucleótidos de interés.

En una realización, el término "flanquea" en el contexto de hibridación de cebador significa que al menos un cebador se hibrida con una secuencia de ADN adyacente en un extremo (por ejemplo el extremo 5') de la secuencia de nucleótidos de interés y al menos un cebador se hibrida con una secuencia de ADN en el otro extremo (por ejemplo el extremo 3') de la secuencia de nucleótidos de interés. Preferentemente, al menos un cebador directo se hibrida con una secuencia de ADN adyacente a un extremo (por ejemplo el extremo 5') de la secuencia de nucleótidos de interés y al menos un cebador inverso se hibrida con una secuencia de ADN en el otro extremo (por ejemplo, el extremo 3') de la secuencia de nucleótidos de interés.

45 En una realización preferida, el término "flanquea" en el contexto de hibridación de cebador significa que al menos un cebador se hibrida con una secuencia diana adyacente en un extremo (por ejemplo el extremo 5') de la secuencia de nucleótidos de interés y al menos un cebador se hibrida con una secuencia diana en el otro extremo (por ejemplo el extremo 3') de la secuencia de nucleótidos de interés. Preferentemente, al menos un cebador directo se hibrida con una secuencia diana adyacente a un extremo (por ejemplo el extremo 5') de la secuencia de nucleótidos de interés y al menos un cebador inverso se hibrida con una secuencia diana en el otro extremo (por ejemplo, el extremo 3') de la secuencia de nucleótidos de interés.

Tal y como se utiliza en el presente documento, el término "cebador" se refiere a un oligonucleótido, tanto de origen natural como en una digestión de restricción purificada o producido sintéticamente, que es capaz de actuar como un punto de inicio de la síntesis cuando se coloca en condiciones en las que la síntesis de un producto de extensión de cebador que es complementario a una cadena de ácido nucleico se induce, (es decir, en presencia de nucleótidos y un agente de inducción tal como ADN polimerasa y a una temperatura y pH adecuados). El cebador es preferentemente monocatenario para conseguir una máxima eficiencia en la amplificación, pero puede ser bicatenario. Si es monocatenario, el cebador se trata primero para separar sus cadenas antes de utilizarse para preparar los productos de extensión. Preferentemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de los productos de extensión en presencia del agente inductor. La longitud exacta de los cebadores dependerá de muchos factores, entre los que se incluye la temperatura, la fuente del cebador y el uso del método.

Adecuadamente, los cebadores tendrán una longitud de al menos 15, preferentemente de al menos 16, 17, 18, 19 o 20, por ejemplo de al menos 25, 30 o 40 nucleótidos. Preferentemente, los cebadores de amplificación tienen una longitud de 16 a 30 nucleótidos.

5 Preferentemente, los cebadores se diseñan para estar tan próximos como sea posible a los sitios de reconocimiento de enzimas de restricción primaria y secundaria que separan la primera secuencia de nucleótidos (diana) y la segunda secuencia de nucleótidos (capturada). Los cebadores pueden diseñarse de tal manera que estén a aproximadamente 100 nucleótidos - tal como aproximadamente 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 o 1 o más nucleótidos de distancia de los sitios de reconocimiento de la enzima de restricción primaria y secundaria.

10 Adecuadamente, los cebadores pueden diseñarse de tal manera que solapen parcial o completamente con los sitios de reconocimiento de enzimas de restricción primaria y secundaria.

15 Adecuadamente, los cebadores de amplificación se diseñan de tal manera que sus extremos 3' se orientan hacia fuera hacia la segunda secuencia de nucleótidos.

En una realización, la amplificación puede combinarse con la adición de secuencias adicionales en los extremos de los productos amplificados. Preferentemente, estas secuencias adicionales son secuencias adaptadoras necesarias para la secuenciación a alto rendimiento. Adecuadamente, los cebadores contienen un saliente (por ejemplo un saliente 5'). Adecuadamente el saliente añade parte de, o la secuencia adaptadora completa, necesaria para la secuenciación a alto rendimiento. Adecuadamente, los cebadores contienen un saliente (por ejemplo un saliente 5') que añade parte de, o la secuencia completa, utilizada para cebar la reacción de secuenciación en la secuenciación a alto rendimiento. Por consiguiente, en una realización, los cebadores solapan parcial o completamente con los sitios de reconocimiento de enzimas de restricción primaria y/o secundaria con las secuencias adaptadoras y cebadoras de secuencia añadidas como salientes 5' en el cebador. Además, o como alternativa, los cebadores contienen una fracción conjugada (por ejemplo biotina) que permite la separación posterior de los productos de PCR del molde 4C genómico.

30 Si el método de amplificación que se utiliza es la PCR inversa, entonces se prefiere que las reacciones de amplificación se realicen sobre un ADN de aproximadamente 100-400 ng de molde 4C (por aproximadamente una mezcla de reacción de PCR de 50 µl) u otras cantidades de ADN para las cuales reacciones de PCR duplicadas dan resultados reproducibles (véase Figura 1) e incluyen un número máximo de eventos de ligamiento por reacción de PCR.

35 Preferentemente, la reacción de amplificación de PCR inversa se realiza utilizando el Sistema PCR Molde de Larga Expansión (Roche), utilizando el Tampón 1 de acuerdo con las instrucciones del fabricante.

#### MUESTRA

40 El término "muestra", tal y como se utiliza en el presente documento, tiene su significado natural. Una muestra puede ser cualquier entidad física que comprenda ADN que esté entrecruzado o que sea capaz de entrecruzarse. La muestra puede ser o puede proceder de material biológico.

45 La muestra puede ser o puede proceder de una o más entidades - tales como una o más células, uno o más núcleos o una o más muestras de tejido. Las entidades pueden ser o pueden proceder de cualquier entidad en las que el ADN - tal como cromatina - esté presente. La muestra puede ser o puede proceder de una o más células aisladas o de una o más muestras de tejido aislado, o de uno o más núcleos aislados.

50 La muestra puede ser o puede proceder de células vivas y/o de células muertas y/o de lisados nucleares y/o de cromatina aislada.

La muestra puede ser o puede proceder de sujetos enfermos y/o no enfermos.

55 La muestra puede ser o puede proceder de un sujeto que se sospecha que padece una enfermedad.

La muestra puede ser o puede proceder de un sujeto que va a analizarse con respecto a la probabilidad de que padezca una enfermedad en el futuro.

60 La muestra puede ser o puede proceder de material del paciente viable o no viable.

La fijación de células y tejidos para su uso en la preparación del molde 3C se describe con detalle en Splinter *et al.*, (2004) *Methods Enzymol.* 375, 493-507.

MARCADOR

Preferentemente, las secuencias de nucleótidos (por ejemplo moldes de ADN 4C amplificados, cebadores o sondas etc.) se marcan para ayudar en sus aplicaciones cadena abajo - tal como hibridación con matriz. Como ejemplo, los moldes de ADN 4C pueden marcarse utilizando cebado al azar o traslación por cortes (*nick*).

Una gran variedad de marcadores (por ejemplo indicadores) puede utilizarse para marcar las secuencias de nucleótidos descritas en el presente documento, particularmente durante la etapa de amplificación. Los marcadores adecuados incluyen radionúclidos, enzimas, agentes fluorescentes, quimioluminiscentes o cromogénicos, así como sustratos, cofactores, inhibidores, partículas magnéticas y similares. Las patentes que enseñan el uso de dichos marcadores incluyen US-A-3817837; US-A-3850752; US-A-3939350; US-A-3996345; US-A-4277437; US-A-4275149 y US-A-4366241.

Los marcadores adicionales incluyen, pero sin limitación,  $\beta$ -galactosidasa, invertasa, proteína fluorescente verde, luciferasa, cloranfenicol, acetiltransferasa,  $\beta$ -glucuronidasa, exoglucanasa y glucoamilasa. Los marcadores fluorescentes también pueden utilizarse, así como reactivos fluorescentes específicamente sintetizados con propiedades químicas particulares. Se dispone de una gran variedad de modos para medir la fluorescencia. Por ejemplo, algunos marcadores fluorescentes muestran un cambio en los espectros de excitación o emisión, algunos presentan transferencia de energía por resonancia en los que un indicador fluorescente pierde fluorescencia, mientras que un segundo gana fluorescencia, algunos presentan una pérdida (desactivación) o aparición de fluorescencia, mientras que algunos indican movimientos rotacionales.

Para obtener material suficiente para marcaje, pueden agruparse amplificaciones múltiples, en lugar de aumentar el número de ciclos de amplificación por reacción. Como alternativa, pueden incorporarse nucleótidos marcados hasta en los últimos ciclos de la reacción de amplificación (por ejemplo 30 ciclos de PCR (sin marcador) + 10 ciclos de PCR (con marcador)).

MATRIZ

En una realización particularmente ventajosa, los moldes de ADN 4C que se preparan de acuerdo con los métodos descritos en el presente documento pueden hibridarse con una matriz. Por consiguiente, puede utilizarse tecnología de matriz (por ejemplo micromatriz) para identificar secuencias de nucleótidos - tales como fragmentos genómicos - que comparten frecuentemente un sitio nuclear con una primera secuencia de nucleótidos (diana).

Las matrices existentes - tales como matrices de expresión y genómicas - pueden utilizarse de acuerdo con la presente invención. Sin embargo, la presente invención también busca proporcionar nuevas matrices (por ejemplo matrices de ADN) como se describe en el presente documento.

Una "matriz" es una colección intencionalmente creada de ácidos nucleicos que pueden prepararse bien sintéticamente o biosintéticamente y explorarse con respecto a la actividad biológica en una variedad de diferentes formatos (por ejemplo, bibliotecas de moléculas solubles y bibliotecas de oligos conectadas a perlas de resina, placas de sílice u otros soportes sólidos). Adicionalmente, el término "matriz" incluye aquellas bibliotecas de ácidos nucleicos que pueden prepararse aplicando ácidos nucleicos de básicamente cualquier longitud (por ejemplo, de 1 a aproximadamente 1000 monómeros de nucleótidos de longitud) en un sustrato.

La tecnología de matriz y las diversas técnicas y aplicaciones asociadas con esta se describe en general en numerosos libros de texto y documentos. Estos incluyen Lemieux *et al.*, 1998, Molecular Breeding 4, 277-289, Schena y Davis. Parallel Analysis with Biological Chips. en PCR Methods Manual (eds. M. Innis, D. Gelfand, J. Sninsky), Schena y Davis, 1999, Genes, Genomes and Chips. In DNA Microarrays: A Practical Approach (ed. M. Schena), Oxford University Press, Oxford, Reino Unido, 1999), The Chipping Forecast (Nature Genetics special issue; enero de 1999 Suplemento), Mark Schena (Ed.), Microarray Biochip Technology, (Eaton Publishing Company), Cortes, 2000, The Scientist 14[17]:25, Gwynne y Page, Microarray analysis: the next revolution in molecular biology, Science, 6 de agosto de 1999; y Eakins y Chu, 1999, Trends in Biotechnology, 17, 217-218.

La tecnología de matriz supera las desventajas de los métodos tradicionales en la biología molecular, que generalmente funcionan en una base de "un gen en un experimento", dando como resultado un bajo rendimiento y la incapacidad de apreciar todo el escenario de la función génica. Actualmente, las aplicaciones principales de la tecnología de matriz incluyen la diferenciación de secuencia (gen/mutación génica) y la determinación del nivel de expresión (abundancia) de genes. El perfil de la expresión génica puede realizarse utilizando tecnología de matriz, opcionalmente en combinación con técnicas proteómicas (Celis *et al.*, 2000, FEBS Lett, 480(1): 2-16; Lockhart y Winzler, 2000, Nature 405(6788): 827-836; Khan *et al.*, 1999, 20(2): 223-9). En la técnica también se conocen otras aplicaciones de tecnología de matriz; por ejemplo, descubrimiento génico, investigación del cáncer (Marx, 2000, Science 289: 1670-1672; Scherf, *et al.*, 2000, Nat Genet; 24(3): 236-44; Ross *et al.*, 2000, Nat Genet. 2000 Mar;24(3): 227-35), análisis de SNP (Wang *et al.*, 1998, Science, 280(5366): 1077-82), descubrimiento de fármacos, farmacogenómica, diagnóstico de enfermedades (por ejemplo, utilizando dispositivos microfluidos; Chemical & Engineering News, 22 de febrero, 1999, 77(8): 27-36), toxicología (Rockett y Dix (2000), Xenobiotica, 30(2): 155-77;



Afshari *et al.*, 1999, *Cancer Res*1; 59(19): 4759-60) y toxicogenómica (un híbrido de genómica funcional y toxicología molecular).

5 En general, cualquier biblioteca puede ordenarse de una manera ordenada en una matriz, separando espacialmente los miembros de la biblioteca. Los ejemplos de bibliotecas adecuadas para la matriz incluyen bibliotecas de ácido nucleico (incluyendo bibliotecas de ADN, ADNc, oligonucleótidos, etc.), bibliotecas de péptidos, polipéptidos y proteínas, así como bibliotecas que comprenden cualquiera de las moléculas, tales como bibliotecas de ligandos, entre otras.

10 Las muestras (por ejemplo, miembros de una biblioteca) generalmente se fijan o inmovilizan en una fase sólida, preferentemente en un sustrato sólido, para limitar la difusión y mezcla de las muestras. En una realización preferida, pueden prepararse bibliotecas de ADN que se unen a ligandos. En particular, las bibliotecas pueden inmovilizarse en una fase sólida sustancialmente plana, incluyendo membranas y sustratos no porosos tales como plástico y vidrio. Adicionalmente, las muestras se disponen preferentemente de tal manera que se facilita el indexado a las paredes o fondo de los pocillos, o a la superficie de la membrana. Puede utilizarse otros medios de disposición y fijación, por ejemplo, transferencia con pipeta, con gotero, mecanismos piezoeléctricos, inyección de tinta y tecnología de chorro de burbujas, aplicación electrostática, etc. En el caso de microplacas basadas en silicio, puede utilizarse la fotolitografía para disponer y fijar las muestras sobre la microplaca.

15 Las muestras pueden disponerse de manera que se “apliquen” sobre el sustrato sólido; esto puede realizarse manualmente o utilizando robots para depositar la muestra. En general, las matrices pueden describirse como macromatrices o micromatrices, siendo la diferencia el tamaño de las aplicaciones de la muestra. Típicamente las macromatrices contienen tamaños de aplicación de muestra de aproximadamente 300 micrómetros o mayores y pueden formarse fácilmente imágenes con gel existente y escáneres de transferencia. Los tamaños de aplicación de la muestra en micromatrices son típicamente menores de 200 micrómetros de diámetro y estas matrices normalmente contienen miles de aplicaciones. Por tanto, las micromatrices pueden requerir robótica especializada y equipo de formación de imágenes, que puede requerir una fabricación a medida. La instrumentación se describe generalmente en una revisión de Cortese, 2000, *The Scientist* 14[11]: 26.

20 En este campo se han descrito técnicas para la producción de bibliotecas inmovilizadas de moléculas de ADN. Generalmente, la mayoría de los métodos de la técnica anteriores que describen cómo sintetizar bibliotecas de moléculas de ácidos nucleicos monocatenarias, utilizando por ejemplo técnicas de enmascaramiento para construir diversas permutaciones de secuencias en diversas posiciones distintas en el sustrato sólido. La Patente de Estados Unidos n.º 5.837.832 describe un método mejorado para la producción de matrices de ADN inmovilizadas en sustratos de silicio basándose en una tecnología de integración a muy gran escala. En particular, la Patente de Estados Unidos n.º 5.837.832 describe una estrategia denominada “mosaico” (*tiling*) para sintetizar conjuntos específicos de sondas en localizaciones espacialmente definidas en un sustrato, que pueden utilizarse para producir las bibliotecas de ADN inmovilizadas de la presente invención. La Patente de Estados Unidos n.º 5.837.832 también proporciona referencias de técnicas anteriores que también pueden utilizarse.

Las matrices también pueden construirse utilizando química de fotodeposición.

25 Las matrices de péptidos (o peptidomiméticos) también pueden sintetizarse en una superficie de una manera que coloque cada miembro de la biblioteca distinto (por ejemplo, una secuencia peptídica única) en una localización distinta previamente definida en la matriz. La identidad de cada miembro de la biblioteca se determina mediante su localización espacial en la matriz. Las localizaciones en la matriz en las que se producen interacciones de unión entre una molécula predeterminada (por ejemplo, una diana o una sonda) y los miembros de la biblioteca reactivos se determinan, mediante lo cual se identifican las secuencias de los miembros de la biblioteca reactivos basándose en la localización espacial. Estos métodos se describen en la Patente de Estados Unidos n.º 5.143.854; en los documentos WO90/15070 y WO92/10092; en Fodor *et al.* (1991) *Science*, 251: 767; Dower y Fodor (1991) *Ann. Rep. Med. Chem.*, 26: 271.

30 Para ayudar a la detección, típicamente se utilizan marcadores (como se ha comentado anteriormente) - tal como cualquier indicador fácilmente detectable, por ejemplo, un indicador fluorescente, bioluminiscente, fosforescente, radiactivo, etc. Dichos indicadores, su detección, acoplamiento con dianas/sondas, etc. se analizan en otra parte de este documento. El marcaje de sondas y dianas también se desvela en Shalon *et al.*, 1996, *Genome Res* 6(7): 639-45.

65

Son ejemplos específicos de matrices de ADN los siguientes:

5 Formato I: la sonda de ADNc (de 500 ~ 5.000 bases de longitud) se inmoviliza en una superficie sólida tal como vidrio utilizando aplicación robótica y se expone a un conjunto de dianas bien por separado o en una mezcla. Se considera ampliamente que este método se ha desarrollado en la Universidad de Stanford (Ekins y Chu, 1999, Trends in Biotechnology, 1999, 17, 217-218).

10 Formato II: una matriz de oligonucleótidos (oligonucleótidos de 20-25 unidades, preferentemente, oligonucleótidos de 40-60 unidades) o sondas de ácidos peptidonucleicos (PNA) se sintetizan bien *in situ* (en microplaca) o mediante síntesis convencional seguido de inmovilización en microplaca. La matriz se expone a ADN de muestra marcado, se hibrida y se determina la identidad/abundancia de las secuencias complementarias. Dicha microplaca de ADN se comercializa en Affymetrix, Inc., con el nombre comercial GeneChip®. Agilent y Nimblegen también proporcionan matrices adecuadas (por ejemplo matrices en mosaico genómicas).

15 Como ejemplos de algunos formatos de micromatriz disponibles en el mercado se indican en la Tabla 1 a continuación (véase también Marshall y Hodgson, 1998, Nature Biotechnology, 16(1), 27-31).

Tabla 1: Ejemplos de formatos de micromatriz de hibridación actualmente disponibles

Compañía	Nombre del producto	Método de micromatriz	Etapas de hibridación	Lectura
Affymetrix, Inc., Santa Clara, California	GeneChip®	Síntesis fotolitográfica <i>in situ</i> (en placa) de ~20-25 oligómeros en obleas de silicio, que se cortan en placas de 1,25 cm <sup>2</sup> o 5,25 cm <sup>2</sup>	10.000-260.000 elementos oligo sondados con fragmentos de 30-40 nucleótidos marcados de ADNc de muestra o ARN antisentido	Fluorescencia
Brax, Cambridge, Reino Unido		Oligo sintético corto, sintetizado fuera de la placa	1000 oligos en un "placa universal" sondadas con ácido nucleico marcado	Espectrometría de masas
Gene Logic, Inc., Columbia, Maryland	READS™			
Genometrix Inc., The Woodlands, Texas	Universal Arrays™			
GENSET, Paris, Francia				
Hyseq Inc., Sunnyvale, California	HyChip™	Muestras de ADN de 500-2000 nt imprimidas sobre membranas de 0,6 cm <sup>2</sup> (HyGnostics) o ~18 cm <sup>2</sup> (Gene Discovery) Oligos de 5 unidades fabricados imprimidos como matrices de 1,15 cm sobre vidrio (HyChip)	Aplicaciones de ADNc de 64 muestras sondadas con 8.000 oligómeros (HyGnostics) o ADNc de muestra ≤55.000 sondadas con 300 oligómeros (Gene Discovery) Universal aplicaciones de 1024 oligo sondadas en ADNc de muestra 10 kb, oligo marcado de 5 meros y ligasa	Radioisótopo  Fluorescencia
Incyte Pharmaceutica ls, Inc., Palo Alto, California	GEM	Impresión piezoeléctrica para aplicación de fragmentos de PCR y síntesis en placa de oligos	≤1000 (eventualmente 10.000) aplicaciones de fragmentos de PCR/oligo sondados con ARN marcado	Fluorescencia y radioisótopo
Molecular Dynamics, Inc., Sunnyvale, California	Storm® FluorImager®	ADNc de 500-5000 nt imprimidos con bolígrafo sobre ~10 cm <sup>2</sup> en portaobjetos de vidrio	~10.000 aplicaciones de ADNc sondado con los ADNc de muestra marcados con de 200-400 nt	Fluorescencia

Compañía	Nombre del producto	Método de micromatriz	Etapas de hibridación	Lectura
Nanogen, San Diego, California	Semiconductor Microchip	Oligos de ~20 unidades prefabricados, capturados sobre aplicaciones electroactivas en obleas de silicio, que se cortan en microplacas de $\leq 1 \text{ cm}^2$	Aplicaciones de 25, 64, 400 (y eventualmente 10.000) oligos polarizados para potenciar la hibridación con ADNc de muestra marcados de 200-400 nt	Fluorescencia
Protogene Laboratories, Palo Alto, California		Síntesis en placa de 40-50 oligómeros en placa de vidrio de $9 \text{ cm}^2$ mediante impresión en una matriz de tensión superficial	$\leq 8.000$ aplicaciones oligo sondadas con ácidos nucleicos de muestra marcados 200-400 nt	Fluorescencia
Sequenom, Hamburgo, Alemania, y San Diego, California	MassArray SpectroChip	Impresión <i>offset</i> de matriz; de oligos de aproximadamente 20-25 unidades	250 localizaciones por SpectroChip interrogado por desorción láser y espectroscopía de masas	Espectrometría de masas
Svnteni, Inc., Fremont, California	UniGEM™	ADNc de 500-5.000 nt imprimidos por punta sobre placa de vidrio de $\sim 4 \text{ cm}^2$	$\leq 10.000$ aplicaciones de ADNc sondadas con ADNc de muestra marcada de 200-400 nt	Fluorescencia
Nimblegen Systems Inc., Madison	Micromatriz de 60 unidades de genoma completo de <i>Homo sapiens</i>	38.000 transcritos con 5 sondas por gen 17,4 mm x 13 mm		Plataforma de barrido de 5 micrómetros
The German Cancer Institute, Heidelberg, Alemania		Macroplaca de PNA prototípica con síntesis en placa de sondas que usan química f-moc o t-moc	Alrededor de 1.000 aplicaciones en una microplaca de 8 x 12 cm	Fluorescencia/espectrometría de masas

5 Para generar datos de ensayos basados en matriz se detectó una señal que significa la presencia de o la ausencia de hibridación entre una sonda y una secuencia de nucleótidos. La presente invención contempla también técnicas de marcaje directo e indirecto. Por ejemplo, el marcaje directo incorpora colorantes fluorescentes directamente en las secuencias de nucleótidos que se hibridan con las sondas asociadas a matriz (por ejemplo se incorporan colorantes en la secuencia de nucleótidos por síntesis enzimática en presencia de nucleótidos marcados o cebadores de PCR). Los esquemas de marcaje directo proporcionan señales fuertes de hibridación, típicamente utilizando familias de colorantes fluorescentes con estructuras químicas y características similares y son sencillas de implementar. En realizaciones preferidas que comprenden el marcaje directo de ácidos nucleicos, se utilizan cianina o análogos de Alexa en múltiples análisis de matriz comparativos con flúor. En otras realizaciones, los esquemas de marcaje indirecto pueden utilizarse para incorporar epítopos en los ácidos nucleicos bien antes o después de la hibridación a las sondas de micromatriz. Se usan uno o más procedimientos de tinción y reactivos para marcar el complejo hibridado (por ejemplo, una molécula fluorescente que se une a los epítopos, mediante lo cual se proporciona una señal fluorescente en virtud de la conjugación de la molécula de colorante con el epítipo de la especie hibridada).

10 Los análisis de datos son también una parte importante de un experimento en el que intervienen matrices. Los datos sin procesar de un experimento de matriz típicamente son imágenes, que requieren transformarse en matrices - tablas en las que las filas representan por ejemplo genes, las columnas representan por ejemplo diversas muestras tales como tejidos o condiciones experimentales y números en cada célula por ejemplo caracterizan la expresión de una secuencia particular (preferentemente, una segunda secuencia que se ha ligado a la primera secuencia de nucleótidos (diana)) en la muestra particular. Estas matrices deben analizarse adicionalmente, si se va a extraer cualquier conocimiento sobre el proceso biológico subyacente. Se desvelan métodos de análisis de datos (incluyendo datos de análisis supervisados y no supervisados así como estrategias bioinformáticas) en Brazma y Vilo J (2000) FEBS Lett 480(1): 17-24.

25 Como se describe en el presente documento la una o más secuencias de nucleótidos (por ejemplo el molde de ADN) que se marca y posteriormente se hibrida con una matriz comprende una secuencia de nucleótidos que está

enriquecido de tramos pequeños de secuencias con una firma distintiva, es decir que abarca la secuencia de nucleótidos entre el sitio de reconocimiento de la enzima de restricción primaria que se ligó durante el procedimiento 3C con la primera secuencia de nucleótidos (diana), y sus sitios de reconocimiento de enzimas de restricción secundarias colindantes respectivos.

5 Una sola matriz puede comprender múltiples secuencias cebo (por ejemplo dos o más).

#### MAPA DE CROMOSOMA

10 En el presente documento se describe adicionalmente un método que implica la caracterización de fragmentos de ADN entrecruzados como consecuencia de su proximidad espacial o física, con el propósito de (re)construcción de mapas de cromosomas (por ejemplo mapas de cromosomas lineales) y la identificación diagnóstica de cambios en estos mapas de cromosomas.

15 Ventajosamente, dichas técnicas pueden aplicarse para fines de diagnóstico - tales como reconstrucción de mapas de cromosomas físicos e identificación de cambios en estos mapas como consecuencia de reordenamientos genómicos.

20 La metodología también puede utilizarse para identificar variación genómica (natural) que no está necesariamente asociada a enfermedad pero que puede predisponer a un sujeto a un determinado rasgo (por ejemplo un rasgo mental o conductual).

25 En un aspecto, se proporciona un método para construir al menos un mapa de cromosoma lineal de un individuo que comprende las etapas de: (a) proporcionar una muestra de ácido nucleico de dicho individuo; (b) entrecruzar el ácido nucleico en dicha muestra; (c) digerir el ácido nucleico con una enzima de restricción primaria; (d) ligar las secuencias nucleotídicas entrecruzadas; (e) invertir el entrecruzamiento; (f) analizar los productos de ligamiento; (g) construir al menos un mapa de cromosoma lineal; y (h) identificar una o más ordenaciones genómicas en el mapa de cromosoma lineal de dicho individuo.

30 En un aspecto adicional, se proporciona un método para diagnóstico de una enfermedad o identificación de un rasgo causado por una o más reordenamientos genómicos en un cromosoma que comprende las etapas de: (a) digerir una muestra de ADN entrecruzado con una enzima de restricción primaria; (b) ligar las secuencias de nucleótidos entrecruzadas; (c) invertir el entrecruzamiento; (d) analizar los productos de ligamiento; (e) construir al menos un mapa de cromosoma lineal; (f) identificar uno o más reordenamientos genómicos en el mapa de cromosoma lineal; y (g) correlacionar la uno o más reordenamientos genómicos con una enfermedad o un rasgo.

#### SONDAS

40 Tal y como se utiliza en el presente documento, el término "sonda" se refiere a una molécula (por ejemplo un oligonucleótido, tanto de origen natural como en una digestión de restricción purificada o producida sintéticamente, de manera recombinante o por amplificación con PCR), que tenga la capacidad de hibridarse con otra molécula de interés (por ejemplo, otro oligonucleótido). Cuando las sondas son oligonucleótidos estas pueden ser mono o bicatenarias. Las sondas son útiles en la detección, identificación y aislamiento de dianas particulares (por ejemplo, secuencias génicas). Como se describe en el presente documento, se contempla que las sondas utilizadas en la presente invención pueden marcarse con marcador de tal manera que sea detectable en cualquier sistema de detección, incluyendo, pero sin limitación, sistemas de enzima (por ejemplo ELISA, así como ensayos histoquímicos basados en enzimas), fluorescentes, radiactivos y luminiscentes.

50 Con respecto a matrices y micromatrices, el término "sonda" se usa para referirse a cualquier material que puede hibridarse que se fije a la matriz con el fin de detectar una secuencia de nucleótidos que se hibrida con dicha sonda. Preferentemente, estas sondas tienen una longitud de 25-60 unidades o son más largas.

En los documentos WO95/11995, EP 717.113 y WO97/29212, se describen estrategias de diseño de sondas

55 Dado que la tecnología 4C permite realizar una exploración imparcial hologenómica de interacciones, es ventajoso preparar una matriz con sondas que indague acerca de cada posible sitio de reconocimiento de enzima de restricción primaria (por ejemplo único/no repetitivo) en el genoma. Por tanto, el diseño de matrices solo depende de la elección de la enzima de restricción primaria y no de las secuencias reales de nucleótidos primaria o secundaria.

60 Aunque existen matrices que pueden utilizarse de acuerdo con la presente invención, se prefiere el uso de configuraciones alternativas.

65 En una configuración, se diseña una o más sondas en la matriz de tal manera que pueden hibridarse en estrecha proximidad con los sitios que se digieren por la enzima de restricción primaria. Más preferentemente, la sonda o las sondas están a una distancia de aproximadamente 20 pb del sitio de reconocimiento de la enzima de restricción

primaria. Más preferentemente, la sonda o las sondas están a una distancia de aproximadamente 50 pb del sitio de reconocimiento de la enzima de restricción primaria.

5 Adecuadamente, la sonda o las sondas están a una distancia de aproximadamente 100 pb (por ejemplo de aproximadamente 0-100 pb, de aproximadamente 20-100 pb) del sitio de reconocimiento de la enzima de restricción primaria.

10 En una configuración preferida, se diseña una sonda única, exclusiva a una distancia de 100 pb a cada lado de los sitios que se digieren por la enzima de restricción primaria.

15 En otra configuración preferida, las posiciones de sitios digeridos por la enzima de restricción secundaria con respecto a las posiciones de sitios digeridos por los sitios de restricción primarios se toman en cuenta. En esta configuración, se diseña una sola sonda exclusiva únicamente en cada lado de los sitios digeridos por la enzima de restricción primaria que tiene el sitio de reconocimiento de enzima de restricción secundaria más cercano a una distancia lo suficientemente larga para diseñar una sonda de una longitud determinada entre el sitio de reconocimiento de la enzima de restricción primaria y secundaria. En esta configuración, por ejemplo, no se diseñan sondas en el lado de un sitio de reconocimiento de enzima de restricción primaria particular que tiene un sitio de reconocimiento de enzima de restricción secundaria a una distancia de 10 pb en el mismo lado.

20 En otra configuración, se diseñan las sondas en la matriz de tal manera que se pueden hibridar en cada lado de los sitios que se digieren por la enzima de restricción primaria. Adecuadamente, una sola sonda a cada lado del sitio de reconocimiento de la enzima de restricción primaria puede utilizarse.

25 En otra configuración adicional, pueden diseñarse dos o más sondas (por ejemplo, 3, 4, 5, 6, 7 u 8 o más) a cada lado del sitio de reconocimiento de la enzima de restricción primaria, que después puede utilizarse para investigar el mismo evento de ligamiento. Para el número y posición de sondas relativas a cada sitio de reconocimiento de enzima de restricción primaria, la localización genómica exacta de su sitio de reconocimiento de enzima de restricción secundaria adyacente puede tenerse en cuenta.

30 En otra configuración adicional, pueden diseñarse dos o más sondas (por ejemplo, 3, 4, 5, 6, 7 u 8 o más) cerca de cada sitio de reconocimiento de enzima de restricción primaria independientemente del sitio de reconocimiento de enzima de restricción secundario más cercano. En esta configuración, todas las sondas deben estar aún cerca de los sitios de reconocimiento de enzima de restricción primaria (preferentemente a una distancia de 300 pb del sitio de restricción).

35 Ventajosamente, el último diseño y también el diseño que utiliza 1 sonda por (lado de un) sitio de reconocimiento de enzima de restricción primaria, permite el uso de diferentes enzimas de restricción secundarias en combinación con una determinada enzima de restricción primaria.

40 Ventajosamente, el uso de múltiples sondas (por ejemplo 2, 3, 4, 5, 6, 7 u 8 o más) por sitio de reconocimiento de enzima de restricción primaria puede minimizar el problema de la obtención de resultados negativos falsos debido a un mal comportamiento de sondas individuales. Adicionalmente, esto también puede aumentar la fiabilidad de los datos obtenidos con un solo experimento de microplaca y reducir el número de matrices necesarios para extraer conclusiones estadísticamente fiables.

45 Las sondas para su uso en la matriz pueden tener una longitud mayor de 40 nucleótidos y pueden ser isotérmicas.

Preferentemente, se excluyen las sondas que contienen secuencias de ADN repetitivas.

50 Se espera que las sondas de diagnóstico para los sitios de restricción que flanquean directamente o que están cerca de la primera secuencia de nucleótidos den fuertes señales de hibridación y también se excluyen del diseño de la sonda.

55 La matriz puede recuperar cualquier genoma incluyendo genoma de mamífero (por ejemplo ser humano, ratón (por ejemplo cromosoma 7)), de vertebrado (por ejemplo pez cebra)) o no vertebrado (por ejemplo de bacterias, levaduras, hongos o insectos (por ejemplo, *Drosophila*)).

60 En una realización adicional preferida, la matriz contiene 2-6 sondas alrededor de cada sitio de restricción primario único y tan cerca como sea posible al sitio de la enzima de restricción.

65 Preferentemente, la distancia máxima del sitio de digestión de la enzima de restricción es de aproximadamente 300 pb.

En una realización adicional preferida de la presente invención, se proporcionan matrices para enzimas de restricción - tales como *HindIII*, *EcoRI*, *BglII* y *NotI* - que abarcan los genomas de mamífero o no mamífero.

Ventajosamente, el diseño de matrices descrito en el presente documento impide la necesidad de rediseñar matrices para cada secuencia diana, siempre que los análisis se realicen en la misma especie.

### CONJUNTOS DE SONDAS

5 Tal y como se utiliza en el presente documento, la expresión “conjunto de sondas” se refiere a un juego o a una colección de sondas que se hibrida con cada uno de los sitios de reconocimiento de enzima de restricción primaria para una enzima de restricción primaria en un genoma.

10 Por consiguiente, en un aspecto adicional se proporciona un conjunto de sondas complementarias en secuencia a la secuencia de ácido nucleico adyacente a cada uno de los sitios de reconocimiento de enzima de restricción primaria para una enzima de restricción primaria en ADN genómico.

15 Adecuadamente, el conjunto de sondas es complementario en secuencia a los primeros 25-60 (por ejemplo, 35-60, 45-60 o 50-60) o más nucleótidos que están adyacentes a cada uno de los sitios de reconocimiento de la enzima de restricción primaria en el ADN genómico. El conjunto de sondas puede ser complementario en secuencia a un lado o a ambos lados (por ejemplo a cualquiera de ellos) del sitio de reconocimiento de enzima de restricción primaria. Por consiguiente, las sondas pueden ser complementarias en secuencia a la secuencia de ácido nucleico adyacente a cada lado de cada uno de los sitios de reconocimiento de enzima de restricción primaria en el ADN genómico.

20 También es posible definir una ventana (por ejemplo de 300 pb o menor - tal como de 250 pb, 200 pb, 150 pb o 100 pb) - desde el sitio de reconocimiento de enzima de restricción primaria) en el que pueden diseñarse una o más sondas para el conjunto. Dichos factores que son importantes en la definición de la ventana dentro de los cuales se diseñan las sondas son, por ejemplo, el contenido de GC, la ausencia de secuencias palindrómicas que pueden formar estructuras en horquilla, el tamaño máximo de tramos de un solo tipo de nucleótido. Por consiguiente, el conjunto de sondas puede ser complementario en secuencia a la secuencia de ácido nucleico que es menor de 300 pb de cada uno de los sitios de reconocimiento de la enzima de restricción primaria en el ADN genómico.

25 También es posible definir una ventana de aproximadamente 100 pb a partir del sitio de reconocimiento de la enzima de restricción primaria para identificar sondas óptimas cerca de cada sitio de restricción.

30 En realizaciones adicionales de la presente invención, el conjunto de sondas es complementario a la secuencia que es menor de 300 pb de cada uno de los sitios de reconocimiento de la enzima de restricción primaria en ADN genómico, complementario a la secuencia que está entre 200 y 300 pb de cada uno de los sitios de reconocimiento de la enzima de restricción primaria en ADN genómico y/o complementario a la secuencia que está entre 100 y 200 pb de cada uno de los sitios de reconocimiento de la enzima de restricción primaria en el ADN genómico.

35 En realizaciones adicionales de la presente invención, el conjunto de sondas es complementario a la secuencia que es de 0 a 300 pb a partir de cada uno de los sitios de reconocimiento de la enzima de restricción primaria en el ADN genómico, complementario a la secuencia que es entre 0 a 200 pb a partir de cada uno de los sitios de reconocimiento de enzima de restricción primaria en el ADN genómico y/o complementario a la secuencia que está entre 0 y 100 pb de cada uno de los sitios de reconocimiento de enzima de restricción primaria en el ADN genómico (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80 o 90 pb de cada uno de los sitios de reconocimiento de la enzima de restricción primaria en el ADN genómico).

40 Incluso pueden diseñarse dos o más sondas que son capaces de hibridar con la secuencia adyacente de cada sitio de reconocimiento de enzima de restricción primaria en el ADN genómico.

45 Las sondas pueden solaparse o solaparse parcialmente. Si las sondas se solapan entonces el solapamiento es preferentemente menor de 10 nucleótidos.

50 También pueden utilizarse fragmentos de PCR que representan los 1-300 primeros nucleótidos (por ejemplo 1-20, 1-40, 1-60, 1-80, 1-100, 1-120, 1-140, 1-160, 1-180, 1-200, 1-220, 1-240, 1-260 o 1-280 nucleótidos), que flanquean cada sitio de reconocimiento de enzima de restricción primaria.

55 También pueden utilizarse fragmentos de PCR que corresponden exactamente a cada sitio genómico que está flanqueado por el sitio de reconocimiento de enzima de restricción primaria y el primer sitio de reconocimiento de enzima de restricción secundaria adyacente. Por consiguiente, la secuencia de la sonda puede corresponder con toda o con parte de la secuencia entre cada uno de los sitios de reconocimiento de la enzima de restricción primaria y cada uno de los sitios de reconocimiento de la enzima de restricción secundaria adyacente.

60 Típicamente, las sondas, matriz de sondas o conjunto de sondas se inmovilizarán en un soporte. Los soportes (por ejemplo soportes sólidos) pueden constituirse de una variedad de materiales - tal como vidrio, sílice, plástico, nylon o nitrocelulosa. Preferentemente los soportes son rígidos y tienen una superficie plana. Típicamente los soportes tienen regiones direccionales espacialmente distintas de aproximadamente 1-10.000.000, o células. Los soportes que tienen aproximadamente 10-1.000.000 o aproximadamente 100-100.000 o aproximadamente 1000-100.000

células son comunes. La densidad de células es típicamente al menos aproximadamente de 1000, 10.000, 100.000 o 1.000.000 de células en un centímetro cuadrado. En algunos soportes, todas las células están ocupadas por mezclas agrupadas de sondas o un conjunto de sondas. En otros soportes, algunas células están ocupadas por mezclas agrupadas de sondas o un conjunto de sondas, y otras células están ocupadas, al menos al grado de pureza obtenible por métodos de síntesis, por un solo tipo de oligonucleótido.

Preferentemente, la matriz descrita en el presente documento comprende más de una sonda por sitio de reconocimiento de enzima de restricción primaria, que en el caso de una enzima de restricción de corte de 6 pb se produce, por ejemplo, aproximadamente 750.000 veces por genoma humano o de ratón.

Para una enzima de restricción que reconoce una secuencia de reconocimiento > 6 pb, puede utilizarse una sola matriz de aproximadamente 2 x 750.000 sondas para abarcar, por ejemplo, el genoma humano o de ratón completo, con 1 sonda a cada lado de cada sitio de restricción.

En un diseño de matriz preferido, el número total de moléculas de sonda de una secuencia de nucleótidos determinada presente en la matriz está en gran exceso con fragmentos homólogos presentes en la muestra 4C a hibridar con dicha matriz. Dada la naturaleza de la tecnología 4C, los fragmentos que representan regiones genómicas cerca de la secuencia de nucleótidos analizada en el molde de cromatina lineal estarán en gran exceso en la muestra de hibridación 4C (como se describe en la Figura 2). Para obtener información cuantitativa sobre eficiencias de hibridación de dichos fragmentos abundantes, puede ser necesario reducir la cantidad de muestra a hibridar y/o aumentar el número de moléculas de una sonda de secuencia de oligonucleótidos determinada en la matriz.

Por tanto, para la detección de elementos de ADN reguladores que contactan frecuentemente, por ejemplo, un elemento de promotor génico este puede ser necesario para utilizar una matriz con sondas que representen solo la región genómica seleccionada (por ejemplo aproximadamente 0,5-10 Mb), pero con cada sonda única presente en posiciones múltiples en la matriz (por ejemplo de aproximadamente 100, 200, 1000). Dichos diseños también pueden preferirse para fines de diagnóstico para detectar reordenamientos genómicos locales (por ejemplo dentro de aproximadamente 10 Mb) - tales como deleciones, inversiones, duplicaciones, etc. - cerca de un sitio (por ejemplo gen de interés).

La matriz puede comprender aproximadamente 3 x 750.000 sondas, 4 x 750.000 sondas, 5 x 750.000 sondas, o preferentemente, 6 x 750.000 sondas. Más preferentemente, la matriz comprende 6 x 750.000 sondas con 2, 3, 4, 5, 6, 7 u 8 o más sondas a cada lado de cada sitio de restricción. Más preferentemente, la matriz comprende 6 x 750.000 sondas con 3 sondas a cada lado de cada sitio de restricción.

Las matrices de sondas o conjuntos de sondas pueden sintetizarse por etapas en un soporte o puede unirse en forma presintetizada. Un método de síntesis es VLSIPS.TM. (como se describe en los documentos US 5.143.854 y EP 476.014), que entraña el uso de luz para dirigir la síntesis de sondas oligonucleotídicas en matrices miniaturizadas de alta densidad. Los algoritmos para el diseño de máscaras para reducir el número de ciclos de síntesis se describen en los documentos US 5.571.639 y US. 5.593.839. Las matrices también pueden sintetizarse de una manera combinatoria para suministrar monómeros a células de un soporte mediante trayectorias de flujo mecánicamente limitadas, como se describe en el documento EP 624.059. Las matrices también pueden sintetizarse aplicando reactivos en un soporte utilizando una impresora de chorro de tinta (véase, por ejemplo el documento EP 728.520).

En el contexto de la presente invención, las expresiones "sustancialmente un conjunto de sondas", "sustancialmente la matriz de sondas" significa que el conjunto o la matriz de sondas comprende al menos aproximadamente 50, 60, 70, 80, 90, 95, 96, 97, 98 o 99 % del conjunto total o completo o matriz de sondas. Preferentemente, el conjunto o la matriz de sondas es un conjunto completo o total de sondas (es decir el 100 %).

En una realización preferida, la matriz comprende una sola sonda única por lado de cada sitio de reconocimiento de enzima de restricción primaria que está presente en un genoma determinado. Si este número de sondas supera el número de sondas que puede contenerse en una sola matriz, la matriz puede preferentemente aún contener una representación del genoma completo de una especie determinada, pero a menor resolución, con por ejemplo una de cada 2, 3, 4, 5, 6, 7, 8, 9, 10, 10<sup>2</sup>, 10<sup>3</sup>, 10<sup>4</sup>, etc. sondas como se ordenan en un molde de cromosoma lineal presente en la matriz. Dichas matrices que abarcan el genoma humano completo u otro genoma a una resolución subóptima pueden preferirse sobre matrices de alta resolución que cubren parte del mismo genoma por ejemplo en casos en los que se encuentran compañeros de translocación.

Preferentemente, la representación del genoma completo de una especie determinada a menor resolución se obtiene mediante sondas en la matriz que cada una representa un solo fragmento de restricción obtenido después de digestión con una enzima de restricción primaria. Preferentemente, esto se obtiene ignorando cada segunda, tercera, cuarta, quinta, sexta, séptima, octava, novena, décima, vigésima, trigésima, cuarentésima, quincuagésima, sexagésima, septuagésima, octogésima, nonagésima o centésima (por ejemplo segunda a centésima) etc. sonda que hibrida con el mismo fragmento de restricción.

Preferentemente, la representación del genoma completo de una especie determinada a menor resolución comprende sondas que se distribuyen igualmente a lo largo de los moldes lineales de cromosoma. Preferentemente, esto se obtiene ignorando una o más sondas en las regiones genómicas que muestran una densidad de sonda más elevada.

### HIBRIDACIÓN

El término “hibridación”, tal y como se utiliza en el presente documento, incluirá “el proceso mediante el cual una cadena de ácido nucleico se une con una cadena complementaria a través del emparejamiento de bases” así como el proceso de amplificación realizado, por ejemplo, con tecnología de reacción en cadena de la polimerasa (PCR).

Las secuencias de nucleótidos capaces de hibridación selectiva serán generalmente al menos 75 %, preferentemente al menos 85 o 90 % y más preferentemente al menos 95 % o 98 % homólogas a la secuencia de nucleótidos complementaria correspondiente sobre una región de al menos 20, preferentemente al menos 25 o 30, por ejemplo al menos 40, 60 o 100 o más nucleótidos contiguos.

“Hibridación específica” se refiere a la unión, formación de dúplex o hibridación de una molécula solo en una secuencia de nucleótidos particular en condiciones rigurosas (por ejemplo a 65 °C y 0,1xSSC {1xSSC = NaCl 0,15 M, Na-citrato 0,015 M pH 7,0}). Las condiciones rigurosas son condiciones en la que una sonda se hibridará con su secuencia diana, pero no con otras secuencias. Las condiciones rigurosas son dependientes de la secuencia y son diferentes en diferentes circunstancias. Las secuencias más largas se hibridan específicamente a temperaturas más altas. Generalmente, se seleccionan condiciones rigurosas para que sean aproximadamente 5 °C menores que el punto de fusión térmico (T<sub>m</sub>) para la secuencia específica a una fuerza iónica y a un pH definidos. La T<sub>m</sub> es la temperatura (a una fuerza iónica, pH y concentración de ácido nucleico definidos) a la cual el 50 % de las sondas complementarias a una secuencia diana se hibrida con la secuencia diana en el equilibrio. (Dado que las secuencias diana están generalmente presentes en exceso, a T<sub>m</sub>, el 50 % de las sondas estarán ocupadas en el equilibrio). Típicamente, las condiciones rigurosas incluyen una concentración salina de al menos aproximadamente de 0,01 a 1,0 M de concentración de iones Na (u otras sales) a un pH de 7,0 a 8,3 y la temperatura es de al menos aproximadamente 30 °C para sondas cortas. También pueden obtenerse condiciones rigurosas con la adición de agentes desestabilizantes - tales como formamida o sales de tetraalquilamonio.

Como entienden los expertos en la técnica, puede utilizarse una hibridación de rigurosidad máxima para identificar o detectar secuencias de nucleótidos idénticas mientras que puede utilizarse hibridación de rigurosidad intermedia (o baja) para identificar o detectar secuencias de polinucleótidos similares o relacionadas.

También se describen métodos para la hibridación de matrices de sondas con secuencias de nucleótidos marcadas o no marcadas. Las condiciones de reacción de hibridación particulares pueden controlarse para alterar la hibridación (por ejemplo, aumentar o disminuir la rigurosidad de unión de sonda/diana). Por ejemplo, temperatura de reacción, concentraciones de aniones y cationes, adición de detergentes y similares, todas pueden alterar las características de hibridación de sondas matriz y moléculas diana.

### FRECUENCIA DE INTERACCIÓN

La cuantificación de las frecuencias de ligamiento de los fragmentos de restricción proporciona una medida de sus frecuencias de entrecruzamiento. Adecuadamente, esto puede realizarse utilizando PCR tal y como se utiliza en tecnología 3C convencional como describe Splinter *et al.* (2004) (citados anteriormente). En resumen, la formación de productos de PCR puede medirse escaneando las intensidades de señal después de la separación en geles de agarosa teñidos con bromuro de etidio, utilizando un formador de imágenes Typhoon 9200 (Molecular Dynamics, Sunnyvale, CA). Adecuadamente, se utilizan diversos controles para corregir la interpretación de datos como también se describe en Splinter *et al.* (2004) (citados anteriormente).

Dado que la tecnología 4C descrita en el presente documento proporciona análisis de alto rendimiento de la frecuencia de interacción de dos o más secuencias de nucleótidos en el espacio nuclear, se prefiere que las frecuencias de ligamientos de los fragmentos de restricción se cuantifiquen utilizando las matrices descritas en el presente documento.

Para la cuantificación, las señales obtenidas de una muestra 4C pueden normalizarse con señales obtenidas de una muestra de control. La muestra de control 4C y la una o más muestras de control, se marcarán con marcadores diferentes y discernibles (por ejemplo colorantes) y se hibridarán simultáneamente con la matriz. La muestra o muestras de control típicamente contendrán todos los fragmentos de ADN (es decir todas las posibles secuencias de nucleótidos segundas que se han ligado con la primera secuencia de nucleótidos (diana)) en cantidades equimolares y, para excluir un sesgo en la eficiencia de hibridación, deben ser de tamaño similar a la segunda secuencia o secuencias de nucleótidos. Por tanto, el molde control típicamente contendrá ADN genómico (del mismo fondo genético al usado para obtener el molde 4C), digerido con tanto la enzima de restricción primaria como la secundaria y marcado por el mismo método (es decir, cebado al azar) que el molde 4C. Dicho molde de control hace posible



corregir las diferencias entre sondas en la eficacia de hibridación. La normalización de señales de matriz 4C para controlar señales de matriz hace posible expresar resultados en términos de enriquecimiento sobre eventos al azar.

5 El molde 4C marcado puede incluso hibridarse con una matriz con o sin una muestra de control diferencialmente marcada y con o sin uno o más moldes 4C distintos marcados diferencialmente. Otros moldes 4C pueden no estar relacionados con este molde 4C, por ejemplo puede obtenerse de diferente tejido y/u obtenerse con un conjunto diferente de cebadores de PCR inversos. Por ejemplo, el primer molde 4C puede ser material del paciente y el segundo molde 4C puede obtenerse a partir de un sujeto sano o de una muestra de control.

10 Dado que deben esperarse llamativos patrones de hibridación para los reordenamientos genéticos no siempre será necesario comparar sujetos enfermos con sujetos sanos. Por consiguiente, múltiples moldes 4C (por ejemplo dos o más), cada uno interrogando un diferente locus del mismo paciente o sujeto pueden hibridarse con una (por ejemplo una o más) matriz.

15 Los moldes 4C pueden marcarse diferencialmente (por ejemplo con hibridación de dos colores o multicolores) y/o pueden marcarse idénticamente en caso de que dichos locus normalmente se encuentren en diferentes cromosomas o en el mismo cromosoma a una distancia lo suficientemente lejos para que se produzca el solapamiento mínimo entre señales de interacción ADN-ADN. Como un ejemplo, el material de un sujeto con leucemia de linfocitos T puede procesarse para obtener moldes 4C para *TCR $\alpha/\delta$*  (marcado en un color para detectar translocaciones) y *MLL*, *TAL1*, *HOX11* y *LMO2* (cada uno marcado en el mismo segundo color para detectar otros reordenamientos genéticos). Estos cinco moldes 4C pueden hibridarse con una matriz, lo que permitirá el análisis simultáneo en locus múltiples para un reordenamiento genómico asociado a la enfermedad.

20 Para la cuantificación de frecuencias de interacción, las intensidades de señal absolutas o ratios sobre la muestra de control también pueden considerarse. Además, las señales de sondas adyacentes en el molde de cromosoma lineal pueden utilizarse para identificar la interacción de regiones cromosómicas. Dicha información posicional se analiza preferentemente ordenando las sondas en el molde de cromosoma lineal y analizando las intensidades de señal absolutas, o los ratios sobre señales de molde de control, mediante estrategias de ventana deslizante, utilizando, por ejemplo, estrategias de media móvil o mediana móvil.

25 La frecuencia de interacción de una o más secuencias de nucleótidos diana con una o más secuencias de nucleótidos de interés (por ejemplo uno o más locus genómicos) obtenida de acuerdo con los métodos descritos en el presente documento pueden utilizarse para reconstruir partes de mapas de cromosomas lineales o enteros e identificar reordenamientos genómicos equilibrados y no equilibrados que se producen dentro y entre cromosomas, en los que dichos reordenamientos son indicativos de un rasgo o enfermedad.

#### MÉTODO DE ENSAYO

30 En un aspecto adicional de la presente invención, se proporciona un método de ensayo para identificar uno o más agentes que modulan una interacción ADN-ADN.

Tal y como se utiliza en el presente documento, el término "modular" se refiere a prevenir, disminuir, suprimir, restaurar, elevar, aumentar o de otra manera afectar la interacción entre ADN-ADN.

45 En algunos casos, puede ser deseable evaluar dos o más agentes conjuntamente para su uso en la modulación de interacción ADN-ADN. En estos casos, los ensayos pueden modificarse fácilmente añadiendo dicho agente o agentes adicionales bien simultáneamente con, o posteriormente al primer agente.

50 El método de la presente invención también puede ser una exploración, mediante la cual diversos agentes se ensayan para la modulación de la actividad de la interacción ADN-ADN.

Se espera que los métodos de ensayo de la presente invención sean adecuados para exploración tanto a gran como a pequeña escala de agentes así como en ensayos cuantitativos.

55 Los usos médicos de dichos agentes terapéuticos se encuentran dentro del ámbito de la presente invención al igual que los programas de desarrollo de fármacos en sí mismos y composiciones farmacéuticas que comprenden dichos agentes. Un programa de desarrollo de fármacos puede, por ejemplo, implicar tomar un agente identificado o que puede identificarse por los métodos descritos en el presente documento, opcionalmente modificándolo (por ejemplo modificando su estructura y/o proporcionando una nueva composición que comprende dicha fracción) y realizando estudios adicionales (por ejemplo, estudios de toxicidad y/o estudios sobre la actividad, estructura o función).

60

Pueden realizarse ensayos clínicos en animales no humanos y pueden eventualmente realizarse en seres humanos. Dichos ensayos clínicos generalmente incluirán la determinación del efecto o efectos de diferentes niveles de dosificación. Los programas de desarrollo de fármacos pueden utilizar ordenadores para analizar fracciones identificadas por exploración (por ejemplo para predecir la estructura y/o función, para identificar posibles antagonistas o agonistas, para investigar otras fracciones distintas que puedan tener estructuras o funciones similares, etc.).

### ENSAYOS DIAGNÓSTICOS

Actualmente, diversos reordenamientos genómicos son difíciles de detectar mediante técnicas citogenéticas moleculares disponibles. Aunque la técnica de matriz de hibridación genómica comparativa (matriz-CGH, *genomic hybridization technique*) es una técnica recientemente desarrollada para la detección de amplificaciones y/o deleciones cromosómicas con una resolución de 35-300 Kb, esta técnica no es adecuada para detectar inversiones y translocaciones cromosómicas equilibradas. Por otro lado, el cariotipado espectral (SKY, *spectral karyotyping*) o cariotipado convencional, a menudo se realiza sobre material de pacientes para la detección de translocaciones cromosómicas, así como de cambios numéricos, aunque la resolución para definir puntos de corte de translocación es baja, normalmente de 10-50 Mb y 5-10 Mb, respectivamente. Por consiguiente, los resultados obtenidos por ambos métodos, y especialmente por SKY, conducirían a experimentos de validación lentos y laboriosos, tales como hibridación con fluorescencia *in situ* (FISH) y estrategias de clonación molecular con puntos de corte.

La tecnología 4C implica un procedimiento que puede detectar cualquiera de los reordenamientos cromosómicos basándose en frecuencias de interacción cambiadas entre secuencias de ADN físicamente unidas. La tecnología 4C es por tanto útil para la identificación de reordenamientos (recurrentes) cromosómicos para la mayoría de las neoplasias/malformaciones congénitas múltiples humanas o retraso mental. Una ventaja importante de la tecnología 4C es que permite un mapeo muy preciso del punto de corte en una región de solo varios miles de pares de bases. Otra ventaja de la tecnología 4C es que no se requiere un conocimiento previo sobre la posición exacta del punto de corte, dados que los puntos de corte serán detectables incluso cuando la secuencia cebo 4C se localice a 1-5 Mb del punto de corte. Esto tiene también la ventaja de que la misma secuencia cebo puede utilizarse para la detección de reordenamientos cromosómicos específicos que cubren grandes áreas de punto de corte. El mapeo preciso de los reordenamientos genómicos por tecnología 4C facilitará enormemente la identificación de uno o más genes expresados de un modo aberrante que subyacen a enfermedades o trastornos genéticos, que contribuirá de un modo importante a entender mejor las enfermedades de genotipo-fenotipo o trastornos genéticos, que contribuirán de un modo importante a un mejor entendimiento de correlaciones entre genotipo-fenotipo, ayudando a tomar decisiones de tratamiento y añadiendo una información sobre el pronóstico importante.

En una realización de la presente invención, para proporcionar una base para el diagnóstico o pronóstico de una enfermedad, se establecen valores normales o estándar de un sujeto. Esto puede realizarse mediante ensayos de muestras extraídas de sujetos normales-tales como animales o seres humanos. La frecuencia de la interacción ADN-ADN puede cuantificarse comparándola con una serie en dilución de controles positivos. Después, los valores estándar obtenidos de las muestras normales pueden compararse con los valores obtenidos de las muestras de sujetos afectados o posiblemente afectados por una enfermedad o un trastorno. La desviación entre los valores estándar y sujetos establece la presencia de la patología.

Dichos ensayos de diagnóstico pueden ajustarse para evaluar la eficacia de un régimen de tratamiento terapéutico particular y pueden utilizarse en estudios en animales, en ensayos clínicos o en el control del tratamiento de un paciente individual. Para proporcionar una base para el diagnóstico de enfermedades, puede establecerse un perfil normal o estándar para la interacción entre ADN-ADN. Los valores estándar obtenidos de muestras normales pueden compararse con valores obtenidos de muestras de sujetos posiblemente afectados por un trastorno o una enfermedad. La desviación entre los valores estándar y sujeto establece la presencia de la patología. Si la enfermedad se establece, un agente terapéutico existente puede administrarse, y puede generarse un perfil de tratamiento o valores. Finalmente, el método puede repetirse regularmente para evaluar si los valores progresan hacia o retornan a un patrón normal o estándar. Pueden utilizarse perfiles de tratamiento sucesivos para mostrar la eficacia del tratamiento durante un período de varios días o de varios meses.

La tecnología 4C detecta con precisión al menos 5 Mb de ADN genómico ligado en *cis* con la secuencia de nucleótidos que se analiza (véanse las Figuras 2-3 y 5). Ventajosamente, la tecnología 4C puede utilizarse para detectar cualquier aberración genómica que venga acompañada por un cambio en la separación del sitio genómico entre secuencias reordenadas y una secuencia 4C (cebo) de elección. Dicho cambio puede ser, por ejemplo, un aumento o disminución en la separación del sitio genómico o puede ser una infrarrepresentación (como ocurre en las deleciones) o una sobrerrepresentación (como ocurre en las duplicaciones) de secuencias proximales (por ejemplo de hasta o mayores de 15 Mb) con respecto a la secuencia 4C (cebo). Típicamente, dichas aberraciones o reordenamientos genómicos son una causa de o están asociadas a enfermedades - tales como cáncer (por ejemplo leucemia) y otras enfermedades genéticas o congénitas como se describe en el presente documento.

Las aberraciones genéticas (por ejemplo aberraciones genómicas o cromosómicas - tales como aberraciones genómicas o cromosómicas equilibradas y/o no equilibradas) incluyen, pero sin limitación, reordenamientos,

translocaciones, inversiones, inserciones, deleciones y otras mutaciones de ácido nucleico (por ejemplo cromosomas) y también pérdida o ganancia de parte o cromosomas completos. Son una causa principal de trastornos o enfermedades genéticas, incluyendo trastornos congénitos y enfermedades adquiridas - tales como neoplasias. En muchos reordenamientos, están implicados dos cromosomas diferentes. En este sentido, genes (o fragmentos de genes) se retiran del contexto fisiológico normal de un cromosoma particular y se localizan en un cromosoma receptor, adyacente a genes o a fragmentos no relacionados de genes (a menudo oncogenes o protooncogenes).

Las neoplasias pueden incluir leucemias agudas, linfomas malignos y tumores sólidos. Son ejemplos no limitantes de alteraciones t(14; 18) que se produce frecuentemente en NHL; t(12; 21) que se encuentra frecuentemente en ALL-B precursor de infancia; y la presencia de aberraciones de 11q23 (gen MLL (leucemia de linaje mixto o leucemia linfóide mieloide)) en leucemias agudas.

El gen MLL en la región cromosómica 11q23 está implicado en diversas translocaciones tanto en ALL como en leucemias mieloides agudas (AML). Hasta ahora, se han identificado al menos diez genes compañeros. Algunas de estas translocaciones, - tales como t(4; 11) (q21; q23), t(11; 19) (q23; p13) y t(1; 11) (p32; q23), predominantemente se producen en ALL, mientras que otras, tales como t(1;11) (q21;q23), t(2;11) (p21;q23), t(6;11) (q27;q23) y t(9;11) (p22;q23) se observan más frecuentemente en AML. Los reordenamientos que implican región 11q23 ocurren muy frecuentemente en leucemias agudas infantiles (aproximadamente 60-70 %), y a un grado mucho más bajo en leucemias juveniles y adultas (cada una aproximadamente 5 %).

Los reordenamientos en neoplasias linfoides a menudo implican genes Ig o TCR. Los ejemplos incluyen los tres tipos de translocaciones (t(8; 14), t(2; 8) y t(8; 22)) que se encuentran en linfomas de Burkitt, en los que el gen MYC está acoplado a segmentos génicos de la cadena pesada de Ig (IGH), kappa Ig (IGK), o lambda Ig (IGL), respectivamente. Otro tipo común de translocación en esta categoría es t(14; 18) (q32; q21) que se observa en aproximadamente el 90 % de los linfomas foliculares, uno de los tipos NHL principales. En esa translocación el gen BCL2 se reordena en regiones dentro del locus IGH dentro o adyacente a los segmentos génicos JH. El resultado de esta aberración cromosómica es la sobreexpresión de la proteína BCL2, que desempeña una función como un factor de supervivencia en el control del crecimiento inhibiendo la muerte celular programada.

El gen BCL2 consta de tres exones, pero estos se dispersan sobre una gran área. De estos el último exón codifica una región no traducida 3' (UTR 3') grande. La UTR 3' es una de las dos regiones en las que muchos puntos de corte t(14; 18) se agrupan y se denomina la "región de punto de corte principal"; la otra región de punto de corte implicada en translocaciones t(14; 18), está localizada 20-30 kb cadena abajo del locus BCL2 y se denomina la "región de agrupación menor". Una tercera área de punto de corte BCL2, la VCR (región de grupo variante), se localiza en el lado 5' del locus BCL2 y está entre otras implicadas en translocaciones variantes, es decir t(2; 18) y t(18; 22) y en los que los segmentos génicos IGK e IGL son los genes homólogos.

Por tanto, como ejemplo, la tecnología 4C puede aplicarse para la exploración de material del paciente de aberraciones genéticas cerca o en locus que se seleccionaron basándose en su asociación frecuente con un fenotipo clínico determinado. Otros ejemplos no limitantes de dichos locus son *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, locus de la inmunoglobulina, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCR $\alpha$ / $\delta$* , *TCR $\beta$* , *HOX* y otros locus en diversas leucemias linfoblásticas.

Ventajosamente, si se sospecha una aberración genética, la tecnología 4C puede aplicarse como la primera y única exploración para verificar y mapear la presencia de la aberración como se explica en el presente documento.

#### *Detección de reordenamientos genómicos*

En una realización particularmente preferida de la presente invención, los métodos descritos en el presente documento pueden utilizarse para la detección de reordenamientos genómicos.

Actualmente, los reordenamientos genómicos - tales como puntos de corte de translocación - son muy difíciles de detectar. Por ejemplo, las micromatrices de hibridación genómica (CGH) comparativas pueden detectar diversos tipos de reordenamientos pero no pueden detectar translocaciones. Si se sospecha translocación en un paciente pero no se conocen los compañeros de cromosomas, puede realizarse cariotipado espectral (SKY) para encontrar los compañeros de translocación y obtener un cálculo estimado aproximado de localizaciones en puntos de corte. Sin embargo, la resolución es muy mala (normalmente no mayor de ~50 Mb) y se requiere normalmente mapeo fino adicional (que es tanto laborioso como costoso). Esto se realiza normalmente utilizando Hibridación con Fluorescencia In Situ (FISH) que de nuevo proporciona una resolución limitada. Utilizando FISH, pueden localizarse puntos de corte en una región de +/- 50 kb a resolución máxima.

Las frecuencias de interacción ADN-ADN principalmente son una función de la separación del sitio genómico, es decir, frecuencias de interacción ADN-ADN son inversamente proporcionales a la distancia lineal (en kilobases) entre dos locus de ADN presentes en el mismo molde de ADN físico (Dekker *et al.*, 2002). Por tanto, una translocación, que crea uno o más moldes de ADN físicos nuevos, se consigue alterando interacciones ADN-ADN

cerca de los puntos de corte, y esto puede medirse por tecnología 4C. Las enfermedades basadas en translocaciones están típicamente causadas por interacciones entre ADN-ADN aberrantes, ya que la translocación es el resultado del ligamiento físico (interacción) de brazos de cromosomas rotos (ADN).

- 5 Por consiguiente, para la detección de translocaciones, la tecnología 4C puede utilizarse para identificar aquellas interacciones ADN-ADN que son diferentes entre sujetos enfermos y no enfermos.

10 Como ejemplo, la tecnología 4C puede aplicarse para la exploración de material del paciente para translocaciones cercanas a locus que se seleccionan basándose en su asociación frecuente con un fenotipo clínico determinado como se describe en el presente documento.

15 Si se sospecha translocación en un paciente pero se desconocen los compañeros de cromosomas, puede realizarse un mapeo inicial utilizando métodos disponibles actualmente como cariotipado espectral (SKY). Esto puede identificar los compañeros de translocación y proporcionar un cálculo muy aproximado de localizaciones de puntos de corte (normalmente no mayor de ~50 Mb de resolución). La tecnología 4C puede después aplicarse, utilizando secuencias "cebo" en esta región localizada por ejemplo en cada 2 Mb, 5 Mb, 10 Mb, 20 Mb (u otros intervalos como se describe en el presente documento), para mapear con precisión el punto de corte e identificar, por ejemplo, el gen o los genes que no se expresan como consecuencia de la translocación.

20 Típicamente se identificará una translocación mediante una transición abrupta de frecuencias de interacción bajas a altas en un cromosoma distinto de uno que contiene la secuencia cebo 4C o en cualquier parte del mismo cromosoma.

25 En una realización preferida, la muestra del sujeto está en un estado premaligno.  
En una realización preferida, la muestra del sujeto consiste en amniocitos cultivados o no cultivados obtenidos por amniocentesis para el diagnóstico prenatal.

30 En un diseño de matriz preferido, las sondas presentes en una sola matriz representan el genoma completo de una especie determinada a una resolución máxima. Por tanto, las matrices para detectar translocaciones y similares mediante tecnología 4C contienen sondas como se describe en el presente documento complementarias en cada lado de cada sitio de reconocimiento de enzima de restricción primaria en el genoma de una especie determinada (por ejemplo ser humano).

35 En otro diseño preferido, las sondas presentes en una sola matriz representan el genoma completo de una especie determinada, pero no a una máxima resolución. Por tanto, las matrices para detectar translocaciones y similares por tecnología 4C contienen sondas como se describe en el presente documento que son complementarias a solo un lado de cada sitio de reconocimiento de enzima de restricción primaria en el genoma de una especie determinada (por ejemplo ser humano).

40 En otro diseño preferido, las sondas presentes en una sola matriz representan el genoma completo de una especie determinada, pero no a resolución máxima. Por tanto, las matrices para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos mediante tecnología 4C contienen sondas como se describe en el presente documento que son complementarias en un lado de cada otro sitio de reconocimiento de enzima de restricción primaria como se ordena a lo largo del molde lineal del genoma de una especie determinada (por ejemplo ser humano).

45 Por tanto, las matrices para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C contienen sondas como se describe en el presente documento que representan cada una un solo fragmento de restricción obtenido después de digestión con una enzima de restricción primaria. Preferentemente, esto se obtiene ignorando cada segunda, tercera, cuarta, quinta, sexta, séptima, octava, novena, décima, vigésima, trigésima, cuadragésima, quincuagésima, sexagésima, septuagésima, octogésima, nonagésima o centésima sonda que hibrida con el mismo fragmento de restricción. Las matrices para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C pueden contener sondas como se describe en el presente documento, que están distribuidas igualmente a lo largo de los moldes de cromosoma lineal. Preferentemente, esto se obtiene ignorando una o más sondas en aquellas regiones genómicas que muestran una mayor densidad de sonda.

50 En otro diseño preferido, las sondas presentes en una sola matriz representan el genoma completo de una especie determinada, pero no a resolución máxima. Por tanto, las matrices para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C contienen sondas como se describe en el presente documento complementarias en un lado de cada tercero, cuarto, quinto, sexto, séptimo, octavo, noveno, décimo, vigésimo, trigésimo, cuadragésimo, quincuagésimo, sexagésimo, septuagésimo, octogésimo, nonagésimo o centésimo, etc., sitio de reconocimiento de enzima de restricción primaria como se ordena a lo largo del molde lineal del genoma de una especie determinada (por ejemplo ser humano). Las matrices para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C pueden contener sondas como se describe en el presente documento, que representan el genoma

completo, pero con una sola sonda cada 100 kilobases. Las matrices para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C, pueden contener sondas, como se describe en el presente documento, que representan cada sitio de reconocimiento de enzima de restricción primaria en el genoma que puede representarse por una secuencia de sonda exclusiva.

5 En otro diseño de matriz preferido, las sondas como se describen en el presente documento en una sola matriz representan regiones genómicas de un tamaño determinado, tal como en aproximadamente 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb o 10 Mb (por ejemplo, de aproximadamente 50 kb-10 Mb) alrededor de todos los locus que se sabe que están implicados en translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos.

15 En otro diseño de matriz preferido, las sondas como se describen en el presente documento en una matriz sencilla representan regiones genómicas de un tamaño determinado - tal como aproximadamente 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb o 10 Mb (por ejemplo de aproximadamente 50 kb-10 Mb) alrededor de una selección de locus que se sabe que están implicados en translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos. Las selecciones pueden realizarse basándose en criterios fundamentados, por ejemplo, pueden representar únicamente los locus que están implicados en un tipo de enfermedad determinada.

20 En otro diseño de matriz preferido, las sondas como se describe en el presente documento en una sola matriz representan una región genómica de interés de, por ejemplo, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 600 kb, 700 kb, 800 kb, 900 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb, 10 Mb, 20 Mb, 70 Mb, 80 Mb, 90 Mb o 100 Mb (por ejemplo 100 kb-10 Mb) (parte de) un cromosoma o cromosomas múltiples, representándose cada sonda múltiples (por ejemplo 10, 100, 1000) veces para representar mediciones cuantitativas de intensidades de señal de hibridación en cada secuencia de sonda.

25 En un diseño experimental preferido, la secuencia 4C (cebo) está dentro de aproximadamente 0 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb, 10 Mb, 11 Mb, 12 Mb, 13 Mb, 14 Mb o 15 Mb (por ejemplo de aproximadamente 0-15 Mb) o más de la secuencia reordenada real (es decir punto de corte en caso de una translocación).

30 En una hibridación preferida, dos moldes 4C diferencialmente marcados obtenidos con una secuencia (cebo 4C) de un sujeto enfermo y no enfermo se hibridan simultáneamente en la misma matriz. Las diferencias en las interacciones de ADN-ADN permiten la detección del punto de corte en *cis* (en el mismo cromosoma que el cebo 4C) y en *trans* (en el compañero de translocación).

35 En una hibridación preferida, se obtienen moldes 4C múltiples marcados diferencialmente con una secuencia (cebo 4C) de sujetos enfermos o no enfermos hibridados simultáneamente con la misma matriz. Las diferencias en las interacciones ADN-ADN permiten la detección del punto de corte en *cis* (en el mismo cromosoma que el cebo 4C) y en *trans* (en el compañero de translocación).

40 Ventajasamente, puede utilizarse el análisis multicolor, en lugar de color dual, en micromatrices permitiendo la hibridación simultánea de más de dos muestras en una sola matriz. Por consiguiente, en la tecnología 4C puede utilizarse hibridación multicolor.

45 En una hibridación preferida, los moldes 4C múltiples marcados diferencialmente obtenidos con una secuencia (cebo 4C) de sujetos enfermos y un molde 4C marcado diferencialmente de un sujeto no enfermo se hibridan simultáneamente en la misma matriz. Las diferencias en las interacciones ADN-ADN permiten la detección del punto de corte en *cis* (en el mismo cromosoma que el cebo 4C) y en *trans* (en el compañero de translocación).

50 En otra hibridación preferida, dos moldes 4C diferencialmente marcados del mismo sujeto no enfermo obtenido con dos secuencias diferentes (cebos 4C) que representan cada uno otro posible compañero de translocación, se hibridan simultáneamente con la misma matriz. Los grupos de fuertes señales de hibridación observados en el molde lineal de cromosomas no relacionados con respecto al cromosoma que lleva la secuencia de interés (cebo 4C) identificará el cromosoma compañero de translocación y el punto de corte en el compañero de translocación.

55 En otra hibridación preferida, los moldes 4C múltiples marcados diferencialmente del mismo sujeto no enfermo, obtenido con secuencias diferentes múltiples (cebos 4C) que representan cada uno otro posible compañero de translocación, se hibridan simultáneamente con la misma matriz. Grupos de señales de hibridación fuertes observadas en el molde lineal de cromosomas no relacionados con el cromosoma que lleva la secuencia de interés (cebo 4C) identificarán el cromosoma compañero de translocación y su punto de corte para la secuencia de interés.

60 El material utilizado para la detección de translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C puede obtenerse por entrecruzamiento (y procesamiento adicional como se describe) de células vivas y/o células muertas y/o lisados nucleares y/o cromatina aislada etc. (como se describe en el presente documento) de sujetos enfermos y/o no enfermos.

65

*Detección de inversiones*

Las inversiones (por ejemplo inversiones equilibradas) no pueden detectarse por métodos - tales como técnicas de hibridación genómica comparativa - pero pueden detectarse por tecnología 4C particularmente cuando la inversión (equilibrada) está próxima a la secuencia 4C (por ejemplo hasta aproximadamente 1-15 Mb o más) (cebo).

La detección de inversiones (equilibradas) se basa en la identificación de aquellas interacciones de ADN-ADN que fueron diferentes entre diferentes sujetos enfermos y no enfermos. Las inversiones cambiarán la posición relativa (en kilobases) del molde de ADN físico de todas las secuencias (excepto las localizadas más centralmente) de la región reorganizada medida contra una secuencia cercana en el mismo cromosoma que se toma como secuencia 4C (cebo). Dado que las frecuencias de interacción ADN-ADN están inversamente relacionadas con la separación de sitios genómicos los sujetos enfermos tendrán patrones invertidos de intensidades de hibridación en todas las sondas localizadas en la región genómica reordenada, en comparación con un sujeto no enfermo. Por tanto, la tecnología 4C permite la identificación de posiciones y tamaños de inversiones (equilibradas).

De acuerdo con este aspecto de la presente invención, un diseño de matriz dedicado preferido comprende sondas en una sola matriz que representa regiones genómicas de un tamaño determinado - tal como de aproximadamente 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb o 10 Mb) (por ejemplo 50 kb-10 Mb) alrededor del locus en el que se sospecha la inversión u otro reordenamiento.

En otro diseño de matriz dedicado preferido, las sondas en una matriz sencilla representan regiones genómicas de un tamaño determinado (50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, etc.) alrededor del locus en el que se sospecha la inversión u otra reordenamiento. Para un análisis cuantitativo fiable de intensidades de señales la cantidad de sondas presentes en la matriz es típicamente un gran exceso con respecto a la cantidad de fragmentos afines que se hibridan con la matriz. Por lo tanto, puede ser necesario tener cada sonda presente múltiples veces (por ejemplo 10, 20, 50, 100, 1000 veces, etc.) en la matriz. Además, puede ser necesario titular la cantidad de molde que está hibridado con la matriz.

*Detección de deleciones*

La detección de deleciones se basa en la identificación de aquellas interacciones ADN-ADN que eran diferentes entre sujetos enfermos y no enfermos. Las deleciones se producirán en ausencia de interacciones de ADN con una secuencia 4C (cebo) localizada cerca de (por ejemplo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 Mb o más) de la región delecionada. Esto puede dar como resultado la ausencia completa de señales de hibridación en todas las sondas localizadas en la región reordenada si la deleción está presente en ambos alelos (homocigotos), o una reducción para sujetos enfermos frente a no enfermos de intensidades de señal si la deleción está presente en únicamente un alelo (heterocigoto). La deleción lleva más secuencias distales a una proximidad más estrecha en el molde de ADN físico en comparación con la secuencia 4C analizada (cebo) que dará como resultado señales de hibridación más fuertes para sondas localizadas directamente más allá de la región delecionada.

*Detección de duplicación o duplicaciones*

La detección de duplicación está típicamente basada en la identificación de aquellas interacciones ADN-ADN que son diferentes entre sujetos enfermos y no enfermos. Las sondas en la región duplicada mostrarán unas señales de hibridación aumentadas con una secuencia 4C (cebo) localizada cerca (por ejemplo aproximadamente 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 o 15 Mb o más) la región reorganizada en comparación con señales de un sujeto de control no enfermo. Las sondas más allá de la región duplicada están lo suficientemente alejadas de la secuencia 4C y por consiguiente mostrarán señales de hibridación disminuidas en comparación con señales de un sujeto no enfermo de control.

Preferentemente, un aumento o una disminución de la frecuencia de interacción ADN-ADN para la muestra del sujeto en comparación con el control, indica una duplicación o inserción.

Preferentemente, un aumento en la frecuencia de interacción ADN-ADN para la muestra del sujeto en comparación con el control y/o una reducción en la frecuencia de interacción ADN-ADN para las regiones distantes indica una duplicación o inserción.

*Diagnóstico prenatal*

Ventajosamente, la tecnología 4C también puede utilizarse en diagnóstico prenatal.

Los ácidos nucleicos pueden obtenerse de un feto utilizando diversos métodos conocidos en la técnica. Como ejemplo, puede utilizarse amniocentesis para obtener líquido amniótico del cual las células fetales en suspensión se extraen y se cultivan durante diversos días (Mercier y Bresson (1995) Ann. Gnt., 38, 151-157). El ácido nucleico de las células puede después extraerse. El conjunto de vellos coriónicos puede posibilitar prescindir de la etapa de

cultivo y evita la colección de líquido amniótico. Estas técnicas pueden aplicarse anteriormente (hasta 7 semanas de gestación para la recogida de vello coriónico y 13-14 semanas para amniocentesis), pero con un riesgo de aborto ligeramente aumentado.

5 Una recogida directa de sangre fetal a nivel del cordón umbilical también puede utilizarse para obtener ácido nucleico, pero típicamente requiere un conjunto de médicos clínicos especializados en esta técnica (Donner *et al.*, (1996) Fetal Diagn. Ther., 10, 192-199).

10 Ventajosamente, las aberraciones genéticas (por ejemplo aberraciones genéticas o cromosómicas) - tales como reordenamientos, translocaciones, inversiones, inserciones, deleciones y otras mutaciones en cromosomas y ácido nucleico - pueden detectarse en esta fase.

15 Preferentemente, aberraciones genéticas (por ejemplo aberraciones genómicas o cromosómicas) - tales como reordenamientos, translocaciones, inversiones, inserciones, deleciones y otras mutaciones en los cromosomas 21, 18, 13, X o Y y también pérdidas o ganancias de parte o todos los cromosomas 21, 18, 13, X o Y pueden detectarse dado que estos son los cromosomas en los cuales se producen la mayoría de las aberraciones en el feto.

#### *Determinación de sitios de integración genómicos*

20 La tecnología 4C también permite la determinación de sitios de integración genómicos de virus y transgenes, etc., también cuando copias múltiples se insertan en diferentes posiciones en el genoma (como se describe en la Figura 3).

#### *Determinación de la predisposición para adquirir determinadas translocaciones*

25 Ventajosamente, la tecnología 4C también puede aplicarse a sujetos no enfermos para medir el entorno genómico de locus frecuentemente implicados en aberraciones genéticas. De esta manera, es posible determinar la predisposición del sujeto para adquirir determinadas aberraciones genéticas.

30 Por tanto, además del uso médico descrito en el presente documento, la presente invención puede utilizarse en diagnóstico.

#### MÚLTIPLE 4C

35 La presente invención permite el análisis simultáneo de la frecuencia de interacciones de secuencias de nucleótidos diana múltiples con una o más secuencias de nucleótidos de interés. La amplificación puede realizarse utilizando PCR múltiple. Dichos métodos permiten una exploración más sesgada para reordenamientos genómicos equilibrados y no equilibrados, es decir, translocaciones, inversiones, deleciones y duplicaciones que pueden haberse producido en cualquier parte del genoma. Los métodos pueden utilizarse para identificar puntos de corte de reordenamientos a una resolución muy alta, típicamente dentro de veinte kilobases (sobre un promedio de 5 kb). El método puede utilizarse en aplicaciones de diagnóstico como se expone anteriormente, por ejemplo, para diagnóstico prenatal, diagnóstico postnatal y para el análisis de muestras de tumor y otras muestras enfermas para identificar reordenamientos genómicos que subyacen a enfermedades o predisposición de sujetos a enfermedades. Las secuencias de nucleótidos amplificadas de interés pueden analizarse en micromatrices 4C a medida (como se ha descrito anteriormente), o en micromatrices en mosaico (*tiling*) genómicas o por secuenciación como se describe adicionalmente en el presente documento.

50 La exploración simultánea de reordenamientos en locus múltiples que se sabe que está asociado con una enfermedad.

55 Para muchas enfermedades, síndromes o fenotipos, se conocen alteraciones de ADN causantes posibles múltiples y los sujetos requieren explorarse para identificar el reordenamiento exacto subyacente de la enfermedad/síndrome/fenotipo. Por ejemplo, en el caso de las leucemias linfoblásticas pueden aplicarse reordenamientos que implican AML1, MLL, MYC, BCL, BCR, ABL, locus de inmunoglobulina, LYL1, TAL1, TAL2, TCR $\alpha/\delta$ , TCR $\beta$ , HOX y posiblemente otros locus frecuentemente subyacentes en la enfermedad; la tecnología 4C puede aplicarse para identificar qué locus y reordenamientos están implicadas en un paciente con la enfermedad determinada.

60 En esta realización, la tecnología 4C se refiere a cada uno de los locus de interés. Cada locus puede analizarse por separado, pero también pueden analizarse locus múltiples simultáneamente en una sola matriz. Por tanto, una exploración 4C puede implicar la amplificación por PCR (inversa) de elementos de ADN que interaccionan con una o más secuencias de nucleótidos diana cerca de cada uno de los locus que requieren analizarse para los reordenamientos. Las secuencias diana en o cerca de estos locus se seleccionan basándose en los criterios mencionados en el presente documento.

65

La amplificación de secuencias de nucleótidos de interés puede realizarse por separado para cada una de las secuencias de nucleótidos diana o puede realizarse simultáneamente en un volumen de reacción por PCR múltiple (inversa). Lo último se prefiere en casos cuando las secuencias de nucleótidos de interés interactúan con diferentes secuencias de nucleótidos diana que pueden marcarse idénticamente sin comprometer los análisis. Esto puede ser el caso por ejemplo cuando los locus se localizan en diferentes cromosomas o cuando los locus se localizan en el mismo cromosoma a una distancia lo suficientemente lejana para que se produzca un solapamiento mínimo entre las señales de interacción ADN-ADN o en casos cuando el solapamiento entre interacciones de ADN-ADN y se amplifica de diferentes secuencias de nucleótidos diana no interfiere con la detección de reordenamientos genómicos.

La amplificación de secuencias de nucleótidos de interés que interactúan con diferentes secuencias de nucleótidos diana se realiza preferentemente por separado cuando cada uno de los conjuntos de interacciones ADN-ADN requiere marcarse diferentemente o cuando los conjuntos de cebadores de PCR inversos interfieren con su eficiencia de amplificación entre sí.

Las secuencias de interés marcadas de manera diferente o idéntica que interactúan con las diversas secuencias de nucleótidos diana se hibridan con una sola micromatriz o múltiples micromatrices que contienen sondas que representan el genoma completo (es decir matrices de alta densidad o matrices en mosaico) o una parte seleccionada del genoma, como se ha descrito anteriormente. Las señales de hibridación se compararán con las obtenidas con una muestra de control, cuando un aumento o una disminución en las frecuencias de interacción ADN-ADN medidas en la muestra de ensayo frente a la muestra de control es indicativa para un reordenamiento de ADN en la muestra de ensayo.

*Una exploración no sesgada de todo el genoma para reordenamientos en posiciones desconocidas en el genoma.*

En una segunda realización, la tecnología 4C se aplica para una exploración de todo el genoma no sesgada para identificar reordenamientos en una muestra de un sujeto enfermo (o no enfermo), cuando los reordenamientos se desconocen previamente y/o se producen en localizaciones desconocidas. En esta realización, las secuencias diana no pueden seleccionarse para que estén próximas al reordenamiento. Las secuencias diana por tanto son desconocidas o se sospecha que están asociadas con la enfermedad. En lugar de esto, se seleccionan para distribuirse a lo largo del genoma seleccionado o sección del genoma, para proporcionar una cobertura suficiente de la cromatina a investigar. Preferentemente, se abarca todo el genoma.

Por ejemplo, las secuencias diana se seleccionan de tal manera que sus secuencias de interacción de interés, sean principalmente las secuencias directamente circundantes de la región diana, es decir, dentro de 5, 10, 15, 20, 25, 30, 35, 40 o 45 a 50 Mb desde la secuencia diana) que abarcan la parte completa o una parte sustancial del genoma o un cromosoma o una parte de un cromosoma. Esto permitirá reconstruir mapas de cromosomas físicos presentes en cualquier sujeto o tipo de célula.

Para que dos secuencias diana adyacentes en el molde de cromosoma lineal tengan secuencias de ADN que interactúan solapando (es decir solapamiento de entornos genómicos) es necesario que estén, por ejemplo, a una distancia de 5, 10, 15, 20, 25, 30, 35, 40, 45 o 50 a 100 Mb entre sí. Por lo tanto, para abarcar todo el genoma (2-3 x 10<sup>9</sup> pb) con secuencias diana separadas ~10 Mb, se requieren 200-300 secuencias diana. Para cubrir el genoma completo con secuencias diana separadas ~50 Mb, se requieren 40-60 secuencias diana. En una realización última, únicamente una secuencia diana por cromosoma se requiere para cubrir el genoma completo.

Todas las secuencias de nucleótidos de interés que interactúan con las secuencias diana se amplifican como en 4C, bien conjuntamente en una sola mezcla de reacción (o en un número limitado de mezclas de reacción) por PCR múltiple (inversa), o en reacciones de PCR distintas (inversas) que pueden agruparse posteriormente. La PCR múltiple es preferida en casos cuando las secuencias nucleotídicas de interés interactúan con diferentes secuencias nucleotídicas diana que pueden marcarse idénticamente sin comprometer los análisis. La amplificación se realizará independientemente para cada secuencia nucleotídica diana cuando cada conjunto de interacciones ADN-ADN que pertenecen a una secuencia diana determinada requiere marcarse diferentemente y/o cuando los conjuntos de cebadores de PCR inversa interfieren con cada eficiencia de amplificación distinta.

*Identificación de reordenamientos intracromosómicos tales como deleciones, inserciones y duplicaciones e inversiones (equilibradas y no equilibradas).*

En una realización preferida, todas las secuencias de ADN amplificadas que interactúan con el conjunto de secuencias diana de una muestra de ensayo (por ejemplo, una muestra de paciente) se marcan idénticamente y las frecuencias de interacción de todo el genoma se comparan con las de una muestra de control (por ejemplo de un sujeto sano). Las muestras de control y de ensayo pueden hibridarse con la misma matriz en diferentes colores, o pueden hibridarse con diferentes matrices y compararse. Un aumento o una disminución en la frecuencia de interacción ADN-ADN para la muestra de ensayo en comparación con la muestra de control es indicativa de una duplicación/inserción o de una deleción en la muestra de ensayo. Esto también puede ser indicativo de una inversión.



Las Figuras 18 y 19 muestran una delección identificada por 4C de esta manera.

En otra realización preferida, las secuencias de ADN amplificadas que interaccionan con el conjunto de secuencias diana de una muestra de ensayo (por ejemplo una muestra de paciente) se marcan en dos colores, alternando los colores para secuencias diana que están adyacentes en el molde del cromosoma lineal. Las secuencias diana adyacentes están lo suficientemente cercanas en el molde de cromosoma lineal para que sus secuencias de interacción se solapen. Por tanto, las secuencias diana adyacentes pueden estar, por ejemplo, separadas 5, 10, 15, 20, 25, 30, 35, 40, 45 a 50 Mb entre sí. Dicho diseño experimental impide que los reordenamientos se pierdan debido a señales de hibridación saturadas cercanas a una secuencia diana determinada. Las frecuencias de interacción de todo el genoma se comparan con las de una muestra de control (por ejemplo de un sujeto sano). Las muestras de control y de ensayo pueden marcarse diferencialmente e hibridarse con la misma matriz, o pueden hibridarse con diferentes matrices y compararse. Un aumento o una disminución en la frecuencia de interacción ADN-ADN para la muestra de ensayo en comparación con la muestra control es indicativa de una duplicación/inserción o de una delección en la muestra de ensayo. También puede ser indicativa de una inversión.

Un análisis 4C posterior dirigido a secuencias diana que flanquean o se encuentran dentro de la parte reorganizada del genoma puede realizarse para identificar inversiones. Un patrón invertido de señales de hibridación en comparación con una muestra de control identifica la inversión en la muestra de ensayo. Esto se representa en la Figura 20.

En una realización adicional preferida, las secuencias de nucleótidos de interés que interaccionan con diferentes secuencias diana yuxtapuestas en el molde del cromosoma se marcan de diferentes tipos. Las reordenamientos se detectan por la aparición o desaparición de señales de ADN de interacción en la muestra del paciente en comparación con la muestra de control. Esto se representa en las Figuras 21 a 23.

*Identificación de translocaciones equilibradas y no equilibradas.*

En una realización preferida adicional, se disponen colorantes múltiples (por ejemplo 48 colorantes) y cada cromosoma se marca con dos colorantes únicos que se usan de tal manera que los colorantes se alternan entre secuencias diana que son adyacentes al molde del cromosoma lineal. Todos los fragmentos de ADN pueden hibridarse entre sí con una matriz que contiene sondas que representan el genoma completo. La identificación de interacciones ADN-ADN entre cromosomas que se producen en las muestras de ensayo pero no en las muestras de control son indicativas de una translocación e identifican los dos cromosomas reorganizados. La transición de señales bajas a altas en un cromosoma no relacionado identifica el sitio de reconocimiento de enzima de restricción primaria cercano al punto de corte del ADN.

Si las translocaciones están equilibradas, cada uno de los dos colorantes específicos del cromosoma debe dar señales de hibridación fuertes en un conjunto de sondas mutuamente exclusivo directamente adyacentes entre sí en el molde lineal del cromosoma no relacionado. El punto de corte en este cromosoma no relacionado se localiza entre los dos conjuntos de sondas que muestran señales de hibridación positivas. Véase la Figura 24.

Cuando se ensaya sobre una muestra del paciente que implica una translocación del cromosoma 1:cromosoma 7 se obtiene el resultado mostrado en la Figura 25. En esta muestra particular las secuencias diana en la matriz representan una selección de secuencias localizadas cerca de los sitios HindIII (véase anteriormente para la derivación de secuencias diana).

Si las translocaciones se producen con pérdida de ADN en los puntos de corte (es decir translocaciones no equilibradas), cada uno de los dos colorantes específicos de cromosoma dará señales de hibridación fuertes en un conjunto de sondas mutuamente exclusivos en el cromosoma no relacionado que no están directamente adyacentes entre sí en el molde lineal del cromosoma no relacionado. Las sondas localizadas en el cromosoma no relacionado entre los dos conjuntos de sondas que muestran interacciones ADN-ADN intercromosómicas representan la región genómica que se perdió. Las sondas más externas de esta región marcan los sitios de transcripción cerca de los puntos de corte del cromosoma no relacionado (véase la Figura 26).

Cuando se ensaya en una muestra del paciente que implica una translocación del cromosoma 4:cromosoma 7 con una cantidad desconocida de ADN delecionado en el punto de corte se obtienen los resultados mostrados en la Figura 27. En esta muestra particular las secuencias diana en la matriz de nuevo representan una selección de secuencias localizadas cerca de los sitios HindIII. Los resultados muestran que tanto la translocación como la delección (aprox. 2 Mb) se detectan.

En una realización aún más preferida, se dispone de colorantes múltiples (por ejemplo 24 colorantes) y se usan para marcar diferencialmente cada cromosoma (Figura 28). Por tanto, todas las secuencias de ADN que interaccionan con secuencias diana presentes en el mismo cromosoma se marcan de modo idéntico y diferente de las localizadas en otros cromosomas. La identificación de interacciones ADN-ADN entre cromosomas que se producen en muestras de ensayo pero no muestras de control son indicativas de translocación e identifican los dos cromosomas reordenados.

Un análisis 4C posterior dirigido específicamente hacia los cromosomas implicados en la translocación permite la identificación de los puntos de corte. En este caso, uno o cada cromosoma está marcado con dos únicos colorantes que se usan de tal manera que alternan entre fragmentos de ADN que interactúan con secuencias diana que son adyacentes al molde de cromosoma lineal.

5 Si las translocaciones están equilibradas, cada uno de los dos colorantes específicos del cromosoma deben dar señales de hibridación fuertes en un conjunto de sondas mutuamente exclusivo directamente adyacentes entre sí en el molde lineal del cromosoma no relacionado. El punto de corte en este cromosoma no relacionado se localiza entre los dos conjuntos de sondas que muestran señales de hibridación positivas.

10 Si las translocaciones se producen con pérdida de ADN en los puntos de corte (es decir translocaciones no equilibradas), cada uno de los dos colorantes específicos de cromosoma dará fuertes señales de hibridación en un conjunto de sondas mutuamente exclusivas en el cromosoma no relacionado que no están directamente adyacentes entre sí en el molde lineal del cromosoma no relacionado. Las sondas localizadas en el cromosoma no relacionado entre los dos conjuntos de sondas que muestran interacciones de ADN-ADN intercromosómicas representan la región genómica que se perdió. Las sondas más externas de esta región marcan los sitios de restricción cercanos a los puntos de corte en el cromosoma no relacionado.

15 En otra realización preferida, se dispone de menos de 24 colorantes, y cada colorante se usa para marcar exclusivamente todos los fragmentos de ADN que interactúan con secuencias diana presentes en el mismo cromosoma. Por tanto, con 1, 2, 3, 4, 6, 8, 10, 12 etc. colorantes, 1, 2, 3, 4, 6, 8, 10, 12 etc. cromosomas pueden marcarse de forma particular. Los fragmentos de ADN marcados que pertenecen a cada uno de estos cromosomas pueden hibridarse entre sí con una matriz que contiene sondas que representan el genoma completo. La aparición de señales de hibridación, que representan interacciones ADN-ADN en un cromosoma no relacionado, identifica este cromosoma como un compañero de translocación. Un análisis 4C posterior dirigido específicamente hacia los cromosomas implicados en la translocación, permite la identificación de puntos de corte (como se ha descrito anteriormente).

20 En otra realización preferida, se dispone de menos de 48 colorantes y cada cromosoma se marca con dos colorantes únicos que se utilizan de tal manera que alternan entre fragmentos de ADN que interactúan con secuencias diana que están próximas en el molde del cromosoma lineal. Por tanto, con 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 etc. colorantes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12 etc. cromosomas pueden marcarse de forma particular. Todos los fragmentos de ADN pueden hibridarse entre sí con una matriz que contiene sondas que representan el genoma completo. Las señales de hibridación en un cromosoma no relacionado representan interacciones de ADN-ADN intercromosómicas e identifican este cromosoma como un compañero de translocación. La transición de señales bajas a altas en un cromosoma no relacionado identifica el sitio de reconocimiento de la enzima de restricción primaria cerca del punto de corte de ADN.

30 En una realización aún preferida, se dispone de menos de 48 colorantes y cada cromosoma está marcado con más de dos colorantes exclusivos que se usan de tal manera que alternan entre fragmentos de ADN que interactúan con secuencias diana que están adyacentes en el molde de cromosoma lineal. Todos los fragmentos de ADN pueden hibridarse entre sí con una matriz que contiene sondas que representan el genoma completo. Las señales de hibridación en un cromosoma no relacionado representan interacciones de ADN-ADN intercromosómicas e identifican este cromosoma como un compañero de translocación. La transición de señales bajas a altas en un cromosoma no relacionado identifica el sitio de reconocimiento de enzima de restricción primaria cerca del punto de corte de ADN.

#### *Cómo identificar otros puntos de corte de translocación*

50 Los puntos de corte identificados en un cromosoma de acuerdo con cualquiera de las estrategias mencionadas anteriormente pueden utilizarse posteriormente para identificar puntos de corte en cromosoma compañero de translocación. Esto puede realizarse por ejemplo mediante secuenciación de productos obtenidos a través de PCR de largo alcance sobre uniones de ADN o PCR inversa o PCR ligada por ligamiento (LM) en círculos de ADN creados por digestión enzimática de restricción y religamiento de ADN genómico (no entrecruzado) utilizando cebadores específicos para un cromosoma que lee en secuencias del otro cromosoma.

En una realización preferida, cada una de las exploraciones anteriormente mencionadas para reordenamientos genómicos pueden ir seguidas de un experimento 4C exclusivo dirigido contra secuencias diana cerca de puntos de corte candidatos para identificarlos de manera inequívoca como tales.

60 Los aspectos de los métodos descritos anteriormente para la detección de diferentes tipos de reordenamientos genómicos pueden combinarse para explorar simultáneamente el genoma con respecto a su aparición.

Si se utilizan matrices genómicas de tipo mosaico (*tiling*), en lugar de matrices 4C, las transiciones en las intensidades de señal para sondas yuxtapuestas en el molde del cromosoma lineal que se observan en muestras de

pacientes pero no en las de control, indican la posición de puntos de corte asociados a un reordenamiento genómico (en lugar del sitio de reconocimiento de la enzima de restricción primaria cerca del punto de corte de ADN).

## SECUENCIACIÓN

5 La secuenciación de ADN de alto rendimiento promete llegar a convertirse en una alternativa asequible y más cuantitativa para micromatrices para analizar grandes conjuntos de secuencias de ADN. Ejemplos de estrategias de secuenciación de alto rendimiento se indican en E. Y. Chan, *Mutation Reseach* 573 (2005) 13-40 e incluyen, pero sin limitación, estrategias de secuenciación a corto plazo tales como estrategias de extensión de ciclos, estrategias de lectura con polimerasa y secuenciación con exonucleasa, estrategias de secuenciación revolucionarias tales como barrido de ADN y secuenciación de nanoporos y análisis lineal directo. Los ejemplos de métodos de secuenciación de alto rendimiento actuales son (piro) secuenciación 454, Sistema de Análisis del Genoma de Solexa, método de secuenciación Agencourt SOLiD (Applied Biosystems), secuenciación MS-PET (Ng *et al.*, 2006, <http://nar.oxfordjournals.org/cgi/content/full/34/12/e84>).

15 La secuenciación puede reemplazar la hibridación de matriz en el análisis de alto rendimiento de los resultados de 4C y otras estrategias basadas en la detección de interacciones genómicas. La frecuencia de la aparición de una secuencia es indicativa de la frecuencia de asociación en el genoma y puede analizarse de la misma manera ya que los resultados de hibridación se analizan en lo anterior.

20 La secuenciación se realiza en el molde proporcionado por las etapas *a-g*, descritas en la reivindicación 1. Como alternativa, la secuenciación puede realizarse en productos de PCR obtenidos por métodos como se describe en Lomvardas *et al.*, *Cell* 126, 403-413, 28 de julio de 2006, o en Ling *et al.*, *Science* 312, 14 de abril de 2006, 269-272.

25 La secuenciación se inicia a partir de uno o ambos extremos de los productos de PCR. Ambos extremos de los productos de PCR constan de secuencias de nucleótidos de composición nucleotídica conocida, siendo al menos uno de los extremos secuencia de nucleótidos diana, y flanquean una secuencia de nucleótidos de interés que interacciona y se liga con la secuencia de nucleótidos diana. Dependiendo del método de secuenciación utilizado, los adaptadores pueden ser necesarios para añadir a uno o ambos extremos de los productos de PCR. Los adaptadores pueden ser secuencias oligonucleotídicas necesarias para el método de secuenciación de interés, que pueden o no contener residuos que por ejemplo permiten capturarlos. Los adaptadores pueden ligarse a los productos de la PCR directamente o después del enromado de los extremos de los productos de PCR. Como alternativa, los conjuntos de cebadores de PCR, como se utilizan en la etapa *g* (reivindicación 1), pueden contener salientes que representan secuencias adaptadoras o pueden contener salientes que introducen sitios de escisión de enzimas de restricción que pueden utilizarse para ligamiento posterior de adaptadores específicos o no específicos para cada extremo del producto de PCR.

35 En experimentos 4C cuyo objetivo es identificar las secuencias de ADN que interaccionan con una sola secuencia diana de nucleótidos, la secuenciación debe leerse a través del evento de ligamiento primario (etapa *c*) y/o del evento de ligamiento secundario (etapa *f*) de tal manera que se obtiene una información de secuencia suficiente para identificar las secuencias de nucleótidos de interés. Típicamente, esto requiere la secuenciación e identificación de tramos de mínimamente 8-30 nucleótidos más allá de la unión de ligamiento en la secuencia de nucleótidos de interés (véase la Figura 28).

45 En la secuenciación 4C múltiple se requiere leer a través del evento de ligamiento primario (etapa *c*) y/o el evento de ligamiento secundario (etapa *f*) de tal manera que se obtiene suficiente información de secuencia para identificar tanto la secuencia de nucleótidos diana como la secuencia de nucleótidos de interés que forman conjuntamente el producto de ligamiento. Típicamente, esto requiere la secuenciación e identificación de tramos de 8-30 nucleótidos en cada lado de la unión de ligamiento. La identificación de la secuencia de nucleótidos diana proporcionará cada producto de ligamiento con un "dirección de domicilio". Los cebadores utilizados en la etapa *g* que hibridan con la secuencia de nucleótidos diana deben localizarse a una distancia desde la unión de ligamiento primario y secundario que está lo suficientemente lejos para que la secuenciación identifique inequívocamente esta "dirección de domicilio". Dependiendo del método de secuenciación, esta distancia puede ser mínimamente de 0, 10, 20, 30, 40, 50, 60, 70, 80, 90 o 100 nucleótidos separados de la unión de ligamiento.

55 En 4C múltiple y simple, cuando los productos de PCR constan de una secuencia de nucleótidos de interés flanqueada en cada lado por secuencias de nucleótidos diana, la lectura desde un extremo del producto de PCR a través de la unión de ligamiento primario, y la lectura desde el otro extremo del producto de PCR a través de la unión del ligamiento secundario, proporciona la misma información. Por tanto, las secuencias obtenidas desde un lado del producto de PCR son suficientes para el análisis de interacciones ADN-ADN. Las secuencias obtenidas desde el otro lado del producto de PCR pueden utilizarse para complementar o verificar datos. Véase la Figura 29.

65 El cebado de las reacciones de secuencia real puede realizarse utilizando los cebadores estándar utilizados habitualmente por las respectivas plataformas de secuenciación de alto rendimiento. Esto puede realizarse mediante cebadores de secuenciación a medida que solapan parcial o completamente con los cebadores de PCR inversos usados para amplificar secuencias capturadas por una secuencia diana determinada. Esto podría prevenir la

relectura de toda la secuencia de cebadores PCR inversa y permitiría la lectura de más nucleótidos del fragmento capturado, que por lo tanto puede identificarse más fácilmente y mapearse en el genoma. Estos cebadores de secuenciación a medida pueden después emparejarse parcialmente con las secuencias adaptadoras normalmente utilizadas como cadena de ADN complementaria al cebador de secuenciación.

5 En una configuración múltiple, los cebadores de secuenciación múltiples que solapan tanto con la secuencia adaptadora como con los extremos externos de los productos de ADN amplificados por PCR inversa pueden utilizarse simultáneamente, de tal manera que cada fragmento diana incluido en el análisis tiene su propio cebador de secuencia exclusivo. El solapamiento con los extremos externos de los productos de ADN amplificados por PCR  
10 inversa (es decir la secuencia de cebador de PCR inversa) debe ser tal que se encuentran aún disponibles nucleótidos suficientes de la secuencia diana para la secuenciación y de tal manera que cada secuencia diana (dirección de domicilio) puede identificarse de forma inequívoca. Dependiendo de la naturaleza y el número de las secuencias diana analizadas simultáneamente, esto significa que en una realización, los cebadores de secuenciación pueden hibridar de 1 a 20 nucleótidos de distancia del sitio de restricción que forma la unión entre la  
15 secuencia diana y la secuencia capturada. La identificación de cada combinación de secuencia diana y secuencia capturada también puede realizarse utilizando secuenciación di-etiqueta que proporciona información de secuencia de ambos extremos de cada fragmento de ADN analizado.

20 En otra realización, los cebadores de secuenciación se diseñan para cada secuencia diana de tal manera que están cerca de los sitios de reconocimiento de la enzima de restricción primaria y/o secundaria analizados y pueden solapar parcial o completamente con los sitios de reconocimiento de la enzima de restricción primaria y/o secundaria analizados.

En una realización, la tecnología de secuenciación de alto rendimiento utilizada es la secuenciación Solexa (Illumina).

25 En una realización, la secuenciación puede dirigirse al lado del sitio de reconocimiento de la enzima de restricción primaria. Esto puede evitar la lectura de eventos de ligamiento al azar que se producen durante la segunda etapa del ligamiento.

30 En una realización, la secuenciación puede dirigirse al lado del sitio de reconocimiento de la enzima de restricción secundaria.

#### *Análisis de datos*

35 Los productos de ligamiento analizados por tecnología 4C están compuestos de una "dirección de domicilio" (secuencia de nucleótidos diana) y de una secuencia nucleotídica de interacción de interés ("dirección de viaje"). En el caso de 4C simple este "dirección de domicilio" se conoce (que es la única secuencia de nucleótidos diana analizada).

40 En casos de 4C múltiple, la "dirección de domicilio" se identifica electrónicamente comparando la secuencia de nucleótidos diana obtenida frente a una biblioteca/base de datos que contiene todas las secuencias de nucleótidos diana incluidas en el análisis. La "dirección de viaje" obtenida (es decir la secuencia de nucleótidos de interés) se identifica electrónicamente comparando su secuencia con una biblioteca/base de datos que contiene todos los fragmentos de ADN genómicos que se localizan entre un sitio de reconocimiento de la enzima de restricción primaria  
45 y secundaria de elección.

Para visualizar los datos y facilitar el análisis, en una realización preferida cada producto de ligamiento secuenciado se representa gráficamente a lo largo del molde lineal de los cromosomas, en la localización genómica de la  
50 "dirección de viaje" (secuencia de nucleótidos de interés). Los códigos de color único revelan la "dirección de domicilio" de cada producto de ligamiento y la frecuencia de detección de cada producto de ligamiento se indica gráficamente. Los reordenamientos genómicos pueden detectarse dividiendo cada frecuencia de interacción medida en una muestra (por ejemplo obtenida de un paciente) sobre las medidas en la otra muestra (por ejemplo en un sujeto sano). Estos valores pueden representarse gráficamente a lo largo de los moldes de cromosoma lineal también.

55 Una disminución en la frecuencia de interacción ADN-ADN en la muestra de ensayo es indicativa de una delección. Típicamente, esto coincide con un aumento en las frecuencias de interacción ADN-ADN para secuencias más allá del punto de corte más distal como se mide a partir de la secuencia diana. Un aumento en la frecuencia de interacción ADN-ADN en la muestra de ensayo es indicativo de una duplicación. Típicamente, esto coincide con una  
60 disminución en las frecuencias de interacción ADN-ADN para secuencias más allá del punto de corte más distal medido desde la secuencia diana. Una inversión en la frecuencia de interacción ADN-ADN en la muestra de ensayo es indicativa de una inversión genómica. La detección de interacciones ADN-ADN a través de cromosomas es indicativa de una translocación. Los puntos de corte se detectan como se describe para el análisis de micromatriz.

65

BIOMARCADORES

La identificación de reordenamientos - tales como translocaciones, inversiones y deleciones, que están asociadas con una enfermedad permite la identificación de biomarcadores que pueden utilizarse para el diagnóstico de la enfermedad. Por ejemplo, pueden diseñarse sondas de hibridación o cebadores de PCR que detecten un reordenamiento determinado y utilizarse para diagnosticar la enfermedad en un paciente. Pueden diseñarse sondas de PCR de acuerdo con técnicas conocidas en la técnica, de tal manera que una región susceptible a reordenamiento en una patología se amplifica utilizando los cebadores; la naturaleza del producto de amplificación será indicativa de la presencia o de la ausencia de la enfermedad. Como alternativa, pueden diseñarse sondas de hibridación o cebadores que se hibridarán exclusivamente en presencia o en ausencia del reordenamiento. Las proteínas de fusión resultantes de reordenamientos pueden detectarse por técnicas tales como detección de anticuerpos con anticuerpos diseñados de acuerdo con técnicas conocidas en la materia o por espectrometría de masas.

SUJETO

El término "sujeto" incluye mamíferos - tales como animales y seres humanos.

AGENTE

El agente puede ser un compuesto orgánico u otro compuesto químico. El agente puede ser un compuesto, que puede obtenerse a partir de o producirse por cualquier fuente adecuada, bien sea natural o artificial. El agente puede ser una molécula de aminoácido, un polipéptido, o un derivado químico de los mismos, o una combinación de los mismos. El agente puede incluso ser una molécula de polinucleótido - que puede ser una molécula con sentido o antisentido, o un anticuerpo, por ejemplo, un anticuerpo policlonal, un anticuerpo monoclonal o un anticuerpo monoclonal humanizado.

Se han desarrollado diversas estrategias para producir anticuerpos monoclonales con carácter humano, que evitan la necesidad de una línea celular humana productora de anticuerpos. Por ejemplo, los anticuerpos monoclonales de ratón útiles se han "humanizado" ligando regiones variables de roedores y regiones constantes de seres humanos (Winter, G. y Milstein, C. (1991) Nature 349, 293-299). Esto reduce la inmunogenicidad anti ratón de ser humano del anticuerpo pero la inmunogenicidad residual se conserva en virtud del marco conservado de la región V exógena. Además, la especificidad de unión a antígeno es esencialmente la del donante murino. El injerto de CDR y manipulación de regiones marco conservadas (documento EP 0239400) tiene una manipulación de anticuerpo mejorada y refinada hasta el punto en el que es posible producir anticuerpos murinos humanizados que son aceptables para uso terapéutico en seres humanos. Los anticuerpos humanizados pueden obtenerse utilizando otros métodos bien conocidos en la materia (por ejemplo como se describe en el documento US-A-239400).

Los agentes pueden unirse a una entidad (por ejemplo una molécula orgánica) mediante un enlazador que puede ser un enlazador bifuncional hidrolizable.

La entidad puede diseñarse u obtenerse a partir de una biblioteca de compuestos, que puede comprender péptidos, así como otros compuestos, tales como moléculas orgánicas pequeñas.

Como ejemplo, la entidad puede ser una sustancia natural, una macromolécula biológica o un extracto fabricado de materiales biológicos tales como células de bacterias, hongos o de animales (particularmente mamíferos) o tejidos, una molécula orgánica o inorgánica, un agente sintético, un agente semisintético, un mimético estructural o funcional, un péptido, un peptidomimético, un péptido escindido de una proteína entera, o péptidos sintetizados de manera sintética (tal como, por ejemplo, bien utilizando un sintetizador peptídico o por técnicas recombinantes o combinaciones de los mismos, un agente recombinante, un anticuerpo, un agente natural o no natural, una proteína de fusión o equivalente de los mismos y mutantes, derivados o combinaciones de los mismos).

Típicamente, la entidad será un compuesto orgánico. Para algunos casos, los compuestos orgánicos comprenderán dos o más grupos de hidrocarbilo. En este caso, la expresión "grupo hidrocarbilo" significa un grupo que comprende al menos C y H y opcionalmente puede comprender uno o más sustituyentes adecuados distintos. Como ejemplos de dichos sustituyentes pueden incluirse halo-, alcoxi-, nitro-, un grupo alquilo, un grupo cíclico, etc. Además de la posibilidad de que los sustituyentes sean un grupo cíclico, una combinación de sustituyentes puede formar un grupo cíclico. Si el grupo hidrocarbilo comprende más de una C entonces aquellos carbonos no requieren necesariamente ligarse entre sí. Por ejemplo, al menos dos de los carbonos pueden ligarse mediante un elemento o grupo adecuado. Por tanto, el grupo hidrocarbilo puede contener heteroátomos. Los heteroátomos adecuados serán obvios para los expertos en la técnica e incluirán, por ejemplo, azufre, nitrógeno y oxígeno. Para algunas aplicaciones, preferentemente la entidad comprende al menos un grupo cíclico. El grupo cíclico puede ser un grupo policíclico, tal como un grupo policíclico no condensado. Para algunas aplicaciones, la entidad comprende al menos uno de dichos grupos cíclicos ligados a otro grupo hidrocarbilo.

La entidad puede contener grupos halo - tales como grupos fluoro, cloro, bromo o yodo.

La entidad puede contener uno o más de grupos alquilo, alcoxi, alqueno, alqueno y alqueno - que pueden ser de cadena ramificada o no ramificada.

#### ENFERMEDAD

5 Los aspectos de la presente invención pueden utilizarse para el tratamiento y/o prevención y/o diagnóstico y/o pronóstico de una enfermedad, tales como las indicadas en el documento WO-A-98/09985.

10 Para facilitar la referencia, parte de ese listado se proporciona ahora: actividad inhibidora de macrófagos y/o de linfocitos T y por tanto, actividad antiinflamatoria; actividad antiinmunitaria, es decir efectos inhibidores contra una respuesta inmunitaria celular y/o humoral, incluyendo una respuesta no asociada con inflamación; enfermedades asociadas con virus y/u otros patógenos intracelulares; inhibir la capacidad de los macrófagos y linfocitos T para adherirse a componentes de la matriz extracelular y fibronectina, así como regular positivamente la expresión de receptor fas en linfocitos T; inhibir reacción inmunitaria no deseada e inflamación incluyendo artritis, incluyendo

15 artritis reumatoide, inflamación asociada a hipersensibilidad, reacciones alérgicas, asma, lupus eritematoso sistémico, enfermedades por colágeno y otras enfermedades autoinmunitarias, inflamación asociada a aterosclerosis, arteriosclerosis, cardiopatía aterosclerótica, lesión por reperfusión, parada cardíaca, infarto de miocardio, trastornos inflamatorios vasculares, síndrome de dificultad respiratoria u otras enfermedades cardiopulmonares, inflamación asociada a úlcera péptica, colitis ulcerosa y otras enfermedades del tracto

20 gastrointestinal, fibrosis hepática, cirrosis hepática u otras enfermedades hepáticas, tiroiditis u otras enfermedades glandulares, glomerulonefritis u otras enfermedades renales y urológicas, otitis u otras enfermedades otorrinolaringológicas, dermatitis u otras enfermedades dérmicas, enfermedades periodontales u otras enfermedades dentales, orquitis o epididimoorquitis, infertilidad, traumatismo orquidal u otras enfermedades testiculares relacionadas con el sistema inmunitario, disfunción de la placenta, insuficiencia de la placenta, aborto habitual, eclampsia, preeclampsia y otras enfermedades ginecológicas relacionadas con el sistema inmunitario y/o

25 con inflamación, uveítis posterior, uveítis intermedia, uveítis anterior, conjuntivitis, coriorretinitis, uveorretinitis, neuritis óptica, inflamación intraocular, por ejemplo, retinitis o edema macular cistoide, oftalmia simpática, escleritis, retinitis pigmentosa, componentes inmunitarios e inflamatorios de enfermedad degenerativa del fondo, componentes inflamatorios de traumatismo ocular, inflamación ocular causada por infección, vitreoretinopatías proliferativas,

30 neuropatía óptica isquémica aguda, cicatrización excesiva, por ejemplo, después de operación de filtración de glaucoma, reacción inmunitaria y/o inflamatoria contra implantes oculares y otras enfermedades oftálmicas relacionadas con el sistema inmunitario e inflamatorias, inflamación asociada con enfermedades o afecciones o trastornos autoinmunitarios en los que, en el sistema nervioso central (SCN) o en cualquier otro órgano, la supresión inmunitaria y/o inflamatoria sería beneficiosa, enfermedad de Parkinson, complicación y/o efectos secundarios del

35 tratamiento de la enfermedad de Parkinson, encefalopatía relacionada con VIH complejo de demencia relacionado con SIDA, enfermedad de Devic, corea de Sydenham, enfermedad de Alzheimer y otras enfermedades, afecciones o trastornos degenerativos del SNC, componentes inflamatorios de ictus, síndrome postpolio, componentes inmunitarios e inflamatorios de trastornos psiquiátricos, mielitis, encefalitis, pan encefalitis esclerosante subaguda, encefalomielitis, neuropatía aguda, neuropatía subaguda, neuropatía crónica, síndrome de Guillain-Barre, corea de

40 Sydenham, miastenia grave, pseudotumor cerebral, Síndrome de Down, enfermedad de Huntington, esclerosis lateral amiotrófica, componentes inflamatorios de compresión del SNC o traumatismo del SNC o infecciones del SNC, componentes inflamatorios de atrofia y distrofias musculares, y enfermedades, afecciones o trastornos relacionados con el sistema inmunitario e inflamatorias del sistema nervioso central y periférico, inflamación postraumática, choque séptico, enfermedades infecciosas, complicaciones inflamatorias o efectos secundarios de

45 cirugía, complicaciones y/o efectos secundarios de trasplante de médula ósea u otro trasplante, complicaciones y efectos secundarios de terapia génica inflamatorios y/o inmunitarios, por ejemplo debido a infección con un portador de virus, o inflamación asociada con SIDA, para suprimir o inhibir una respuesta inmunitaria humoral y/o celular, para tratar o mejorar enfermedades proliferativas de monocitos o leucocitos, por ejemplo leucemia, reduciendo la cantidad de monocitos o linfocitos, para la prevención y/o el tratamiento de rechazo de injerto en casos de trasplante

50 de células, tejidos y órganos naturales o artificiales tales como córnea, médula ósea, órganos, lentillas, marcapasos, tejido de piel natural o artificial. Los trastornos específicos relacionados con cáncer incluyen, pero sin limitación: tumores sólidos; tumores portados en la sangre tales como leucemias; metástasis tumoral; tumores benignos, por ejemplo hemangiomas, neuromas acústicos, neurofibromas, tracomias y granulomas piogénicos; artritis reumatoide; psoriasis; enfermedades angiogénicas oculares, por ejemplo, retinopatía diabética, retinopatía de prematuro, degeneración macular, rechazo de injerto de córnea, glaucoma neovascular, fibroplasia retrolental, rubeosis; Síndrome Osler-Webber; angiogénesis miocárdica; neovascularización de placa; telangiectasia; articulaciones hemofílicas; angiofibroma; granulación de heridas; colaterales coronarios; colaterales cerebrales; malformaciones arteriovenosas; angiogénesis de extremidades isquémica; glaucoma neovascular; fibroplasia retrolental; neovascularización diabética; enfermedades relacionadas con *Helicobacter*, fracturas, vasculogénesis,

60 hematopoyesis, ovulación, menstruación y placentación.

Preferentemente, la enfermedad es cáncer - tal como leucemia linfocítica aguda (ALL), leucemia mieloide aguda (AML), cáncer adrenocortical, cáncer anal, cáncer de vejiga, cáncer de sangre, cáncer de hueso, tumor cerebral, cáncer de mama, cáncer del sistema genital de la mujer, cáncer del sistema genital del hombre, linfoma del sistema nervioso central, cáncer de cuello uterino, rhabdomyosarcoma juvenil, sarcoma juvenil, leucemia linfocítica crónica (CLL), leucemia mieloide crónica (CML), cáncer de colon y rectal, cáncer de colon, cáncer endometrial, sarcoma

65

5 endometrial, cáncer esofágico, cáncer ocular, cáncer de vesícula biliar, cáncer gástrico, cáncer de tracto gastrointestinal, tricoleucemia, cáncer de cabeza y cuello, cáncer hepatocelular, enfermedad de Hodgkin, cáncer hipofaríngeo, sarcoma de Kaposi, cáncer de riñón, cáncer laríngeo, leucemia, cáncer de hígado, cáncer de pulmón, histiocitoma fibroso maligno, tioma maligno, melanoma, mesotelioma, mieloma múltiple, mieloma, cáncer del seno  
 10 paranasal y de la cavidad nasal, cáncer nasofaríngeo, cáncer del sistema nervioso, neuroblastoma, linfoma no Hodgkin, cáncer de la cavidad bucal, cáncer orofaríngeo, osteosarcoma, cáncer de ovario, cáncer pancreático, cáncer paratiroideo, cáncer de pene, cáncer faríngeo, tumor de la hipófisis, neoplasia de células plasmáticas, linfoma primario del SNC, cáncer de próstata, cáncer de recto, cáncer del sistema respiratorio, retinoblastoma, cáncer de glándulas salivales, cáncer de piel, cáncer de intestino delgado, sarcoma de tejido blando, cáncer de estómago, cáncer testicular, cáncer tiroideo, cáncer de sistema urinario, sarcoma uterino, cáncer vaginal, sistema vascular, macroglobulinemia de Waldenstrom y tumor de Wilms.

KITS

15 Los materiales para su uso en los métodos de la presente invención se ajustan adecuadamente para la preparación de kits.

20 Dicho kit puede comprender envases, cada uno con uno o más de los diversos reactivos (típicamente en forma concentrada) utilizados en los métodos descritos en el presente documento, incluyendo, por ejemplo, una enzima de restricción primaria, una enzima de restricción secundaria, un agente de entrecruzamiento, una enzima de ligamiento (por ejemplo una ligasa) y un agente para invertir el entrecruzamiento (por ejemplo proteinasa K).

25 También pueden proporcionarse oligonucleótidos en los envases que pueden estar en cualquier forma, por ejemplo liofilizado o en solución (por ejemplo, una solución de agua destilada o tamponada), etc.

30 En un aspecto preferido de la presente invención, se proporciona un kit que comprende un conjunto de sondas como se describe en el presente documento, una matriz y opcionalmente uno o más marcadores.

También se incluirá típicamente un conjunto de instrucciones.

USOS

35 Ventajosamente, la presente invención puede utilizarse para obtener información sobre la organización espacial de secuencias de nucleótidos - tales como locus genómicos *in vitro* o *in vivo*.

40 Como ejemplo, la tecnología 4C puede utilizarse para estudiar la organización tridimensional de uno o más locus génicos. En particular, esta tecnología puede utilizarse para estudiar la función de uno o más factores de transcripción en la organización tridimensional de uno o más locus génicos.

45 Como ejemplo adicional, la tecnología 4C puede utilizarse para estudiar la función de factores que actúan en *trans* y elementos de ADN reguladores que actúan en *cis*.

50 Como un ejemplo adicional, la tecnología 4C puede utilizarse para estudiar la regulación de genes de largo alcance *in vitro* o *in vivo*.

55 Como ejemplo adicional, la tecnología 4C puede utilizarse para estudiar la proximidad e interacción intracromosómica.

60 Como ejemplo adicional, la tecnología 4C puede utilizarse para estudiar la proximidad e interacción intercromosómica.

65 Como ejemplo adicional, la tecnología 4C puede utilizarse para identificar secuencias de nucleótidos que actúan con un promotor, potenciador, silenciador, aislante, región de control de locus, origen de replicación, MAR, SAR, centrómero, telómero u otra secuencia de interés en una red reguladora.

Como ejemplo adicional, la tecnología 4C puede utilizarse para identificar genes responsables de un fenotipo (enfermedad) en los que se produce una mutación y/o delección que afecta a un elemento regulador distante y su mapeo por lo tanto no proporciona dicha información.

60 Como ejemplo adicional, la tecnología 4C puede utilizarse para reconstruir eventualmente la conformación espacial de locus génicos, regiones genómicas grandes o incluso cromosomas completos.

65 Como un ejemplo adicional, la tecnología 4C puede utilizarse para definir posibles secuencias de anclaje que mantienen determinados cromosomas juntos en el espacio nuclear.

Como un ejemplo adicional, la tecnología 4C puede utilizarse para reconstruir eventualmente a alta resolución el posicionamiento de cromosomas unos con respecto a otros.

5 Como ejemplo adicional, la tecnología 4C puede utilizarse en el diagnóstico (por ejemplo diagnóstico prenatal) para detectar o identificar reordenamientos y/o aberraciones genómicas - tales como translocaciones, deleciones, inversiones, duplicaciones.

#### TÉCNICAS GENERALES DE METODOLOGÍA DE ADN RECOMBINANTE

10 La presente invención emplea, a menos que se indique de otra manera, técnicas convencionales de química, biología molecular, microbiología, ADN recombinante e inmunología, que se encuentran dentro de las capacidades de un experto habitual en la técnica. Dichas técnicas se explican en la bibliografía. Véase, por ejemplo, J. Sambrook, E. F. Fritsch, y T. Maniatis, 1989, *Molecular Cloning: A Laboratory Manual*, Segunda Edición, Libros 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. *et al.* (1995 y suplementos periódicos; *Current Protocols in Molecular Biology*, cáp. 9, 13 y 16, John Wiley and Sons, Nueva York, N. Y.); B. Roe, J. Crabtree, y A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Techniques*, John Wiley and Sons; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical Approach*, Irl Press; y, D. M. J. Lilley y J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA Methods in Enzymology*, Academic Press. Cada uno de estos textos generales se incorpora en el presente documento por referencia.

20 Ahora se describirá la invención adicionalmente como ejemplos, cuya intención es ayudar a uno de los expertos habituales en la técnica a realizar la invención y no pretende de ninguna manera limitar su alcance.

#### EJEMPLO 1

25 *Sección de materiales y métodos que va con las Figuras 2, 13, 14, 15, 16, 17, 19.*

#### Tecnología 4C

30 Las etapas iniciales del procedimiento de la tecnología 3C se realizaron como se ha descrito anteriormente (Splinter *et al.* (2004). *Methods Enzymol* 375, 493-507 (2004), proporcionando productos de ligamiento entre los fragmentos HindIII. Este molde 3C ligado a HindIII (~50 µg) se digirió durante una noche a 100 ng/µl con 50 U de una enzima de restricción, de corte frecuente, secundaria que es bien DpnII (HS2, Rad23A) o NlaIII (β mayor). Para impedir la mutación en la formación de círculos del ADN (Rippe *et al.* (1995) *Trends Biochem Sci* 20, 500-6), se tuvo cuidado de seleccionar una enzima de restricción secundaria que no cortaba aproximadamente 350-400 pb desde el sitio de restricción HindIII que demarca el fragmento de restricción de interés (es decir el "cebo"). Después de la digestión con la enzima de restricción secundaria, el ADN se extrajo con fenol, se precipitó con etanol y posteriormente se ligó a baja concentración (muestra de 50 µg en 14 ml utilizando 200 U de ligasa (Roche), 4 horas a 16 °C) para promover la formación de DpnII- o círculo de DpnII. Los productos de ligamiento se extrajeron con fenol y se precipitaron con etanol, utilizando glucógeno (Roche) como un transportador (20 µg/ml). Los círculos de interés se linealizaron digiriendo durante una noche con 50 U de una enzima de restricción terciaria que cortaba el cebo entre los sitios de reconocimiento de la enzima de restricción primaria y secundaria, utilizando las siguientes enzimas de restricción: SpeI (HS2), PstI (Rad23A) y PflmI (β-mayor). Esta etapa de linealización se realizó para facilitar hibridación con cebadores posterior durante las primeras rondas de amplificación por PCR. Los productos digeridos se purificaron utilizando una columna de retirada de nucleótidos QIAquick (250) (Qiagen).

Las reacciones de PCR se realizaron utilizando el sistema de PCR Molde Largo Expandido (Roche), utilizando las condiciones cuidadosamente optimizadas para garantizar la amplificación lineal de fragmentos de un tamaño de hasta 1,2 kb (80 % de fragmentos de PCR 4C son más pequeños que 600 pb). Las condiciones de la PCR fueron las siguientes: 94 °C durante 2 minutos, 30 ciclos de 94 °C durante 15 segundos, 55 °C durante 1 minuto y 68 °C durante 3 minutos, seguido de una etapa final de 68 °C durante 7 minutos. Se determinó la cantidad máxima de molde que aún muestra un rango de amplificación lineal. Para esto, se añadieron diluciones en serie de molde a las reacciones de PCR, el material de ADN amplificado se procesó en un gel de agarosa y los productos de PCR se cuantificaron utilizando el programa informático ImageQuant. Típicamente, 100-200 ng de molde por 50 µl de reacción de PCR dieron productos en el intervalo de amplificación lineal. Se agruparon de 16 a 32 reacciones de PCR y se purificó este molde 4C utilizando el sistema de retirada de nucleótidos QIAquick (250) (Qiagen). El molde 4C purificado se marcó y se hibridó con matrices de acuerdo con protocolos de micromatriz ChIP convencionales (Nimblegen Systems of Iceland, LLC). El ADN genómico marcado diferencialmente, que se digirió con la enzima primaria y secundaria utilizada en el procedimiento 4C, sirvió como un molde de control para corregir las diferencias en las eficiencias de hibridación. Para cada experimento, se marcaron dos muestras independientemente procesadas con orientaciones de colorantes alternativas.

Secuencias de cebadores 4C utilizadas:  
 HS2: 5'-ACTTCCTACACATTAACGAGCC-3',  
 5'- GCTGTTATCCCTTTCTCTTCTAC-3'



Rad23A: 5'- TCACACGCGAAGTAGGCC-3',  
5'- CCTTCTCCACCATGATGA-3'

β-mayor: 5'-AACGCATTTGCTCAATCAACTACTG-3',  
5'-GTTGCTCCTCACATTTGCTTCTGAC-3'

#### Matrices 4C

5 Las matrices y análisis se basaron en construcción NCBI m34. Se seleccionaron sondas (de 60 meros) a partir de las secuencias de 100 pb cadena arriba y abajo de sitios HindIII. El contenido de CG se optimizó hacia 50 %, para señales de hibridación uniformes. Para impedir la hibridación cruzada, las sondas que tenían cualquier similitud con repeticiones muy abundantes (RepBase 10,09)<sup>3</sup> se retiraron del conjunto de sondas. Además, las sondas que dieron más de dos aciertos BLAST en el genoma también se retiraron del conjunto de sondas. Los alineamientos de secuencias se realizaron utilizando MegaBLAST (Zhang *et al.* (2000) J Comput Biol 7, 203-14) utilizando ajustes convencionales. Se definió un éxito como un alineamiento de 30 nt o mayor.

#### Análisis de datos 4C

15 Para cada sonda se calculó la proporción de señal entre muestra 4C/ADN genómico y los datos se visualizaron con un programa informático SignalMap proporcionado por Nimblegen Systems. Los datos se analizaron utilizando el paquete R (<http://www.r-project.org>), Spotfire y Excel. Las proporciones de hibridación no procesadas mostraron grupos de 20-50 señales 4C positivas a lo largo del molde cromosómico. Para definir estas agrupaciones, se aplicó una media móvil. Se utilizaron diversos tamaños de ventana, que variaban de 9-39 sondas, que identificaron todos los mismos grupos. Los resultados mostrados se basaron en un tamaño de ventana de 29 sondas (con un promedio de 60 kb) y se compararon con la media móvil realizada a través de los datos al azar. Esto se realizó para cada matriz por separado. Por consiguiente, todas las mediciones se apreciaron con respecto a la amplitud y ruido de esa matriz específica. La proporción de descubrimientos falsos (FDR), definida como (número de positivos falsos)/(número de positivos falsos + número de positivos verdaderos) se determinó de la siguiente manera: (número de positivos en el conjunto aleatorizado)/(número de positivos en los datos). El nivel de umbral se determinó utilizando una estrategia vertical para establecer el valor mínimo para el cual: FDR <0,05.

20 A continuación, se compararon los experimentos biológicos duplicados. Las ventanas que cumplían el umbral de ambos duplicados se consideraron positivas. Cuando se compararon los datos al azar, ninguna ventana estaba por encima del umbral en ninguno de los duplicados. Las ventanas positivas adyacentes directamente al molde cromosómico se juntaron (no se permitieron huecos), creando áreas positivas.

#### Análisis de expresión

35 Para cada tejido, se realizaron tres micromatrices independientes de acuerdo con el protocolo Affymetrix (matrices de ratón 430\_2). Los datos se normalizaron utilizando catools RMA; ([www.bioconductor.org](http://www.bioconductor.org)) y para cada conjunto de sondas las mediciones de las tres micromatrices se promediaron. Además, cuando múltiples conjuntos de sondas representaban el mismo gen, también se promediaron. Se utilizó Mas5calls (biblioteca Affy: [www.bioconductor.org](http://www.bioconductor.org)) para establecer llamadas “presentes”, “ausentes” y “marginales”. Los genes con una llamada “presente” en las tres matrices y un valor de expresión mayor de 50 se denominaron expresados. Los genes con una llamada “presente” en las tres matrices y un valor de expresión mayor de 50 se denominaron expresados. Los “genes específicos de hígado fetal” se clasificaron como genes que cumplían los criterios de expresarse en hígado fetal y tenían valores de expresión más de cinco veces mayores en comparación con el cerebro fetal. Para proporcionar una medición de actividad transcripcional global alrededor de cada gen, se aplicó una suma móvil. Para esto, se utilizaron valores de expresión transformados por logaritmo. Para cada gen se calculó la suma de la expresión de todos los genes encontrados en una ventana 100 kb cadena arriba del inicio y 100 kb cadena abajo del final del gen, incluyendo el propio gen. Los valores resultantes para genes activos encontrados dentro de las regiones 4C positivas (n = 124, 123 y 208 respectivamente para HS2 en hígado, Rad23A en cerebro y Rad23A en hígado) se compararon con los valores obtenidos para genes activos fuera de las áreas 4C positivas (n = 153, 301 y 186, respectivamente, donde n = 153 corresponde al número de genes activos, que no interaccionan, presentes entre la región de interacción más centromérica y el telómero del cromosoma 7); los dos grupos se compararon utilizando un ensayo de suma de rangos de Wilcoxon de una cola.

#### Sondas FISH

55 Se utilizaron los siguientes clones BAC (BACPAC Resources Center); RP23-370E12 para Hbb-1, RP23-317H16 para crom.7at 80,1 Mb (grupo de genes OR), RP23-334E9 para Uros, RP23-32C19 para crom.7 a 118,3 Mb, RP23-143F10 para crom.7 a 130,1 Mb, RP23-470N5 para crom.7 a 73,1 Mb, RP23-247L11 para crom.7 a 135,0 Mb (grupo de genes OR), RP23-136A15 para Rad23A, RP23-307P24 para crom.8 a 21,8 Mb y RP23-460F21 para crom.8 a 122,4 Mb. Para una sonda específica de centrómero del cromosoma 7 se usó el clon P1 5279 (Genome Systems Inc.) que se empareja con el segmento de ADN D7Mit21. Se prepararon sondas marcadas con cebador aleatorio utilizando el Sistema de Marcaje Genómico CGH de Matriz BioPrime (Invitrogen). Antes de marcar, el ADN se digirió

con DpnII y se purificó con un limpiador de ADN y un kit concentrador 5 (Zymo research). El ADN digerido (300 ng) se marcó con SpectrumGreen dUTP (Vysis) o Alexa fluor 594 dUTP (*molecular probes*) y se purificó a través de un ADN de PCR GFX y un kit de Purificación de Banda en Gel (Amersham Biosciences) para retirar los nucleótidos no incorporados. La especificidad de las sondas marcadas se analizó en propagaciones de metafase preparadas a partir de células ES murinas.

### Crio-FISH

Se realizó Crio-FISH como se ha descrito anteriormente<sup>5</sup>. En resumen, muestras de hígado y cerebro E14.5 se fijaron durante 20 minutos en paraformaldehído al 4 %/HEPES 250 mM, pH 7,5 y se cortaron en bloques pequeños de tejido seguido de otra etapa de fijación de 2 horas en paraformaldehído al 8 % a 4 °C. Los bloques de tejido fijado se sumergieron en sacarosa 2,3 M durante 20 minutos a temperatura ambiente, se montaron en un portaobjetos y se ultracongelaron con nitrógeno líquido. Los bloques de tejido se conservaron en nitrógeno líquido hasta su seccionamiento. Las criosecciones ultrafinas de aproximadamente 200 nm se cortaron utilizando un ultramicrotomo E de Reichert equipado con criounión (Leica). Utilizando un asa cargada con sacarosa, las secciones se transfirieron a cubreobjetos y se conservaron a -20 °C. Para la hibridación, las secciones se lavaron con PBS para retirar la sacarosa, se trataron con RNasa 250 ng/ml en 2xSSC durante 1 h a 37 °C, se incubaron durante 10 min en HCl 0,1 M, se deshidrataron en una serie de etanol y se desnaturalizaron durante 8 min a 80 °C en formaldehído al 70 %/2x SSC, pH 7,5. Las secciones se deshidrataron de nuevo directamente antes de la hibridación de la sonda. 500 ng de sonda marcada se coprecipitaron con 5 µg de ADN Cot1 de ratón (Invitrogen) y se disolvió en hybmix (formamida al 50 %, sulfato de dextrano 10 %, 2x SSC, tampón fosfato 50 mM, pH 7,5). Las sondas se desnaturalizaron durante 5 min a 95 °C, se volvieron a emparejar durante 30 min a 37 °C y se hibridaron durante al menos 40 horas a 37 °C. Después de lavados de posthibridación, los núcleos se tiñeron con tinción de contraste con DAPI 20 ng/ml (Sigma) en PBS/Tween 20 al 0,05 % y se montaron en un reactivo antidestinción Gold Prolong (Molecular Probes).

Se recogieron imágenes con un microscopio epifluorescente Zeiss Axio Imager Z1 (x 100 plano apocromático, objetivo de aceite 1,4), dotado con una cámara CCD y un programa FISH Imaging System Isis (Metasystems). Se analizó un mínimo de 250 alelos de β globina o Rad23A y se puntuaron como solapante o no solapante con BAC localizados en cualquier parte en el genoma, por una persona que no conocía la combinación de sondas aplicada a las secciones. Se realizaron ensayos de bondad de ajuste (estadística G)<sup>6</sup> duplicados, para evaluar el significado de las diferencias entre los valores medidos para regiones 4C positivas frente a 4C negativas. En la Tabla 2 se proporciona una visión general de los resultados.

Aunque se encontraron diferencias estadísticamente significativas entre las frecuencias de interacción de fondo (0,4-3,9 %) y verdaderas (5-20,4 %), puede estar claro que las frecuencias medidas por crio-FISH son más bajas que las medidas por otros que utilizan diferentes protocolos FISH. El seccionamiento puede separar algunos locus de interacción y por lo tanto las mediciones crio-FISH infraestimarán ligeramente las frecuencias de interacción verdaderas. Por otro lado, los procedimientos FISH 2D y 3D actuales sobreestimarán estos porcentajes debido a una resolución limitada en la dirección z. En el futuro, técnicas microscópicas mejoradas en combinación con sondas FISH más específicas revelarán mejores frecuencias de interacción verdaderas.

### EJEMPLO 2

El procedimiento 3C (es decir fijación en formaldehído, digestión con enzima de restricción (primaria), religamiento de fragmentos de ADN entrecruzados y purificación de ADN) se realiza esencialmente como se ha descrito (Splinter *et al.*, (2004) *Methods Enzymol.* 375: 493-507) produciendo una mezcla de ADN ("molde 3C") que contiene fragmentos de restricción que están ligados porque estaban originalmente muy cercanos en el espacio nuclear.

Se realizó PCR inversa para amplificar todos los fragmentos ligados a un fragmento de restricción determinado ("cebo"; seleccionado debido a que contiene un promotor, potenciador, aislante, región de unión a matriz, origen de replicación o cualquier otra secuencia nucleotídica primaria (diana)).

Para esto, se crean círculos de ADN digiriendo el molde 3C con una enzima de restricción secundaria (preferentemente un cortador frecuente que reconoce secuencias tetra- o pentanucleotídicas) seguido de ligamiento en condiciones diluidas de tal manera que se favorecen las interacciones intramoleculares. Para minimizar un sesgo en la formación de círculos debido a limitaciones topológicas (Rippe *et al.*, (2001) *Trends in Biochem. Sciences* 26, 733-40), se seleccionó una enzima de restricción secundaria que preferentemente corta el cebo a >350-400 pb desde el sitio de restricción primario. Para aumentar la eficiencia de la amplificación por PCR inversa y reproducibilidad, los círculos se linealizaron mejor antes de la amplificación por PCR con una enzima de restricción (por ejemplo, un cortador de 6 o más pb) que corta el cebo entre el sitio de restricción de diagnóstico primario y secundario.

La digestión del molde 3C con la enzima de restricción secundaria, la circularización a través del ligamiento en condiciones diluidas y linealización de círculos que contienen cebo se realizaron en condiciones convencionales

para dichas manipulaciones de ADN para producir un molde de ADN para amplificación por PCR inversa (“molde 4C”)

5 Por consiguiente, 10 µg de molde 3C se digirieron en 100 µl con 20 U de la enzima de restricción secundaria (durante una noche), seguido de termoinactivación de la enzima y purificación de ADN. El ligamiento se realizó en 10 ml (ADN 1 ng/µl) con 50 U de T4 ligasa (4 horas a 16 °C, 30 minutos a TA), seguido de purificación de ADN. Finalmente, la linealización de los círculos de interés se realizó en 100 µl con 20 U de enzima de restricción (durante una noche), seguido de nuevo de purificación de ADN.

10 Para la PCR inversa, se diseñaron dos cebadores específicos de cebo, cada uno tan cercano como sea posible al sitio de reconocimiento de la enzima de restricción primaria y directamente adyacente a la secundaria, respectivamente, y cada uno con su extremo 3' orientado externamente para que la extensión proceda inmediatamente a través de los sitios de restricción en un fragmento ligado al cebo. La PCR inversa con estos  
15 cebadores se realiza preferentemente en 100-400 ng de ADN de molde 4C (por 50 µl de mezcla de reacción de PCR), para incluir un número máximo de eventos de ligamiento por reacción de PCR. Se realizó PCR inversa aplicando el Sistema PCR de Molde Largo Expandido (Roche), utilizando el tampón 1 de acuerdo con los procedimientos del fabricante.

Se realizaron los siguientes ciclos de PCR:

- 20
1. 2 min 94 °C
  2. 15 s 94 °C
  3. 1 min 55 °C
  4. 3 min 68 °C
  - 25 5. repetir la etapa 2-4 29x (o cualquiera entre 25-40x)
  6. 7 min 68 °C
  7. finalizar

30 Se realizó electroforesis en gel para analizar la reproducibilidad entre reacciones de PCR individuales. Típicamente, deben obtenerse patrones de productos idénticos.

Para obtener suficiente material para marcar por cebado al azar e hibridación de matriz, las reacciones de PCR múltiples (cada una obtenida después de 30 ciclos de PCR) pueden agruparse, (en lugar de aumentar el número de ciclos de PCR por reacción). Como una alternativa de marcaje cebado al azar, los nucleótidos marcados pueden incorporarse en los últimos ciclos de PCR (por ejemplo 30 ciclos (sin marcador) + 10 ciclos (con marcador)).

### EJEMPLO 3

#### *Detección de translocación utilizando tecnología 4C*

40 La tecnología 4C se utiliza para medir las frecuencias de interacción de una secuencia X determinada presente en un cromosoma A determinado en células de un sujeto sano y en células de un paciente que lleva una sola translocación, recíproca, entre los cromosomas A y B, estando el punto de corte próximo a la secuencia X (como se muestra en la Figura 8).

45 En células normales este análisis revela señales de hibridación elevadas (es decir interacciones frecuentes con X) para (casi) cada sonda localizada en 0,2-10 Mb de la secuencia X en el cromosoma A (el tamaño real de la región cromosómica que muestra fuertes señales de entrecruzamiento depende en gran medida de la complejidad de la muestra que se hibridó con la matriz). En cualquier parte en el mismo cromosoma A, así como en otros cromosomas, no se observó dicha región amplia (en el molde de ADN lineal) de sondas con elevadas señales de hibridación.

50 Sin embargo, en las células del paciente, las señales de hibridación con todas las sondas del cromosoma A localizadas al otro lado del punto de corte se reducen en ~50 % (una copia del cromosoma A sigue intacta y producirá señales normales), mientras que se observa una única concentración (es decir, no presente en células normales) de señales de hibridación elevadas para las sondas que bordean el punto de corte en el cromosoma B. De hecho, la brusca transición entre las sondas que no muestran señales de hibridación frente a las que muestran señales de hibridación fuertes en el cromosoma B revela la localización del punto de corte en el cromosoma B.

### EJEMPLO 4

#### *Análisis de resultados de tecnología 4C*

65 La tecnología 4C se utilizó para caracterizar el entorno genómico de la región de control de locus (LCR, *locus control region*) de la β globina de ratón, centrándose en un fragmento de restricción que contenía su sitio hipersensible 2 (HS2, *hypersensitive site* 2). La LCR es un fuerte elemento regulador de la transcripción, específico de células

eritroides, necesario para obtener altos niveles de expresión del gen de la  $\beta$  globina. El locus de la  $\beta$  globina está presente en el cromosoma 7 en la posición 97 Mb, donde reside en un grupo grande, de 2,9 Mb, de genes de receptores olfativos, que se transcriben únicamente en neuronas olfativas. Las interacciones se analizaron en dos tejidos: hígado fetal E14.5, en el que la LCR es activa y los genes de P globina se transcriben altamente, y cerebro fetal E14.5, en el que la LCR es inactiva y los genes de globina son silenciosos. En ambos tejidos, la inmensa mayoría de las interacciones se encontraron con secuencias en el cromosoma 7 y se detectaron muy pocas interacciones LCR con seis cromosomas no relacionados (8, 10, 11, 12, 13, 14) (Figura 12a). Las señales más fuertes en el cromosoma 7 se encontraron en una región de 5-10 Mb centrada alrededor de la posición cromosómica de la  $\beta$  globina, de conformidad con la idea de que las frecuencias de interacción son inversamente proporcionales a la distancia (en pares de bases) entre secuencias de ADN físicamente unidas. No fue posible interpretar las interacciones en esta región de una manera cuantitativa. Se dedujo que estas secuencias cercanas eran, junto con la  $\beta$  globina, tan frecuentes que su gran sobrerrepresentación en las muestras de hibridación saturó las sondas correspondientes. Esto se confirmó cuando se realizaron hibridaciones con muestras diluidas 1:10 y 1:100 y se descubrió que la intensidad de la señal se redujo en sondas fuera y en el borde, pero no dentro de esta región (datos no mostrados).

El procedimiento 4C produjo datos muy reproducibles. La Figura 2b-c muestra proporciones no procesadas de señales 4C sobre señales de hibridación de control para dos regiones de 1,5 Mb en el cromosoma 7, separadas aproximadamente 25 Mb y 80 Mb del gen de la  $\beta$  globina. A este nivel de resolución, los resultados de muestras independientemente procesadas eran casi idénticos. Tanto en hígado como en cerebro fetal, se identificaron grupos de señales positivas en el cromosoma 7, a menudo en localizaciones cromosómicas decenas de megabases separadas de la P globina. Estos grupos típicamente consistían en mínimamente 20-50 sondas con proporciones de señal aumentadas yuxtapuestas en el molde del cromosoma (Figura 12b-c). Cada sonda en la matriz analiza un evento de ligamiento independiente. Además, únicamente dos copias del fragmento de restricción HS2 están presentes por célula, cada una de las cuales solo puede ligarse a otro fragmento de restricción. Por lo tanto, la detección de eventos de ligamiento independientes con 20 o más fragmentos de restricción adyacentes indica fuertemente que el locus correspondiente contacta con la LCR de la  $\beta$  globina en células múltiples.

Para determinar el significado estadístico de estos grupos, se ordenaron datos de experimentos individuales en mapas cromosómicos y se analizaron utilizando un algoritmo de media móvil con un tamaño de ventana de aproximadamente 60 kb. La distribución de la media móvil de datos barajados al azar se utilizó para establecer un valor umbral, que permitió una proporción de descubrimientos falsos del 5 %. Este análisis identificó 66 grupos en hígado fetal y 45 en cerebro que se encontraron reproduciblemente en experimentos por duplicado (Figura 12d-f). De hecho, FISH de alta resolución confirmó que dichos grupos representaban verdaderamente locus que interactuaban frecuentemente (véase más adelante).

Por tanto, la tecnología 4C identifica locus de interacción de largo alcance por detección de eventos de ligamiento independientes con fragmentos de restricción múltiples agrupados en una posición cromosómica.

Se realizó una serie completamente independiente de experimentos 4C con un conjunto de cebadores de PCR inversa diferente que investigó el entorno genómico del gen P mayor localizado ~50 kb cadena abajo de HS2. En hígado fetal, el gen P mayor se transcribe altamente y contacta frecuentemente con la LCR. Se descubrieron grupos casi idénticos de interacciones de largo alcance con  $\beta$  mayor así como con HS2, tanto en hígado fetal como en cerebro, justificando adicionalmente que estos locus frecuentemente contactan con el locus de la  $\beta$  globina (Figura 16).

#### EJEMPLO 5

*El locus activo e inactivo de  $\beta$  globina ocupa distintos entornos genómicos.*

Una comparación entre dos tejidos reveló que el locus de la  $\beta$  globina transcrito de un modo activo en hígado fetal interactúa con un conjunto de locus completamente diferente en comparación con su homólogo transcripcionalmente silencioso en cerebro ( $r = -0,03$ ; correlación de Rango de Spearman) (Figura 12f). Esto excluyó que los resultados estuvieran influenciados por la composición de secuencias de las sondas. En hígado fetal, los segmentos de ADN de interacción se localizaron dentro de una región de 70 Mb centrada alrededor del locus de la  $\beta$  globina, con la mayoría (40/66) localizados hacia el telómero del cromosoma 7. En cerebro fetal, los locus de interacción se encontraron en distancias similares o incluso más largas de la  $\beta$  globina en comparación con cerebro fetal y con la inmensa mayoría de interacciones (43/45) localizadas hacia el centrómero del cromosoma 7. Estos datos demuestran que el locus de  $\beta$  globina activo e inactivo contacta con diferentes partes del cromosoma 7.

Se representaron seis cromosomas distintos (8, 10, 11, 12, 13 y 14) en las micromatrices. Señales de hibridación fuertes en estos cromosomas eran raras, típicamente aparecían aisladas en el molde de ADN lineal y a menudo estaban ausentes en experimentos duplicados. Además, los niveles de media móvil a través de estos cromosomas nunca están reproduciblemente cercanos a los niveles puntuados para el cromosoma 7 (Figura 17). Por tanto, los datos mostraron que el locus de la  $\beta$  globina contactaba en la inmensa mayoría con locus en otra parte del mismo cromosoma, de acuerdo con la localización preferida de este locus dentro de su propio territorio cromosómico. Se

observa que el locus de la  $\alpha$  globina también estaba presente en la matriz (cromosoma 11) y que no puntuaba positivo para la interacción con  $\beta$  globina, de acuerdo con la demostración reciente por FISH de que la  $\alpha$  y  $\beta$  globina de ratón no se encuentran frecuentemente en el espacio nuclear (Brown, J. M *et al.* (2006) *J Cell Biol* 172, 177-87).

5 Para entender mejor la relevancia de las interacciones de largo alcance observadas en el cromosoma 7, se compararon los locus de interacción con las posiciones cromosómicas de genes. Además, se realizó análisis de matriz de expresión Affymetrix para determinar la actividad de transcripción en estas posiciones en los dos tejidos. Aunque el tamaño promedio de las áreas de interacción en hígado fetal y cerebro era comparable (183 kb y 159 kb, respectivamente), se observaron diferencias drásticas en su contenido génico y actividad. En hígado fetal, el 80 %  
10 de los locus que interactúan con  $\beta$  globina contenidos en uno o más genes activamente transcritos, mientras que en cerebro fetal la inmensa mayoría (87 %) no mostró actividad génica detectable (Figura 14). Por tanto, el locus de  $\beta$  globina está embebido en un entorno genómico muy diferente en los dos tejidos. En cerebro, donde el locus no es activo, principalmente se pone en contacto con locus transcripcionales silenciosos localizados hacia el centrómero del cromosoma 7. En hígado fetal, donde el locus es muy activo, interacciona preferentemente con regiones activamente transcritas localizadas más prominentemente hacia el lado telomérico del cromosoma 7. Cabe destacar, que la tecnología 4C identificó tanto Uros como Eraf (~30 Mb lejos de  $\beta$  globina) como genes que interactúan con el locus activo de  $\beta$  globina en hígado fetal, de acuerdo con observaciones previas realizadas por FISH (Osborne, C. S. *et al.* (2004) *Nat Genet* 36, 1065-71 (2004)). Cabe destacar que, en cerebro, se observaron contactos con los otros dos grupos de genes receptores olfativos en el cromosoma 7 que se localizaron en cada lado de, y a 17 y  
20 37 Mb de distancia de la  $\beta$  globina.

Ninguna de todas las regiones transcritas en el cromosoma 7 interacciona con el locus de  $\beta$  globina activo en hígado fetal. Por lo tanto, se buscó un denominador compartido exclusivamente por el locus de interacción pero no con otras regiones activas en hígado fetal. Los genes de  $\beta$  globina, Uros y Eraf ambos eran genes específicos de eritroides que pueden regularse por el mismo conjunto de factores de transcripción y es una idea atractiva que estos factores coordinen la expresión de sus genes diana en el espacio nuclear. Se compararon los datos de la matriz de Expresión Affymetrix de hígado fetal E14.5 con los de cerebro fetal para identificar genes expresados preferencialmente (> 5 veces más) en hígado fetal. Como tales, el 28 % de los genes activos en el cromosoma 7 se clasificaron como "específicos de hígado fetal", de los cuales el 25 % se encontraban en un área de colocalización.  
25 Por tanto, se encontró que no había enriquecimiento de genes específicos de hígado fetal en las áreas de colocalización. Y lo que es más importante, el 49 de 66 (74 %) de las regiones de interacción no contenían ninguna "específica de hígado fetal" y por lo tanto se llegó a la conclusión de que los datos no mostraron ninguna prueba de la expresión coordinada de genes específicos de tejido en el espacio nuclear. Los genes de  $\beta$  globina se transcriben a proporciones altas excepcionales y después se preguntó si el locus interaccionaba preferencialmente con otras regiones de alta actividad transcripcional, siendo genes que se expresan altamente o áreas con una alta densidad de genes activos. Utilizando recuentos Affymetrix como una medida de la actividad génica, se realizó un algoritmo de suma móvil para medir la actividad transcripcional total dentro de regiones de 200 kb alrededor de genes activamente transcritos. Este análisis reveló que la actividad transcripcional alrededor de genes que interactúan no era mayor que alrededor de genes activos que no interactúan en el cromosoma 7 ( $p = 0,9867$ ; suma de Rango de Wilcoxon).  
30  
35  
40

#### EJEMPLO 6

##### *El entorno genómico de un gen conservador se conserva enormemente entre los tejidos*

45 A continuación se investigó si un gen que se expresaba de modo similar en ambos tejidos también cambiaba su entorno genómico. Rad23A es un gen expresado de manera ubicua que reside en un grupo génicamente denso de genes mayormente constitutivos en el cromosoma 8. Tanto en hígado fetal E14.5 como en cerebro, este gen y muchos de sus vecinos directos son activos. Se realizó análisis 4C y se identificaron muchas interacciones de largo alcance con locus hasta a 70 Mb de distancia de Rad23A. Cabe destacar, que las interacciones con Rad23A estaban muy correlacionadas entre hígado y cerebro fetal ( $r = 0,73$ , correlación de Rango de Spearman) (Figura 14a). Una característica compartida de estos locus fue de nuevo que contenían genes activamente transcritos. Por tanto, en ambos tejidos aproximadamente el 70 % contenía al menos un gen activo (Figura 14b-c). Las regiones alrededor de genes de interacción mostraron niveles más altos estadísticamente significativos de actividad génica en comparación con genes activos en otra parte del cromosoma, determinado mediante un algoritmo de suma móvil ( $p < 0,001$  para ambos tejidos). Por tanto, a diferencia del locus de la  $\beta$  globina, el gen Rad23A que está localizado en una región rica en gen preferencialmente interacciona sobre la distancia con otras regiones cromosómicas de actividad transcripcional aumentada. Se observó por FISH que el área cromosómica que contiene Rad23A reside principalmente en el borde de (90 %) o fuera (10 %) de su territorio cromosómico (no publicado, D. Noordermeer, M. Branco, A. Pombo y W. de Laat). Sin embargo, el análisis 4C solamente reveló interacciones intracromosómicas reveladas y ningún área en el cromosoma 7, 10, 11, 12, 13 o 14 cumplía de un modo reproducible los criterios de rigurosidad para interacciones. Por tanto, Rad23A está mayormente implicado en interacciones intracromosómicas que son similares en dos tejidos muy diferentes. Si Rad23A tiene locus adyacentes preferidos en estos cromosomas no relacionados, no interactúan con suficiente frecuencia para detectarse en condiciones utilizadas en el presente documento para tecnología 4C.  
50  
55  
60  
65

EJEMPLO 7*Validación de tecnología 4C por microscopía de alta resolución*

5 Para validar los resultados obtenidos por tecnología 4C, se realizaron experimentos crio-FISH. Crio-FISH es una técnica microscópica recientemente desarrollada, que tiene la ventaja sobre los protocolos FISH 3D actuales de que preserva mejor la ultraestructura nuclear aunque ofrece una resolución mejorada en el eje z a través de la preparación de criosecciones ultrafinas (Branco, M. R. y Pombo, A (2006). PLoS Biol 4, e138). Los datos 4C se verificaron midiendo con qué frecuencia los alelos de  $\beta$  globina o Rad23A (siempre  $n > 250$ ) se colocaron con más de 15 regiones cromosómicas seleccionadas en secciones ultrafinas de 200 nm preparados de cerebro e hígado E14.5. Cabe destacar, que todas las frecuencias de interacción medidas por crio-FISH coincidían perfectamente con los resultados 4C (Figura 16). Por ejemplo, regiones distantes que se identificó que interactuaban con la  $\beta$  globina por tecnología 4C se colocaron más frecuentemente que las áreas intermedias no detectadas por 4C (7,4 % y 9,7 %, frente a 3,6 % y 3,5 %, respectivamente). Además, los dos grupos de genes receptores olfatorios distantes que se identificó por tecnología 4C que interactuaban con  $\beta$  globina en cerebro fetal pero no hígado tuvieron una puntuación de frecuencia de localización respectivamente de 12,9 % y 7 % en cerebro, frente a 3,6 % y 1,9 % en secciones de hígado. En resumen, las frecuencias de colocación medidas por locus positivamente identificados por tecnología 4C eran todas significativamente más elevadas que las frecuencias medidas para locus de fondo ( $p < 0,05$ ; ensayo G). Se llegó a la conclusión de que la tecnología 4C identificaba por completo los locus de ADN de interacción. Finalmente, se utilizó crio-FISH para demostrar que los locus que se identificó que interactuaban con la  $\beta$  globina también contactaban frecuentemente entre sí. Esto era cierto para dos regiones activas separadas sobre una gran distancia cromosómica en hígado fetal (Figura 17) así como para dos grupos de genes OR inactivos lo suficientemente lejanos en el cromosoma en cerebro (Figura 16). Cabe destacar que también se encontraron contactos frecuentes entre estos dos grupos de genes OR distantes en hígado fetal, donde no interactuaba con el grupo de genes OR que contenía el locus de P globina activamente transcrito. Estos datos indicaron que las interacciones nucleares entre distintos grupos de genes OR no eran una peculiaridad del tejido de cerebro fetal analizado. Es tentador especular que dichos contactos espaciales facilitan la comunicación entre muchos genes OR requiere asegurar que únicamente un solo alelo se transcribe por neurona olfativa (Shykind, B. (2005) Hum Mol Genet 14 Spec No 1, R33-9).

EJEMPLO 8*Organización nuclear de dominios de cromatina activos e inactivos*

35 Las observaciones descritas en el presente documento demuestran que no solamente las regiones genómicas activas, sino también inactivas forman distintas regiones en el espacio nuclear que implican muchos contactos de largo alcance, lo que sugiere fuertemente que cada segmento de ADN tiene su propio conjunto de interacciones preferido. Los datos sugieren que cuando el locus de la  $\beta$  globina se activa, esto lleva a un entorno genómico silencioso transcripcional y entra en un área nuclear donde las interacciones con dominios activos están favorecidas. Se espera que dicho reposicionamiento drástico tras la activación transcripcional pueda ser un rasgo solo de genes específicos de tejido que alcanzan un nivel determinado de expresión y, lo que es más importante, se encuentran aislados de otros genes activos en el molde cromosómico lineal, como es el caso de la  $\beta$  globina. Se propone que la red extensiva de interacciones de largo alcance que se identifican tanto en locus genómicos activos como entre inactivos, refleja diferencias entre células en conformaciones cromosómicas más que ser una consecuencia de movimientos dinámicos durante la interfase (Chakalova *et al.* (2005) Nat Rev Genet 6, 669-77 (2005)). Presumiblemente, diferentes grados de descondensación después de la división celular alejan las regiones genómicas activas de la cromatina inactiva (Gilbert, N. *et al.* (2004) Cell 118, 555-66 (2004)) y contacta entre locus distantes de similar composición de cromatina que se estabilizan principalmente a través de afinidades entre proteínas de unión a cromatina. La yuxtaposición espacial entre locus distantes puede ser funcional, pero también puede simplemente ser consecuencia de los patrones de un cromosoma no plegado. Aunque los locus individuales pueden moverse dentro de un volumen nuclear restringido, la conformación general de un cromosoma se mantendría en su mayor parte por todo el ciclo celular y requeriría división celular para el restablecimiento. Esta idea está de acuerdo con estudios de formación de imágenes de células vivas que muestran movimiento restringido en locus de ADN etiquetados en el interior del núcleo (Chubb *et al.* (2002) Curr Biol 12, 439-45 (2002)) y se ajusta bien con estudios que muestran que la información de la posición de la cromatina nuclear está frecuentemente propagada durante la división celular sin conservarse en la población de células (Essers, J. *et al.* Mol Biol Cell 16, 769-75 (2005); Gerlich, D. *et al.* Cell 112, 751-64 (2003)).

EJEMPLO 9*Estudio de viabilidad: la tecnología 4C detecta con precisión deleciones en una muestra de pacientes (Figura 19)*

60 La presencia de una deleción presente en un paciente con leucemia se reveló por 4C utilizando una secuencia de nucleótidos diana que está bien cadena arriba ("hacia la izquierda") 2 Mb (A) o 1,3 Mb (B) desde el primer punto de corte. Obsérvese que las deleciones causan una reducción de las señales de interacción de ADN en la región delecionada, pero también causan un aumento en las frecuencias de interacción ADN:ADN para secuencias

directamente cadena abajo ("a la derecha") del último punto de corte. Esto es particularmente obvio cuando las interacciones con la secuencia B de nucleótido diana se examina estrechamente (véase los dos gráficos inferiores). Basándose en datos 4C se diseñaron cebadores a cada lado de la región delecionada y se identificó el punto de corte secuenciando: el texto plano es secuencia cadena arriba de deleción, en negrita indica un nucleótido insertado, subrayado es la secuencia cadena abajo de la deleción.

#### EJEMPLO 10

*Estudio de viabilidad: la tecnología 4C detecta con precisión una translocación equilibrada en una muestra del paciente (Figura 25).*

Estudio de viabilidad para la detección de translocaciones equilibradas. La detección de la translocación t(1;7) como se describe en (R. Burnett *et al.*, Blood, Vol 84, n.º 4 (15 de agosto), 1994: págs. 1232-1236). Las secuencias de nucleótidos diana flanquean el locus TCRb en el cromosoma 7, representando las señales rojas interacciones ADN:ADN con la secuencia diana que está localizada cadena arriba del locus TCRb, y las señales en azul que representan interacciones de ADN:ADN con la secuencia diana que está localizada cadena arriba del locus TCRb. Se representan las señales de ADN de interacción encontradas en el cromosoma 1. El panel superior muestra la distribución de señal teórica. El panel central e inferior muestra la distribución de señales real. El panel inferior muestra señales a una resolución de sondas individuales yuxtapuestas en el molde del cromosoma. Obsérvese que en el caso de una translocación equilibrada las secuencias de nucleótido diana que flanquean el punto de corte mostrarán un conjunto mutuamente exclusivo de señales de interacción de ADN intercromosómico que directamente se bordean entre sí sobre el molde de cromosoma lineal del cromosoma patrón de translocación. La posición del punto de corte secuenciado (descrito en Burnett *et al.*, 1994) se indica mediante una flecha en el panel inferior.

#### EJEMPLO 11

*Estudio de viabilidad: la tecnología 4C detecta con precisión una translocación no equilibrada en una muestra de paciente (Figura 27).*

Detección de translocaciones no equilibradas. Detección de translocación t(4;7) como se describe en (RJ Galjaard *et al.*, Am J Med Genet A. 30 de agosto de 2003; 121(2): 168-73). Las secuencias de nucleótidos diana se localizan en el cromosoma 7; las señales de ADN de interacción representadas se localizan en el cromosoma 4. Se utilizaron dos secuencias diana localizadas cadena arriba (5') y cadena abajo (3') del punto de corte en el cromosoma 7. Las señales de ADN de interacción localizadas en el cromosoma 4 indican (para ambas secuencias diana en azul). La región entre los grupos de fragmentos de ADN de interacción sobre el cromosoma 4 se han delecionado en este paciente. Parte superior: señales para el cromosoma 4 completo. Panel inferior datos 4C: señales en una región 11,5 MB alrededor de los puntos de corte en el cromosoma 4. Basándose en estos datos 4C, el fragmento de restricción HindIII sobre el cromosoma 4 que contenía el punto de corte de translocación se identificó y se utilizó para mapear el punto de corte por secuenciación.

La secuencia se proporciona en la parte inferior de la figura, en la que la secuencia subrayada es del cromosoma 4, la negrita se encuentra tanto en 7 como en 4 y la secuencia plana es del cromosoma 7.

#### EJEMPLO 12

*Identificación rápida a alta resolución de reordenamientos genómicos equilibrados por tecnología 4C*

##### Sumario

Las técnicas actuales para estudiar la variación genética fracasan o no identifican con precisión reordenamientos cromosómicos equilibrados (inversiones, translocaciones) que frecuentemente se producen en la población humana y que pueden causar enfermedades. En el presente documento se demuestra que la tecnología 4C detecta inversiones y translocaciones equilibradas, así como translocaciones y deleciones no equilibradas, a una resolución (~7 kilobases) que permite la secuenciación inmediata de los puntos de corte. La tecnología 4C se utiliza para caracterizar reordenamientos subyacentes de anomalías congénitas y leucemia. El gen *LMO3* se identificó como un nuevo compañero de translocación del gen  $\beta$  del receptor de linfocitos T (*TCRB*) en leucemia linfoblástica aguda de linfocitos T (T-ALL). Estos resultados establecen que la tecnología 4C es poderosa como una herramienta de investigación clínica nueva para el análisis preciso de reordenamientos genómicos, importante para el diagnóstico de enfermedades, pronóstico y, finalmente, en un cuidado óptimo para el paciente.

##### Introducción

Los reordenamientos cromosómicos (deleciones, amplificaciones, inversiones, translocaciones) pueden ser la causa de enfermedad, particularmente cuando afectan a la expresión de genes debido a la ganancia o pérdida de genes, la creación de genes de fusión o reposición de elementos de ADN reguladores de la transcripción. Los reordenamientos que surgen en la línea germinal pueden dar lugar a defectos congénitos, los de tejido somático

pueden dar como resultado neoplasia. Después de la finalización de los proyectos de secuenciación del genoma humano, se ha convertido ahora en una tarea principal caracterizar variantes estructurales en el genoma humano, ya que está cada vez más claro que la diversidad genómica se produce naturalmente en la población humana y puede relacionarse con susceptibilidad a enfermedad (1-6).

La hibridación genómica comparativa basada en micromatriz (CGH-matriz) es un enfoque de genómica de alto rendimiento ampliamente usado que puede detectar amplificaciones o deleciones cromosómicas en una resolución de algunas kilobases o incluso menor. La CGH se basa en la medición de cambios en el número de copias de ADN tal como se observa en deleciones o amplificaciones y por lo tanto no identifica translocaciones e inversiones que se producen sin pérdida o ganancia de contenido de ADN. Se desconoce con cuánta frecuencia dichos eventos equilibrados ocurren pero se estima que constituyen hasta el 20 % de todas las variaciones estructurales (7). Hasta ahora su detección depende enormemente de estrategias citogenéticas tales como cariotipado cromosómico, que tienen la desventaja de que se pierden eventos (aproximadamente 20 %) y proporcionan una resolución limitada (un máximo de 5-10 megabases). Esto necesita análisis intensivos adicionales para identificar las aberraciones genéticas naturales subyacentes de la enfermedad. En este caso, se demuestra que la tecnología (8) de Captura en placa de Conformación de Cromatina (4C) identifica reordenamientos genómicos tanto equilibrados como no equilibrados a una resolución (~7 kb, véase más adelante) que permite la clonación y secuenciación inmediata de los puntos de corte cromosómicos. Cabe destacar que esta estrategia de alta resolución requiere el uso de una sola micromatriz para explorar todo el genoma y por lo tanto es rentable.

Nuestros resultados iniciales con 4C mostraron que independientemente del plegamiento en la cromatina y las interacciones de largo alcance de un locus determinado, fragmentos de ADN cerca del molde del cromosoma lineal están siempre capturados de un modo eficaz, dando como resultado una señal de hibridación fuerte y frecuentemente incluso saturada para sondas dentro de una región de al menos 5-10 megabases alrededor de la secuencia diana. Esto está de acuerdo con la idea de que los segmentos de ADN localizados más cerca de una fibra de cromatina flexible interactuarán más frecuentemente (9). Los fragmentos locales también están capturados mucho más eficazmente que segmentos a megabases de distancia que forman bucle frecuentemente hacia la secuencia diana (8). La captura al azar de fragmentos de restricción es poco habitual, como se demuestra por la muy baja frecuencia de señales de sondas localizadas en cromosomas no relacionados. Por tanto, la tecnología 4C permite la reconstrucción de mapas físicos de moldes de cromosomas alrededor de secuencias diana y por tanto debe ser capaz de identificar cambios en estos mapas como resultado de reordenamientos genómicos.

#### Materiales y Métodos

##### *Preparación de las muestras*

Las muestras de pacientes T-ALL y muestras de linfocitos T de control sano se manipularon como se ha descrito previamente (Vlierberghe *et al.*, *Leukemia* 20, 1245 (Jul, 2006); Simonis *et al.*, *Nat Genet* 38, 1348 (Nov, 2006)). La línea celular transformada con EBV del paciente PAP se cultivó y manipuló como se ha descrito anteriormente (Simonis *et al.*, *Nat Genet* 38, 1348 (Nov, 2006), Galjaard y col., *Am J Med Genet A* 121, 168 30, 2003).

##### *Diseño de matriz 4C*

Se diseñaron sondas de 60 pb dentro de 100 pb de un sitio HindIII, utilizando criterios descritos anteriormente que por ejemplo seleccionan únicamente secuencias de ADN exclusivas (Simonis *et al.*, *Nat Genet* 38, 1348 (Nov, 2006)). Para ser capaz de abarcar todo el genoma con 400.000 sondas que se ajustan a una micromatriz Nimblegen, se realizó una selección. Los números de sondas se redujeron primero manteniendo solamente una sonda por fragmento HindIII, en lugar de una a cada lado. En segundo lugar, las sondas se seleccionaron de tal manera que el espaciado de las sondas era tan igual como fuera posible por todo el genoma.

##### *Análisis 4C*

Se realizaron análisis 4C como se ha descrito anteriormente (2), utilizando las siguientes secuencias cebadoras:

extremo 5' de *TCRB* CATGAAGAAACGAGCACCC CCTTGATGTTTCTCCCTTTACC

extremo 3' de *TCRB* TGTCAGGCTCTTCTCCTACAC GTCGTCCAGAACACTCACC

t(4;7) Centromérico AATCCAGGGCTACTTCCAG CCGTGATGCTATCTGCCA

t(4;7) Telomérico TGTTGGAAGACCAGGTGAAG TGTCGTGGAAGCGAGTG

Deleción 9 CAATCCCAGATACATTCTCATACAAATACTTTCCAAGACTGG AC

3' de *TCRA* GAATATGTTATGCTTGATCC TTCCATGAGAGAAGTCTAG



Los datos 4C se visualizaron utilizando programa informático SignalMap. Para crear dibujos de vista de cromosomas completos de los datos 4C, se calculó una media móvil con un tamaño de ventana de 29 sondas utilizando el paquete R (<http://www.r-project.org>).

#### 5 *Secuenciación final emparejada con fragmento de restricción*

Se digirieron 10 µg de ADN genómico en primer lugar en 500 µl con 10 U de una enzima que reconoce 6 bases (HindIII, BglII o EcoRI) (37 °C durante 2 horas). Las muestras se purificaron mediante extracción con fenol-cloroformo y precipitación con etanol. Posteriormente, las muestras se ligaron en 2 ml con 40 U de ligasa (Roche) durante 4 horas a 16 °C y 30 minutos a 20 °C.

Las muestras ligadas se purificaron mediante extracción con fenol-cloroformo y precipitación con etanol.

Se realizó una segunda digestión con una enzima de restricción que reconocía 4 bases (por ejemplo, NlaIII o DpnII) en las mismas condiciones que las descritas para la enzima de reconocimiento de 6 bases. El ligamiento posterior también fue como se ha descrito anteriormente. Las muestras se purificaron mediante extracción con fenol-cloroformo y precipitación con etanol. Los fragmentos seleccionados se amplificaron por PCR de 50-100 ng de ADN, utilizando las siguientes condiciones: 94 °C durante 3 minutos, seguido de 30 ciclos de 15 segundos a 94 °C, 1 minuto a 55 °C y 2 minutos a 72 °C en una etapa final de 7 minutos a 72 °C.

### Resultados

#### *Identificación de puntos de corte de translocación*

Para ensayar esto, se aplicó primero tecnología 4C a la línea celular HSB-2 T-ALL, que contenía una translocación recíproca t(1; 7) (p35;q35) entre el locus del *receptor*  $\beta$  de *linfocitos T* (TCRB) en el locus 7q35 y *LCK* en 1p35 (10). Se realizaron dos experimentos 4C independientes, cada uno de ellos analizando interacciones con un fragmento de restricción diferente en el cromosoma 7 localizado a cada lado del punto de corte del locus *TCRB* y a una distancia de 462 kb y 239 kb, respectivamente. En ambos casos, se observaron fuertes señales de hibridación alrededor del locus *TCRB* en el cromosoma 7 en una muestra de control sana y en la muestra HSB-2. La muestra de control no mostró ninguna señal, o mostró señal de fondo, en todos los otros cromosomas (Fig. 30A, Fig. 33). Por otro lado, la muestra HSB-2 mostró señales adicionales muy altas específicamente a través de una región de megabases en 1p35 (Fig. 30B). Estas señales representaban fragmentos de restricción en el cromosoma 1 capturados por los dos fragmentos en el cromosoma 7 en HSB-2, indicando que partes de estos cromosomas se pusieron en estrecha proximidad. La secuencia diana TCRB más telomérica en el cromosoma 7 capturó fragmentos de restricción hacia el lado centromérico del gen *LCK* en el cromosoma 1. A la inversa, la secuencia diana *TCRB* más centromérica capturó fragmentos hacia el lado telomérico de *LCK*. Esto está de acuerdo con la orientación de la translocación t(1;7). Además, los primeros fragmentos de restricción capturados en el cromosoma 1 en ambos experimentos flanquean directamente el punto de corte cromosómico identificado previamente. Por tanto, 4C localiza puntos de corte de translocación en una posición entre el par de sondas que representa la transición desde los fragmentos de restricción capturados a no capturados. En este caso, los dos experimentos 4C que analizan interacciones con secuencias diana del cromosoma 7 identifican cada uno el punto de corte en el cromosoma 1 implicado en la translocación equilibrada t(1;7) dentro de 27 kb.

#### 45 *Identificación de inversiones*

A continuación, se ensayó si 4C también podía identificar inversiones. Se aplicó tecnología 4C a una muestra de un paciente T-ALL pediátrico que, basándose en estudios de expresión de micromatriz y FISH, se sospechaba que llevaba una inversión en el cromosoma 7, inv(7)(p15q35). Esta anomalía condujo al reordenamiento del locus *TCRB* en el grupo de genes *HOXA*, como se ha descrito previamente para otros pacientes (11, 12). De nuevo se realizaron dos experimentos, utilizando el mismo conjunto de fragmentos diana *TCRB* que identificaron la translocación t(1;7) descrita anteriormente. Las dos secuencias diana capturaron eficazmente muchos fragmentos que abarcaban megabases de ADN del otro lado del cromosoma 7 alrededor del locus *HOXA* solo en la muestra de 4 pacientes (Fig. 30C-D). Además, cada secuencia diana capturó una región cromosómica distinta alrededor del grupo de genes *HOXA*, indicando que los dos fragmentos están ligados a diferentes partes del cromosoma 7. El fragmento diana *TCRB* más 3' capturó los fragmentos *HOXA* más 3' mientras que el fragmento diana *TCRB* más 5' capturó los fragmentos *HOXA* más 5', revelando de este modo una inversión entre los locus. Las dos regiones capturadas alrededor de *HOXA* se flanquean directamente entre sí, mostrando que la inversión estaba equilibrada y no venía acompañada por una pérdida (extensiva) de secuencias *HOXA*. Las dos sondas que marcan la transición entre los fragmentos no capturados y capturados revelaron la posición del punto de corte, que se localizaba en una región de 6 kb cercana del gen *HOXA9* del grupo *HOXA* (Fig. 30D). La confirmación de esta región que de hecho lleva el punto de corte se obtuvo por secuenciación, utilizando una estrategia de secuenciación de extremos emparejados con fragmentos de restricción (Fig. 34). Por lo tanto se llega a la conclusión de que la tecnología 4C es la primera estrategia genómica de alto rendimiento que puede detectar translocaciones e inversiones equilibradas. La resolución proporcionada por la tecnología 4C permite la clonación inmediata de los puntos de corte. La tecnología 4C es por lo tanto la primera técnica que puede detectar eventos genéticos equilibrados a dicha alta resolución.

*Identificación de translocaciones desequilibradas*

El potencial de la tecnología 4C se exploró adicionalmente aplicándolo a una línea celular transformada con EBV procedente de un paciente con Polidactilia Postaxial (PAP). PAP es un trastorno hereditario dominante autosómico caracterizado por dígitos ulnares o fibulares extra. Las células del paciente se caracterizaron previamente por cariotipado y FISH que contenía una translocación desequilibrada entre los cromosomas 4 y 7 con una microdelección, t(4;7)(p15.2;q35). Sin embargo, la resolución limitada de FISH impidió la definición del grado de la delección y las posiciones exactas de los puntos de corte (13). Se realizaron dos experimentos 4C, cada uno de ellos analizando interacciones de ADN con un fragmento diana localizado en otro lado de la parte reordenada del cromosoma 7, uno de los cuales estaba a 4 megabases de distancia del punto de corte más cercano (véase más adelante). En ambos experimentos, se capturaron fragmentos genómicos no solo en el cromosoma 7 sino también en el cromosoma 4, en 4p15.2 (Fig. 31A, Fig. 35). A diferencia de lo que se observaba para la translocación equilibrada, los fragmentos del cromosoma 4 capturados por las dos secuencias diana no estaban directamente adyacentes entre sí. Un punto de corte se localizó en la posición 17,28 Mb (NCBI 36) y el otro se encontró en la posición 20,08 Mb del cromosoma 4, mostrando que la translocación t(4;7) venía acompañada por una delección de 2,8 Mb en el cromosoma 4. Para verificar que las transición de fragmentos de restricción capturados a no capturados de hecho marcan las posiciones de los puntos de corte, el punto de corte que era menos obvio a partir de los datos 4C, localizado en 20,08 Mb, se clonó y secuenció (Fig. 31B). Esto confirmó la posición del punto de corte del cromosoma 4 en 20,08 Mb dentro del gen *SLIT2* y reveló que estaba reordenado con una secuencia intergénica en el cromosoma 7 que estaba a 4 Mb de distancia de la secuencia diana usada para identificar el punto de corte. Esto muestra que las secuencias diana 4C pueden capturar fragmentos de ADN e identificar reordenamientos incluso cuando los puntos de corte están a varias megabases de separación. Cuando los análisis 4C se dirigen a ambos lados de un punto de corte genómico, este puede identificar inmediatamente si una translocación o inversión está equilibrada o viene acompañada por reordenamientos adicionales tales como una delección (es decir, no equilibrada).

*Identificación de delecciones*

A continuación se investigó si la tecnología 4C podía identificar una delección que no estuviese asociada con una translocación. Para ello, se analizó otra muestra de un paciente T-ALL pediátrico, basándose en datos de matriz-CGH, que contenían una delección homocigota de los locus *p15/p16* en el cromosoma 9p21. El tamaño exacto de la delección y los puntos de corte reales no se conocen. Se definió un fragmento diana localizado a ~ 2 Mb de distancia de uno de los puntos de corte estimados. Como se esperaba, se observó una región que carecía de señales sonda, demarcando el área delecionada (Fig. 31C). Se observó un aumento en las señales de hibridación para la región inmediatamente cadena abajo de la delección en la muestra del paciente frente a la muestra de control sano. Esto es debido a que la región está en estrecha proximidad en el molde lineal con el fragmento diana debido a la delección. Basándose en los datos 4C, se diseñaron cebadores PCR que flanqueaban la región delecionada ~2 Mb que permitían la amplificación a través de los puntos de corte; la secuenciación de los productos de PCR confirmó las posiciones de los dos puntos de corte que flanqueaban la delección (Fig. 31D). Se llega a la conclusión de que la tecnología 4C puede identificar delecciones homocigotas. Las delecciones revelan en sí mismas regiones que contienen señales de hibridación reducida en combinación con más secuencias cadena abajo que muestran unas señales de hibridación aumentadas.

*Muestras de pacientes no caracterizadas*

La tecnología 4C se aplicó para explorar muestras de pacientes T-ALL no caracterizadas para reordenamientos genéticos asociados con el locus *TCRB* o el locus del Receptor  $\alpha$  de linfocitos T (*TCRA*). En T-ALL, las translocaciones cromosómicas frecuentemente se producen durante recombinación VDJ tentativa de los locus *TCR*. Las muestras de cinco pacientes T-ALL, que previamente se había demostrado que no llevaban ninguna de las anomalías genéticas recurrentes asociadas con T-ALL (datos no mostrados), se analizaron por tecnología 4C, utilizando una secuencia diana cerca de *TCRB* y una cerca de *TCRA*. Ninguna de las muestras mostró reordenamientos con el locus *TCRA* (datos no mostrados) y cuatro de cinco muestras de pacientes también mostraron que no había reordenamientos con *TCRB* (fig. 36), que se confirmó posteriormente por FISH. Sin embargo, la muestra de un paciente mostró una translocación entre *TCRB* y el brazo p del cromosoma 12 (Fig. 32A; fig. 36). Además, se descubrió que el paciente que llevaba una gran delección en el cromosoma 12, que se confirmó mediante un experimento de oligomatrix-CGH (datos no mostrados). Esta delección se localizaba a ~3 Mb de distancia del punto de corte de translocación, mostrando de nuevo que las secuencias diana 4C podían identificar reordenamientos sobre grandes distancias. La translocación t(7;12)(q35;p12.3) no se describió antes en T-ALL. Las dos sondas en el cromosoma 12 que marcaban la transición entre los fragmentos de restricción capturados y no capturados tenían estaban a una distancia de 6 kb y se localizaban justo cadena abajo del gen solamente de dominio Lim *LMO3*. Se utilizó secuenciación de extremos emparejados con fragmentos de restricción para confirmar que estas sondas demarcaban la región que contenía el punto de corte. La secuenciación de los puntos de corte presentes en ambos cromosomas derivados demostró que el cromosoma 12 se disponía sin la pérdida de un solo nucleótido, mientras que la rotura del cromosoma 7 venía acompañada por una delección de casi 400 kb de secuencias *TCRB* (Fig. 32B). Esta delección se debía posiblemente a eventos de delección asociados con una recombinación VDJ tentativa en el locus *TGRB*. Ambos puntos de corte también contenían pares de bases intermedios de origen desconocido (de 4 y 18 pb, respectivamente), que podían representar los nucleótidos al azar

que normalmente se incorporaban también durante la recombinación VDJ. De manera interesante, las posiciones de translocación del potenciador de *TCRB* 70 kb cadena abajo del gen *LMO-3* (Fig. 32C), que es comparable a su posición normal relativa a *TCRB*. Los datos de expresión de micromatriz mostraron que aunque *LMO3* está normalmente fuera en las muestras de pacientes T-ALL, el gen se expresa altamente en esta muestra T-ALL (Fig. 37). Los miembros de la familia de proteínas *LMO-1* y *LMO-2*, pero no *LMO3*, se han descubierto previamente como compañeros de translocación oncogénicos de los locus *TCR* en T-ALL. De manera interesante, se descubrió recientemente que *LMO3* actuaba como un oncogén en neuroblastoma (14). Por tanto, la tecnología 4C aplicada a la exploración de muestras de pacientes no caracterizadas para aberraciones genéticas condujo al descubrimiento de una translocación previamente no detectada y estableció *LMO-3* como un supuesto nuevo oncogén de linfocitos T.

### Conclusiones

Estos datos establecen tecnología 4C como la primera estrategia genómica que puede identificar anomalías genéticas equilibradas tales como translocaciones e inversiones recíprocas. Además, es obvio que la tecnología 4C puede identificar deleciones homocigotas y deleciones asociadas con translocaciones. La tecnología 4C también puede identificar eventos no equilibrados heterocigotos basándose en cambios en números de copias de ADN capturados. Una ventaja principal de la tecnología 4C sobre las estrategias de secuenciación de extremos emparejados (15) o incluso secuenciación de todo el genoma (de alto rendimiento) es que la identificación de reordenamientos equilibrados no se basa en la captura de fragmentos de secuencias sencillas que llevan el punto de corte; en lugar de ello, la tecnología 4C identifica reordenamientos equilibrados basándose en la captura de muchos fragmentos que cubren varias megabases a través del punto de corte. Por ejemplo, se identifican translocaciones recíprocas basándose en la captura de muchos fragmentos que se localizan en una parte de un cromosoma por una secuencia diana localizada en otro cromosoma. Esto hace que la tecnología sea más poderosa que otras estrategias. La resolución de la tecnología 4C es alta y permite la clonación inmediata y la secuenciación de puntos de corte genéticos, como se demostró para cuatro reordenamientos diferentes. Incluso puntos de corte localizados a varias megabases de distancia de la secuencia diana pueden identificarse fácilmente. La resolución es esencialmente idéntica al tamaño promedio de los fragmentos creados por la enzima de restricción. El uso de una enzima que crea fragmentos más pequeños debe incluso aumentar adicionalmente la resolución de esta técnica. La tecnología 4C puede aplicarse para todos los tipos de células tales como células de cáncer de sangre, tumores sólidos, amniocitos recogidos para diagnóstico prenatal, etc., siempre que puedan aislarse (o cultivarse) suficientes células que contengan ADN intacto. Actualmente se comienza con aproximadamente 10 millones de células, pero esta cantidad puede reducirse, ya que se hibridan uniones de ligamiento amplificadas por PCR que originan de ~0,5 millones de equivalentes de genoma. La tecnología 4C actual requiere la selección de una secuencia diana y es por lo tanto particularmente adecuada para la exploración de reordenamientos cerca de locus frecuentemente implicados en una enfermedad, tal como los genes receptores de linfocitos T en T-ALL o los locus de cadena ligera y pesada del receptor de linfocitos B (*BCR*) en linfomas humanos. La técnica es también muy útil para el mapeo fino de reordenamientos mal caracterizados, por ejemplo de translocaciones o inversiones que se descubrió que se basaban en el cariotipo cromosómico.

### EJEMPLO 13

#### *Secuenciación 4C múltiple*

#### Diseño experimental

Este ejemplo se basa en una de las diversas tecnologías de secuenciación a alto rendimiento (Solexa), pero puede modificarse para ajustarse a otras plataformas. Para utilizar la secuenciación Solexa (Illumina) para el análisis de los resultados de 4C, la secuenciación se dirige a uniones de ligamiento. Por lo tanto, se diseñaron cebadores de PCR inversa para cada secuencia diana de tal manera que estaban cerca de los sitios de reconocimiento de la enzima de restricción primaria y secundaria analizados. En este caso, se diseñaron cebadores de PCR inversos cortos (18-meros) que cada uno se solapa parcial o completamente con los sitios de reconocimiento de la enzima de restricción primaria (*HindIII*) y secundaria (*DpnII*) analizados. Las secuencias del cebador de secuencia y adaptadores Solexa se añadieron como salientes 5' a cebadores de PCR inversa (Fig. 38).

Tres secuencias diana (cebos) se seleccionaron en el cromosoma 7 humano, respectivamente en las posiciones 85, 105 y 139 Mb. Los conjuntos de cebadores utilizados para amplificar los fragmentos capturados por cada uno de estos cebos son los siguientes:

85 Mb:

*DpnII*-cebador: atgtgactcctctagatc

*DpnII*-cebador con adaptador: aatgatacggcgaccaccgaacactctttccctacacgacgctctccgatct-atgtgactcctctagatc

*HindIII*-cebador: ccctgaacctcttgaagct

*HindIII*-cebador con adaptador: caagcagaagacggcatacga-ccctgaacctcttgaagct

105 Mb:

*DpnII*-cebador: cggcctccaattgtgatc  
*DpnII*-cebador con adaptador: aatgatacggcgaccaccggaacactctttccctacacgacgctctccgatct-cggcctccaattgtgatc  
*HindIII*-cebador: gaattgcttttggaagcctt  
*HindIII*-cebador con adaptador: caagcagaagacggcatacga-gaattgcttttggaagcctt

139 Mb:

*DpnII*-cebador: ttttagccctgacagatc  
*DpnII*-cebador con adaptador: aatgatacggcgaccaccggaacactctttccctacacgacgctctccgatct-ttttagccctgacagatc  
*HindIII*-cebador: agtcaaacataagcctaagc  
*HindIII*-cebador con adaptador: caagcagaagacggcatacga-agtcaaacataagcctaagc

Cada conjunto de cebadores (con adaptadores) se usó en una reacción de PCR distinta; se realizaron 3 reacciones de PCR (utilizando 200 ng de molde por reacción) en condiciones estándar (descritas en Simonis *et al.*, Nature Methods 2007, vol. 4, 895-901). Se preparó molde 4C a partir de la línea celular T-ALL HSB-2, que contenía una translocación recíproca t(1;7)(p35;q35) entre el locus beta del receptor de linfocitos T (TCRB) en 7q35 y el locus LCK en 1p35 (Burnett, R. C., *et al.*, Blood 84, 1232-6 (1994)). En la Figura 39 se muestran los productos de la PCR.

Posteriormente, las reacciones de PCR se agruparon por conjunto de cebadores y se purificaron sobre columnas de limpieza Amersham. La concentración de ADN se midió y se mezclaron cantidades equimolares de material amplificado por cada conjunto de cebadores. Esta mezcla se analizó mediante secuenciación Solexa.

Resultados de la secuenciación Solexa

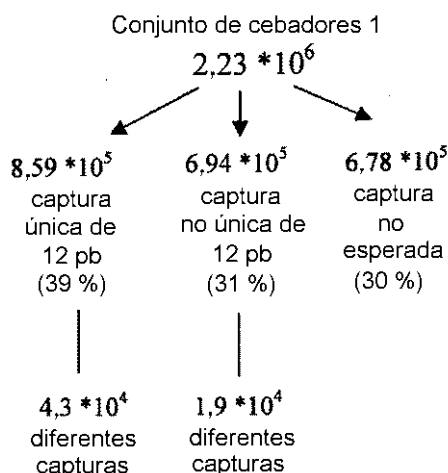
Número total de lecturas (un carril):  $4,9 * 10^6$  secuencias.  
 93 % de las secuencias comienzan con una de las secuencias de cebadores esperadas.

Cuando se analizan los 12 primeros pares de bases (pb) a través de la unión de ligamiento *DpnII* y se comparan estos 12 pares de bases con una base de datos local que contiene todos los fragmentos genómicos de 12 pb que flanquean los sitios *DpnII* relevantes en el genoma (es decir aquellos directamente adyacentes a un sitio *HindIII*) se encontró que: el 37 % contienen captura de 12 pb únicos, 34 % contienen captura de 12 pb no únicos y 29 % contienen captura de 12 pb no esperados.

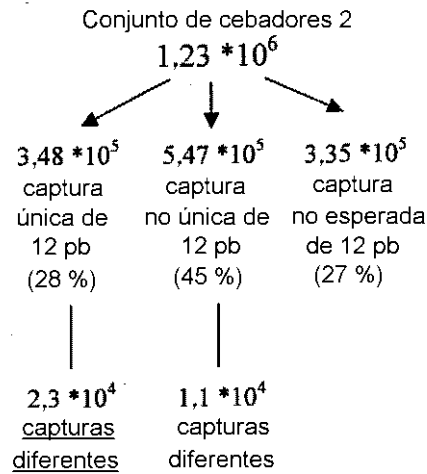
De las  $4,9 * 10^6$  secuencias: 34 % contenían cebador + captura única, 32 % contenía cebador + captura no única. Por lo tanto en total el 66 % contenía como se esperaba una primera secuencia de 30 pb.

Por conjunto de cebadores, el número total de secuencias y la naturaleza de estas secuencias es el siguiente:

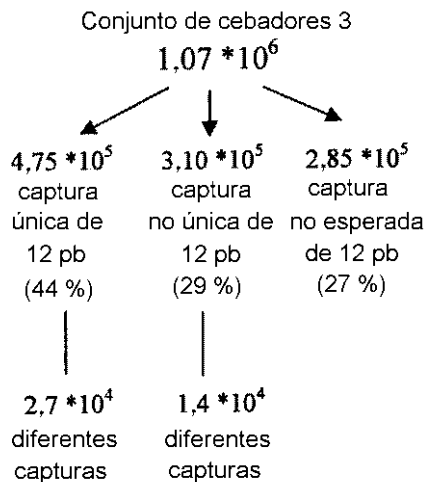
Conjunto de cebadores 1 (85 Mb):



Conjunto de cebadores 2 (105 Mb):



Conjunto de cebadores 3 (139 Mb):



5

Cuando las secuencias capturadas se representan en sus posiciones cromosómicas, los datos se muestran en las Figuras 40 y 41.

10 En resumen, los datos proporcionan un estudio de viabilidad para secuenciación 4C múltiple. Para cada secuencia, el cebo y la secuencia capturada pueden identificarse. La técnica puede mejorarse adicionalmente:

1. La secuenciación puede dirigirse mejor al lado *HindIII* (en lugar de al lado *DpnII*); esto impide la lectura de eventos de ligamiento al azar que se producen durante la segunda etapa de ligamiento.
- 15 2. El uso menos frecuente de cortadores (7 u 8 cortadores en lugar de 6 cortadores) aumentará la distancia genómica capturada (abarcada) por cada cebo.
3. El uso de una enzima de restricción que no digiere ADN repetitivo (o que está relativamente infrarrepresentada en ADN repetitivo) aumentará el número de lecturas interpretables.

20 Aspectos adicionales 1

1. Un método para analizar la frecuencia de interacción de una secuencia de nucleótidos diana con una o más secuencias de nucleótidos (por ejemplo uno o más locus genómicos) que comprenden el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

25

2. Un método para identificar una o más interacciones ADN-ADN que son indicativas de una patología particular o de un estado portador que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

30

3. Un método de diagnóstico o pronóstico de una enfermedad o síndrome o estatus portador causado por o asociado con un cambio en un ADN-ADN que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

5 4. Un método de ensayo para identificar uno o más agentes que modulan una interacción ADN-ADN que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

10 5. Un método para detectar la localización de un punto de corte (por ejemplo una translocación) que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

15 6. Un método para detectar la localización de una inversión que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

20 7. Un método para detectar la localización de una deleción que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

8. Un método para detectar la localización de una duplicación que comprende el uso de una secuencia de nucleótidos o una matriz de sondas o un conjunto de sondas o una matriz como se describe en el presente documento.

25 9. Un método para analizar la frecuencia de interacción de una secuencia de nucleótidos diana con una o más secuencias de nucleótidos de interés (por ejemplo uno o más locus genómicos) que comprende las etapas de:

- (a) proporcionar una muestra de ADN entrecruzado;
- (b) digerir el ADN entrecruzado con una enzima de restricción primaria;
- (c) ligar las secuencias de nucleótidos entrecruzadas;
- (d) invertir el entrecruzamiento;
- (e) opcionalmente digerir las secuencias de nucleótidos con una enzima de restricción secundaria;
- (f) opcionalmente ligar las secuencias de nucleótidos;
- (g) amplificar la una o más secuencias de nucleótidos de interés que están ligadas con la secuencia de nucleótidos diana utilizando al menos dos cebadores oligonucleotídicos, en el que cada cebador se hibrida con una secuencia de ADN conocida que flanquea las secuencias de nucleótidos de interés;
- (h) hibridar la secuencia o secuencias amplificadas con una matriz o secuenciar las secuencias amplificadas; y
- (i) determinar la frecuencia de interacción entre secuencias de ADN.

40 Aspectos adicionales 2

Incluso más aspectos adicionales de la presente invención se exponen a continuación en los párrafos numerados.

45 1. Una secuencia de nucleótidos circularizada que comprende una primera y una segunda secuencia de nucleótidos separada por sitios de reconocimiento de enzimas de restricción primarias en el que cada primera secuencia de nucleótidos es una secuencia de nucleótidos diana y dicha segunda secuencia de nucleótidos puede obtenerse mediante el entrecruzamiento de ADN genómico.

50 2. La secuencia de nucleótidos circularizada de acuerdo con el párrafo 1, en la que la secuencia de nucleótidos diana se selecciona del grupo que consiste en un promotor, un potenciador, un silenciador, un aislante, una región de unión a matriz, una región de control de locus, una unidad de transcripción, un origen de la replicación, un punto caliente de recombinación, un punto de corte de translocación, un centrómero, un telómero, una región densa en genes, una región pobre en genes, un elemento repetitivo y un sitio de integración (vírico).

55 3. La secuencia de nucleótidos circularizada de acuerdo con el párrafo 1, en la que la secuencia de nucleótidos diana es una secuencia de nucleótidos que está asociada a o causa una enfermedad, o está localizada a menos de 15 Mb en un molde de ADN lineal de un locus que está asociado con una enfermedad o la causa.

60 4. La secuencia de nucleótidos circularizada de acuerdo con cualquiera de los párrafos 1-3, en la que la secuencia de nucleótidos diana se selecciona del grupo que consiste en: *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, *IGH*, *LYL1*, *TAL1*, *TAL2*, *TCR $\alpha$ / $\delta$* , *TCR $\beta$* , *HOX* u otros locus asociados con la enfermedad como se describe en "Catalogue of Unbalanced Chromosome Aberrations in Man" 2ª edición. Albert Schinzel. Berlín: Walter de Gruyter, 2001. ISBN 3-11-011607-3.

5. La secuencia de nucleótidos circularizada de acuerdo con cualquiera de los párrafos 1-4 en la que el sitio de reconocimiento de la enzima de restricción primaria es un sitio de reconocimiento de 6-8 pb, preferentemente seleccionado del grupo que consiste en *Bgl*III, *Hind*III, *Eco*RI, *Bam*HI, *Spe*I, *Pst*I y *Nde*I.
- 5 6. La secuencia de nucleótidos circularizada de acuerdo con cualquiera de los párrafos anteriores, en la que el sitio de reconocimiento de la enzima de restricción secundaria es un sitio de reconocimiento de secuencia de 4 o 5 pb.
- 10 7. La secuencia de nucleótidos circularizada de acuerdo con cualquiera de los párrafos anteriores, en la que el sitio de reconocimiento de la enzima de restricción secundaria se localiza a más de aproximadamente 350 pb desde el sitio de restricción primario.
- 15 8. La secuencia de nucleótidos circularizada de acuerdo con cualquiera de los párrafos anteriores, en la que la secuencia de nucleótidos está marcada.
9. Un método de preparación de una secuencia de nucleótidos circularizada que comprende las etapas de:
- (a) proporcionar una muestra de ADN entrecruzado;
- 20 (b) digerir el ADN entrecruzado con una enzima de restricción primaria;
- (c) ligar las secuencias de nucleótidos entrecruzadas para la circularización;
- 25 10. Un método de acuerdo con el párrafo 9, en el que la secuencia de nucleótidos entrecruzada se amplifica utilizando PCR.
11. Un método de acuerdo con el párrafo 10, en el que la secuencia de nucleótidos entrecruzada se amplifica utilizando PCR inversa.
- 30 12. Un método de acuerdo con el párrafo 10 o párrafo 11, en el que se utiliza el Sistema PCR de Molde Largo Expandido (Roche).

TABLA 2

Interacción	en 4C	N	% de solapamiento	en Crio-FISH	Valor de P
B-globina - Cro.7 73,1 Mb	+	258	7,4	+	P < 0,001
B-globina - Cro.7 80,1 Mb (OR)	-	254	3,6	-	
B-globina - Cro.7 118,3 Mb	-	255	3,5	-	
B-globina - Cro.7 127,9 Mb (Uros)	+	259	6,6	+	P < 0,001
B-globina - Cro.7 130,1 Mb	+	413	9,7	+	P < 0,001
B-globina - Cro.7 135,0 Mb (OR)	-	261	1,9	-	
B-globina - D7Mit21	x	258	0,4	-	
Cro.7 80,1 Mb - Cro.7 135,0 Mb	x	253	5,9	+	P < 0,05
Cro.7 73,1 Mb - Cro.7 130,1 Mb	x	254	5,5	+	P < 0,05
Rad23A - Cro. 8 21,8 Mb	+	255	5,9	+	P < 0,05
Rad23A - Cro. 8 122,4 Mb	+	261	8	+	P < 0,001
B-globina - Cro.7 73,1 Mb	-	256	3,9	-	
B-globina - Cro.7 80,1 Mb (OR)	+	256	12,9	+	P < 0,001
B-globina - Cro.7 118,3 Mb	-	242	4,1	-	
B-globina - Cro.7 130,1 Mb	-	263	3	-	
B-globina - Cro.7 135,0 Mb (OR)	+	256	7	+	P < 0,05
B-globina - D7Mit21		258	6,2	+	P < 0,05
Cro.7 80,1 Mb - Cro.7 135 Mb		261	5	+	P < 0,1
Rad23A - Cro. 8 21,8 Mb	-	260	3,8	-	
Rad23A - Cro. 8 122,3 Mb	+	258	8,1	+	P < 0,001

Referencias

- 35 Blanton J, Gaszner M, Schedl P. 2003. Protein:protein interactions and the pairing of boundary elements in vivo. *Genes Dev* 17: 664-75.
- 40 Dekker, J., Rippe, K., Dekker, M., y Kleckner, N. 2002. Capturing chromosome conformation. *Science* 295: 1306-11.
- Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* 18: 2485-90.

Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T. 2005. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet* 37: 31-40.

5 Murrell A, Heeson S, Reik W. 2004. Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nat Genet* 36:889-93.

Palstra, R. J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., y de Laat, W: 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet* 35: 190-4.

10 Patrinos, G. P., de Krom, M., de Boer, E., Langeveld, A., Imam, A.M.A, Strouboulis, J., de Laat, W., y Grosveld, F. G. (2004). Multiple interactions between regulatory regions are required to stabilize an active chromatin hub. *Genes & Dev.* 18: 1495-1509.

15 Spilianakis CG, Flavell RA. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 5: 1017-27.

Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F., y de Laat, W. 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell* 10: 1453-65.

20 Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA. 2005.

Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell.* 17: 453-62

25 Referencias para el ejemplo 12

1. *Nature* 447, 661 (7 de junio, 2007).

2. D. F. Easton *et al.*, *Nature* 447, 1087 (28 de junio, 2007).

3. L. Feuk, A. R. Carson, S. W. Scherer, *Nat Rev Genet* 7, 85 (febrero, 2006).

30 4. A. J. Sharp, Z. Cheng, E. E. Eichler, *Annu Rev Genomics Hum Genet* 7, 407 (2006).

5. A. J. Iafrate *et al.*, *Nat Genet* 36, 949 (septiembre 2004).

6. M. R. Mehan, N. B. Freimer, R. A. Ophoff, *Hum Genomics* 1, 335 (agosto, 2004).

7. E. E. Eichler *et al.*, *Nature* 447, 161 (10 de mayo, 2007).

35 8. M. Simonis *et al.*, *Nat Genet* 38, 1348 (noviembre, 2006).

9. K. Rippe, *Trends Biochem Sci* 26, 733 (diciembre, 2001).

10. R. C. Burnett, M. J. Thirman, J. D. Rowley, M. O. Diaz, *Blood* 84, 1232 (15 de agosto, 1994).

11. F. Speleman *et al.*, *Leukemia* 19, 358 (marzo, 2005).

12. J. Soulier *et al.*, *Blood* 106, 274 (1 de julio, 2005).

13. R. J. Galjaard *et al.*, *Am J Med Genet A* 121, 168 (30 de agosto, 2003).

40 14. M. Aoyama *et al.*, *Cancer Res* 65, 4587 (1 de junio, 2005).

15. E. Tuzun *et al.*, *Nat Genet* 37, 727 (julio, 2005).

Todas las publicaciones mencionadas en la memoria descriptiva anterior se incorporan por referencia en el presente documento. Diversas modificaciones y variaciones de los métodos descritos y del sistema de la invención serán obvias para los expertos en la técnica sin alejarse del alcance y espíritu de la misma. Aunque la invención se ha descrito junto con realizaciones específicas preferidas, debe entenderse que la invención reivindicada no debe limitarse indebidamente a dichas realizaciones específicas. De hecho diversas modificaciones de los modos descritos para llevar a cabo la invención, que son obvias para los expertos en biología molecular o en campos relacionados, pretenden incluirse en el alcance de las siguientes reivindicaciones.

50

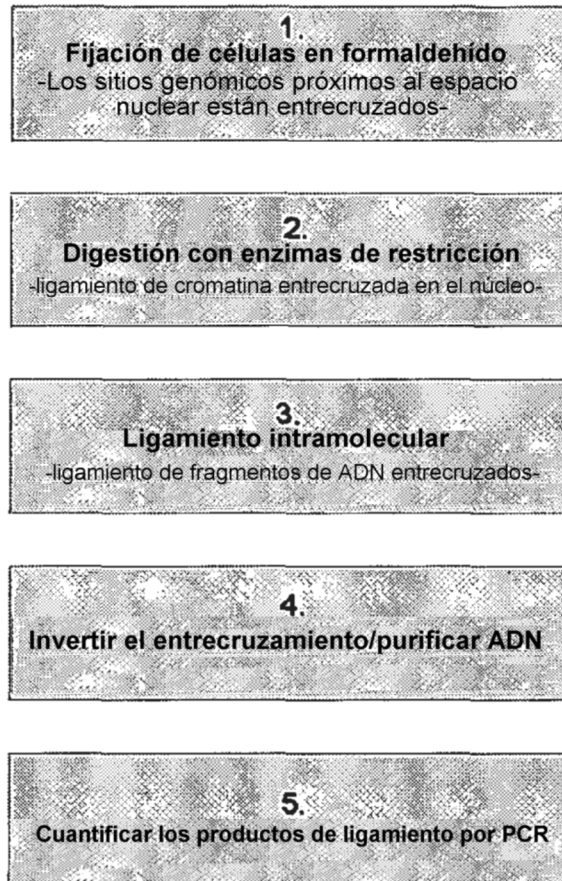


## REIVINDICACIONES

1. Un método para analizar la frecuencia de interacción de dos o más secuencias de nucleótidos cebo con una o más secuencias de nucleótidos de interés (por ejemplo uno o más locus genómicos) que comprende las etapas de:
- 5 (a) proporcionar una muestra de ADN entrecruzado;  
 (b) digerir el ADN entrecruzado con una enzima de restricción primaria;  
 (c) ligar las secuencias de nucleótidos entrecruzadas;  
 (d) invertir el entrecruzamiento;  
 10 (e) opcionalmente, digerir las secuencias de nucleótidos con una enzima de restricción secundaria;  
 (f) opcionalmente, ligar una o más secuencias de ADN de composición nucleotídica conocida, con el sitio o sitios de digestión de la enzima de restricción secundaria disponibles que flanquea(n) la una o más secuencias de nucleótidos de interés;  
 15 (g) amplificar la una o más secuencias de nucleótidos de interés utilizando al menos dos cebadores oligonucleotídicos, hibridándose cada cebador con las secuencias de ADN que flanquean las secuencias de nucleótidos de interés utilizando PCR múltiple;  
 (h) hibridar la secuencia o secuencias amplificadas con una matriz; y  
 (i) determinar la frecuencia de interacción entre las secuencias de ADN, en el que los productos de ligamiento entre las secuencias cebo y las secuencias entrecruzadas se analizan mediante secuenciación de alto  
 20 rendimiento.
2. El método de acuerdo con la reivindicación 1, en el que la reacción de ligamiento de la etapa (c) o (f) da como resultado la formación de círculos de ADN.
- 25 3. El método de acuerdo con cualquiera de las reivindicaciones anteriores para analizar la frecuencia de interacción de dos o más secuencias de nucleótidos cebo con una o más secuencias de nucleótidos de interés, que comprende el agrupamiento de algunos o de todos los productos de PCR obtenidos para cada una de las secuencias cebo en la etapa (g) y análisis simultáneo posterior de sus interacciones con el ADN.
- 30 4. El método de acuerdo con la reivindicación 3, en el que las dos o más secuencias amplificadas se marcan diferencialmente antes del agrupamiento y del análisis por hibridación con una matriz.
5. El método de acuerdo con la reivindicación 3 o la reivindicación 4, en el que las dos o más secuencias  
 35 amplificadas se marcan idénticamente y se analizan por hibridación con una matriz cuando las secuencias se encuentran en cromosomas diferentes.
6. El método de acuerdo con la reivindicación 3, en el que las dos o más secuencias amplificadas se marcan idénticamente cuando las secuencias se encuentran en el mismo cromosoma a una distancia suficiente para permitir un solapamiento mínimo entre las señales de interacción ADN-ADN.
- 40 7. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que se utiliza secuenciación a alto rendimiento para analizar las uniones de ligamiento formadas entre las secuencias cebo y las secuencias de interés capturadas.
- 45 8. El método de acuerdo con la reivindicación 7, en el que la secuenciación se dirige a las uniones de ligamiento formadas entre las secuencias cebo y las secuencias de interés capturadas, por la adición de secuencias adaptadoras necesarias para la secuenciación con los extremos de las secuencias amplificadas.
- 50 9. El método de acuerdo con la reivindicación 8, en el que la secuenciación se dirige a las uniones de ligamiento formadas entre las secuencias cebo y las secuencias de interés capturadas, por la adición de las secuencias adaptadoras completas, o de una parte de ellas, necesarias para la secuenciación, como salientes 5' en los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés.
- 55 10. El método de acuerdo con la reivindicación 8, en el que la secuenciación se dirige a las uniones de ligamiento formadas entre las secuencias cebo y las secuencias de interés capturadas, por la conjugación de una sustancia de biotina u otra fracción con los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés, seguido de purificación mediada por estreptavidina o de cualquier otra manera, del material amplificado por PCR.
- 60 11. El método de acuerdo con cualquiera de las reivindicaciones 7 a 10, en el que la secuenciación se dirige a las uniones de ligamiento entre las secuencias cebo y las secuencias de interés capturadas, diseñando los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés entre 400, 300, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1 nucleótidos desde el sitio o sitios de reconocimiento de la enzima de restricción primaria y/o secundaria analizados.
- 65

- 5 12. El método de acuerdo con cualquiera de las reivindicaciones 7 a 10, en el que la secuenciación se dirige a las uniones de ligamiento entre las secuencias cebo y las secuencias de interés capturadas, diseñando los cebadores oligonucleotídicos utilizados para amplificar la una o más secuencias de nucleótidos de interés de tal manera que solapan parcial o completamente con el sitio de reconocimiento de la enzima de restricción primaria y/o secundaria analizado.
- 10 13. El método de acuerdo con cualquiera de las reivindicaciones 7 a 12, en el que las secuencias se leen a través de la unión de ligamiento de tal manera que, cuando se analizan muestras de PCR múltiples o agrupadas, se obtiene una información de secuencia suficiente (por ejemplo, de 12 nucleótidos o más) en cualquier lado de la unión de ligamiento para identificar de modo inequívoco cada secuencia cebo y cada secuencia de interés capturada.
- 15 14. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que la secuencia de nucleótidos cebo se selecciona del grupo que consiste en un reordenamiento genómica, un promotor, un potenciador, un silenciador, un aislante, una región de unión a matriz, una región de control de locus, una unidad de transcripción, un origen de replicación, un punto caliente de recombinación, un punto de corte de translocación, un centrómero, un telómero, una región densa en genes, una región pobre en genes, un elemento repetitivo, un sitio de unión con un factor de transcripción y un sitio de integración (vírico).
- 20 15. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que la secuencia de nucleótidos cebo es una secuencia de nucleótidos que está asociada a una enfermedad o la causa, o que está localizada a una distancia de hasta 15 Mb, o mayor, en un molde de ADN lineal de un locus que está asociado a una enfermedad o la causa.
- 25 16. El método de acuerdo con cualquiera de las reivindicaciones anteriores, en el que la secuencia de nucleótidos cebo se selecciona del grupo que consiste en: *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, *IGH*, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCRα/δ*, *TCRβ* y *HOX* u otros locus asociados a la enfermedad.
- 30 17. Un método para identificar una o más interacciones ADN-ADN que son indicativas de una patología particular, que comprende la etapa de realizar las etapas (a)-(i) de las reivindicaciones 1-16, en el que en la etapa (a) se proporciona una muestra de ADN entrecruzado de una célula enferma y de una célula no enferma, y en el que una diferencia entre la frecuencia de interacción entre las secuencias de ADN de las células enfermas y no enfermas indica una diferencia en la organización lineal de los moldes de cromosoma (por ejemplo un reordenamiento genómica), que es indicativa de una característica o de una patología particular.
- 35 18. Un método de diagnóstico o de pronóstico de una enfermedad o de un síndrome causado por, o asociado a, un cambio en una interacción ADN-ADN que comprende la etapa de realizar las etapas (a)-(i) de cualquiera de las reivindicaciones 1-16, en la que la etapa (a) comprende proporcionar una muestra de ADN entrecruzado de un sujeto; y en el que la etapa (i) comprende comparar la frecuencia de interacción entre las secuencias de ADN con la de un control no afectado; en el que una diferencia entre el valor obtenido del control y el valor obtenido del sujeto es indicativa de que el sujeto padece la enfermedad o el síndrome o es indicativa de que el sujeto padecerá la enfermedad o el síndrome.
- 40 19. Un método de acuerdo con cualquier reivindicación anterior, para la detección de un reordenamiento, translocación, inversión, deleción, duplicación o inserción, equilibrada y/o no equilibrada.
- 45 20. Un método para analizar la frecuencia de interacción de una secuencia de nucleótidos cebo con una o más secuencias de nucleótidos (por ejemplo, uno o más locus genómicos) que comprende las etapas de:
- 50 (a) proporcionar una muestra de ADN entrecruzado;
- (b) digerir el ADN entrecruzado con una enzima de restricción primaria;
- (c) ligar las secuencias de nucleótidos entrecruzadas;
- (d) invertir el entrecruzamiento;
- (e) opcionalmente digerir las secuencias de nucleótidos con una enzima de restricción secundaria;
- (f) circularizar las secuencias de nucleótidos;
- 55 (g) amplificar la una o más secuencias de nucleótidos que están ligadas con la secuencia de nucleótidos cebo;
- (h) analizar las secuencias amplificadas mediante secuenciación a alto rendimiento; y
- (i) determinar la frecuencia de interacción entre las secuencias de ADN.
- 60 21. Un método de acuerdo con cualquiera de las reivindicaciones 1 a 19, en el que la etapa de hibridación con la matriz se reemplaza por una etapa de secuenciación.

# FIGURA 1



**FIGURA 2**

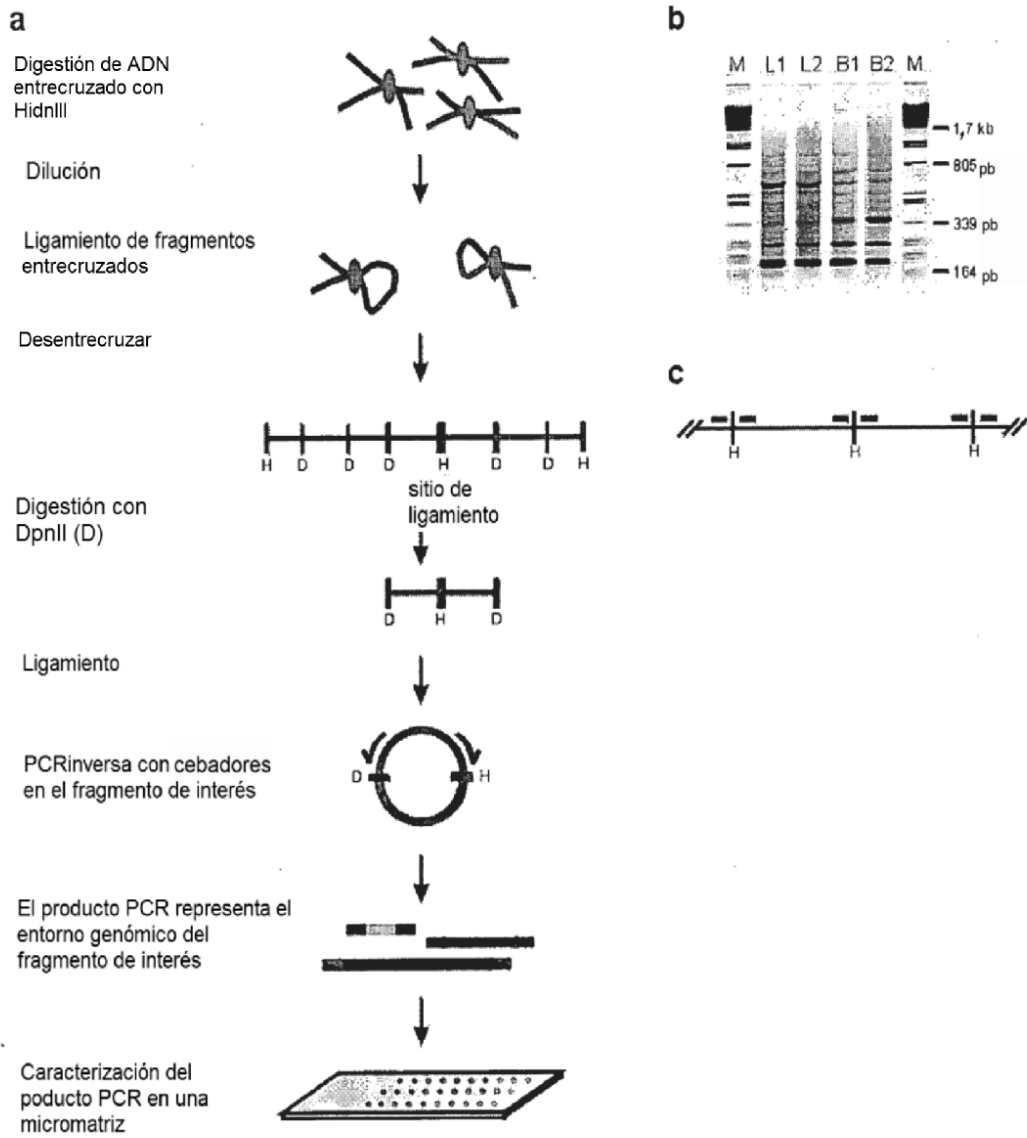
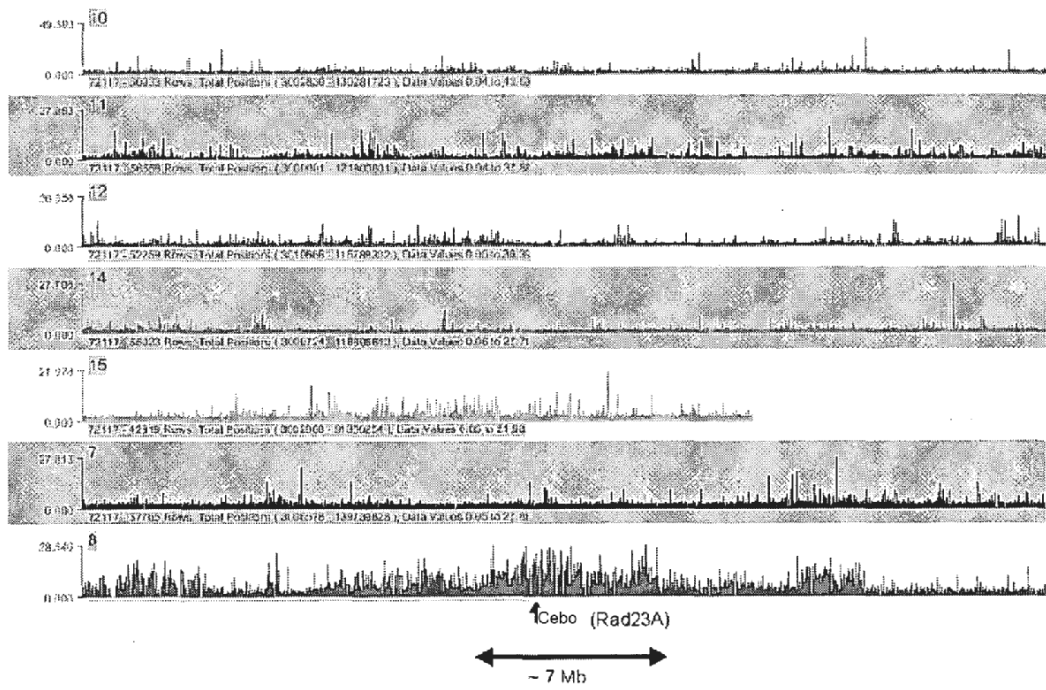


FIGURA 3



**FIGURA 4**

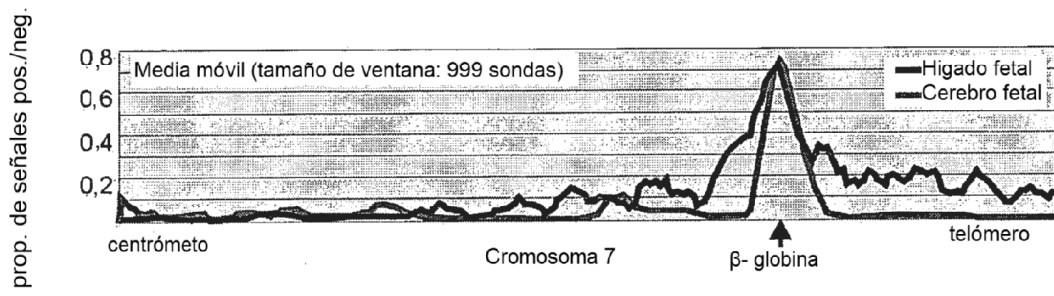


FIGURA 5

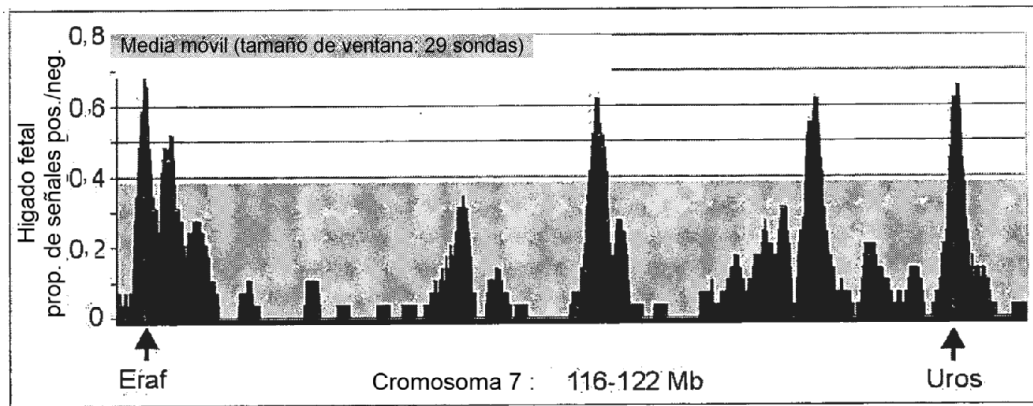
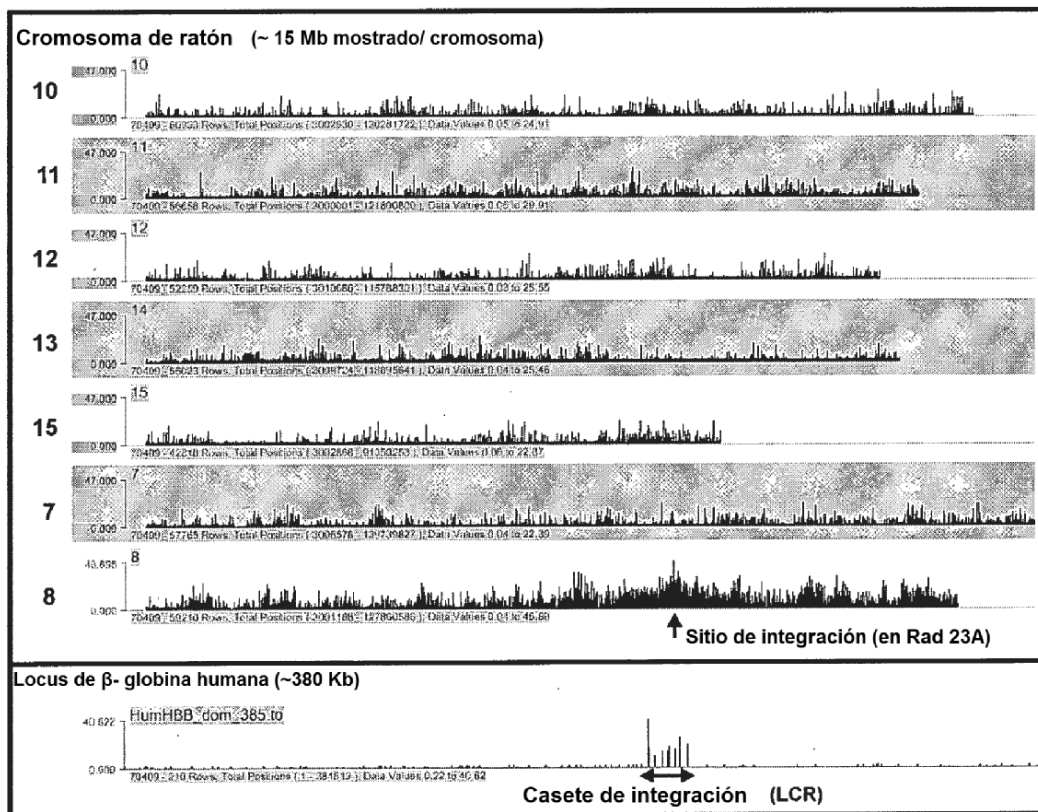


FIGURA 6





**FIGURA 7**

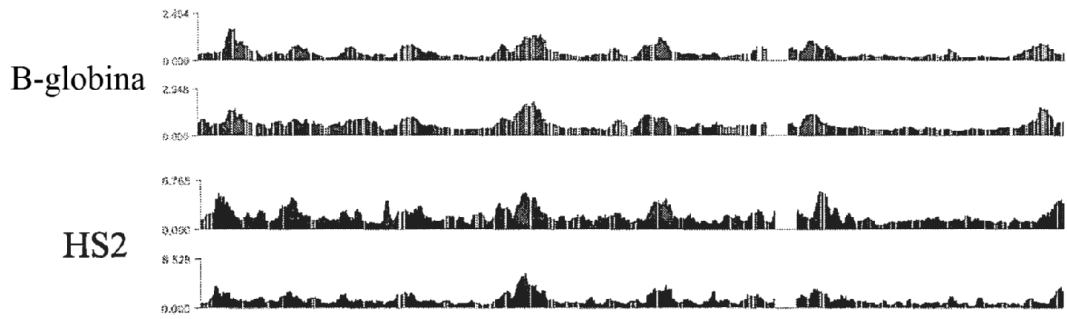
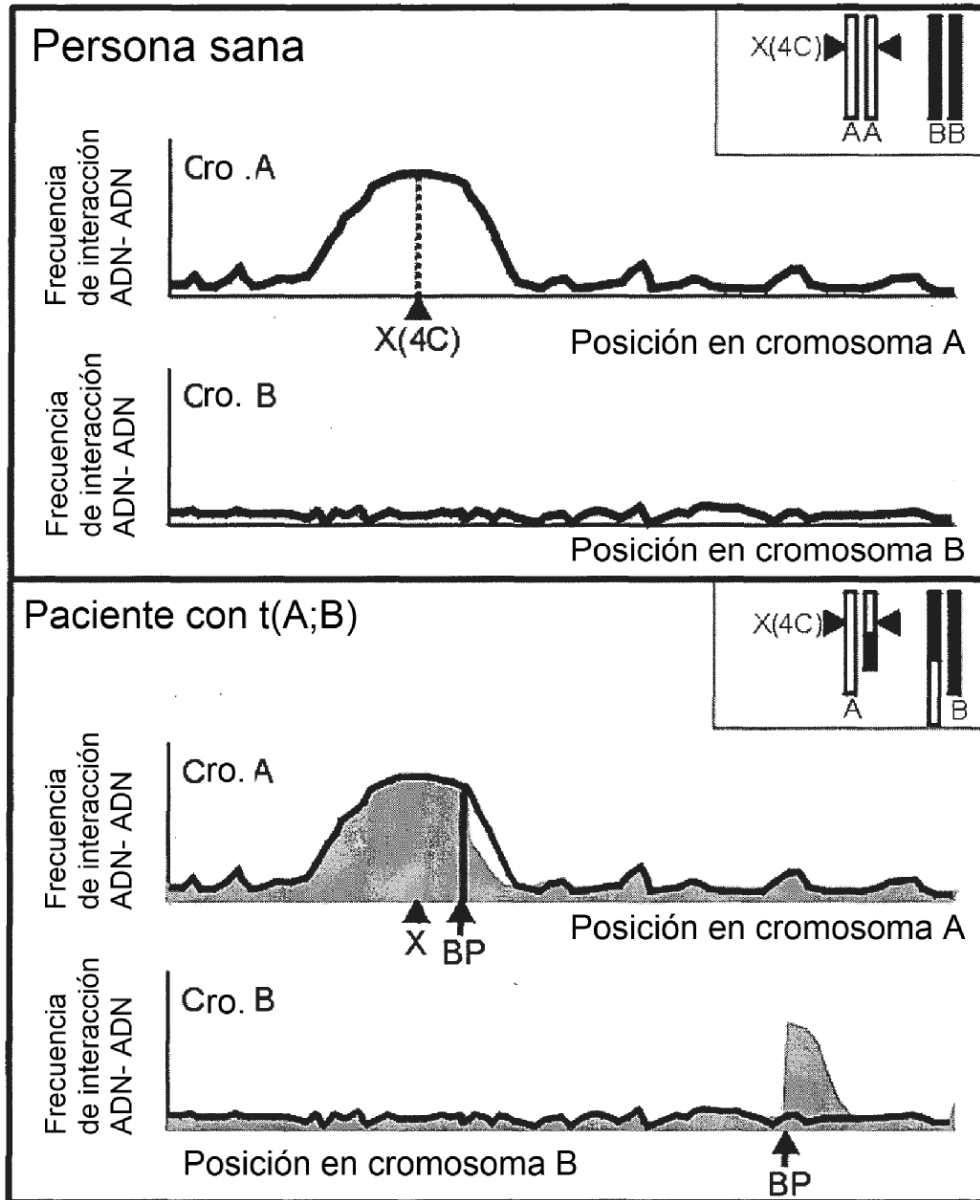


FIGURA 8



**FIGURA 9**

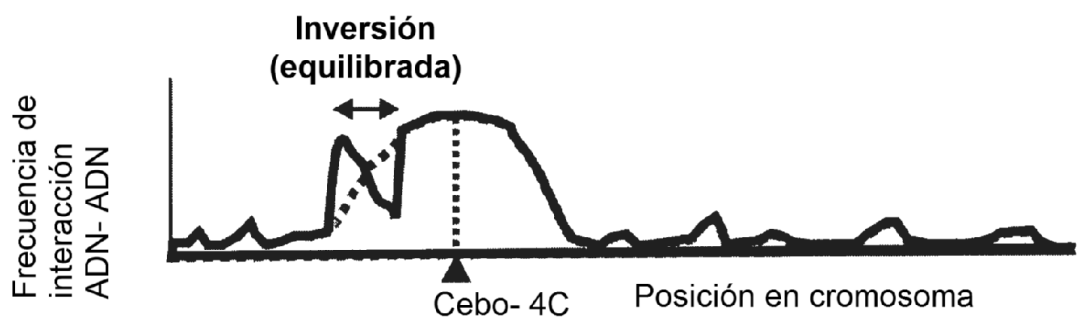


FIGURA 10

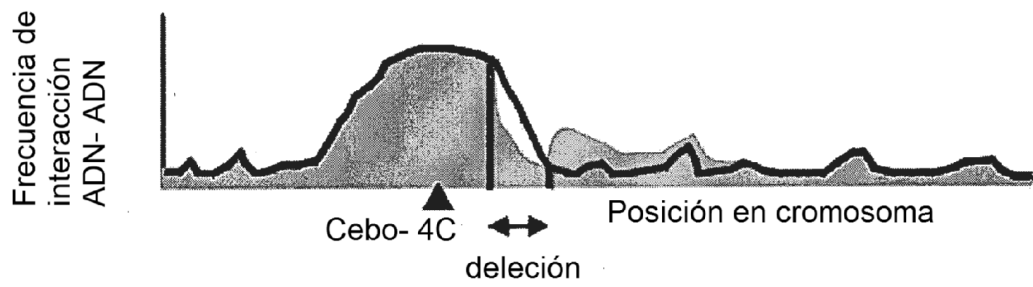


FIGURA 11

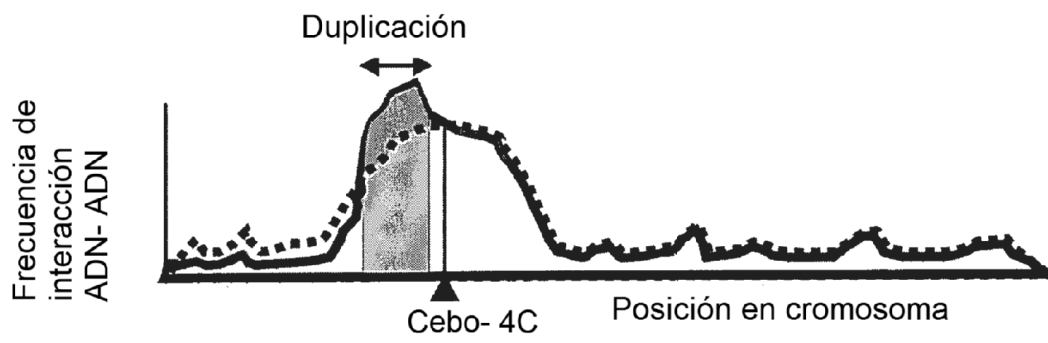
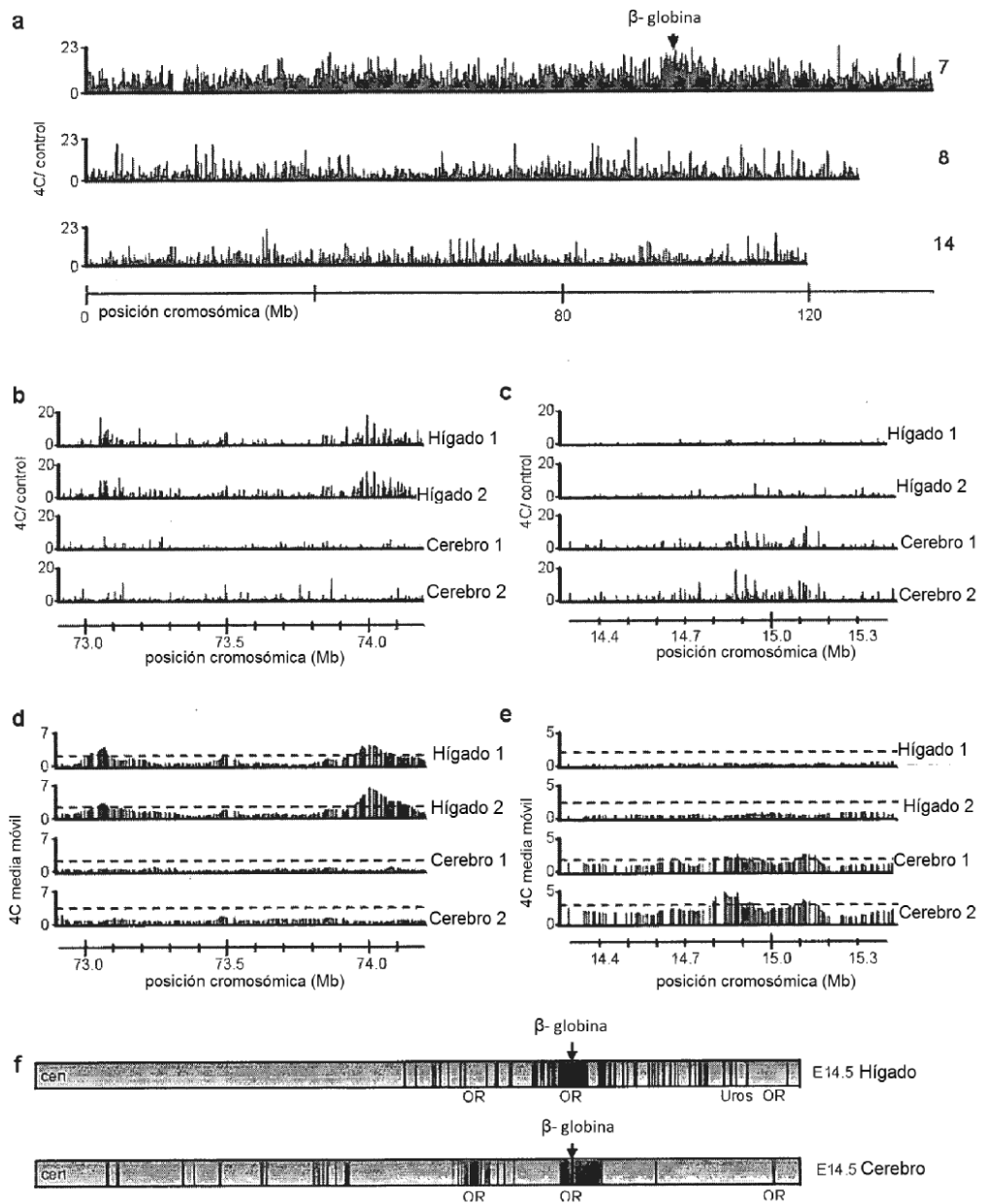
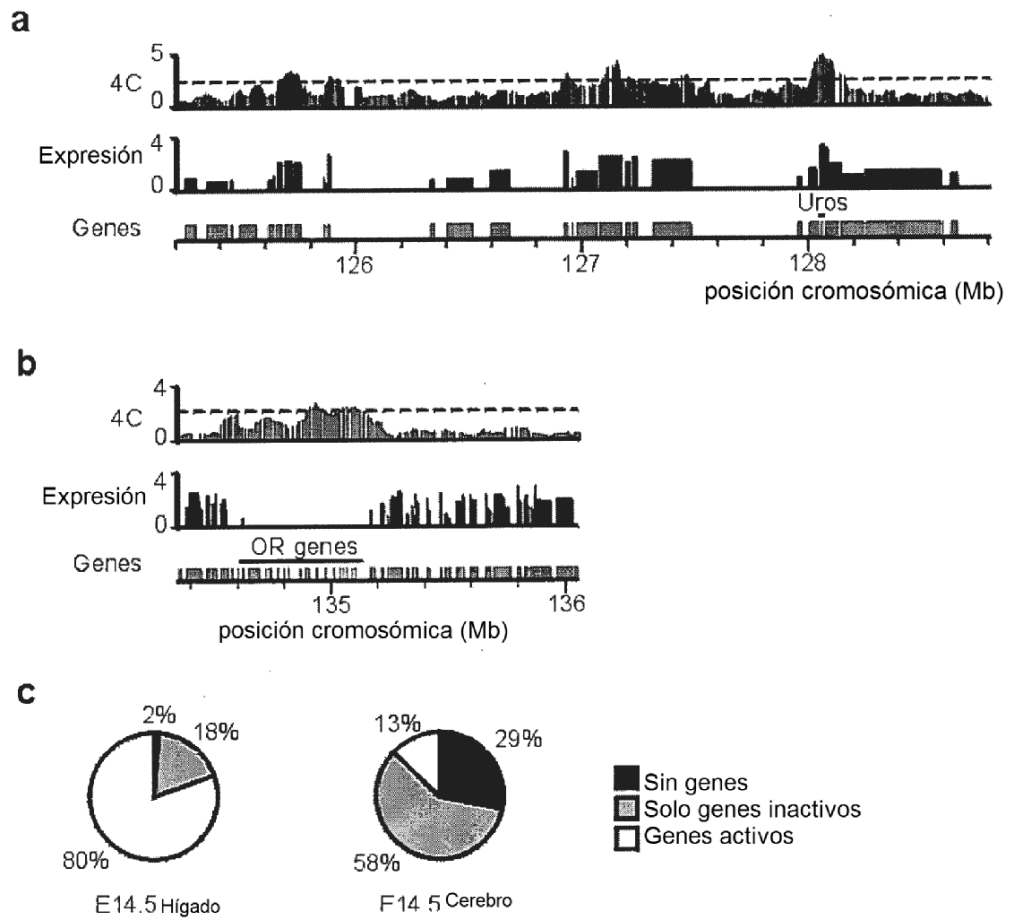


FIGURA 12



**FIGURA 13**



**FIGURA 14**

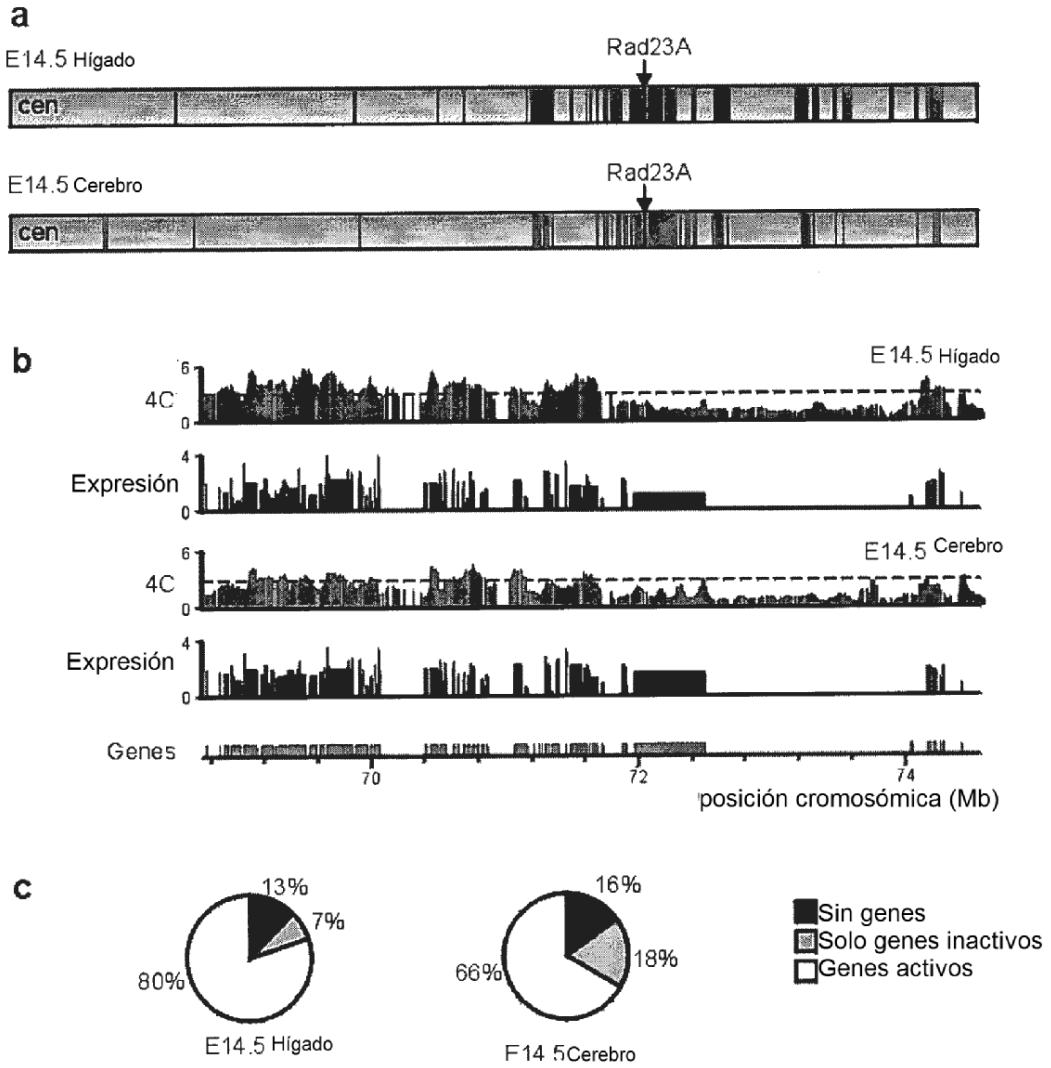




FIGURA 15

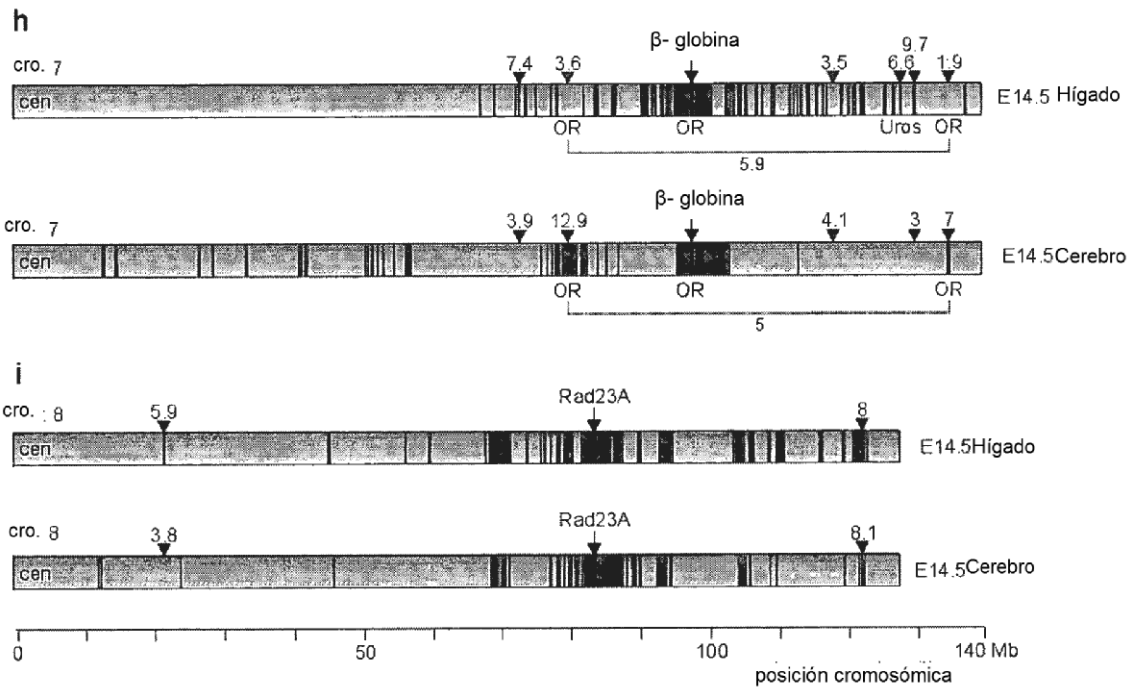
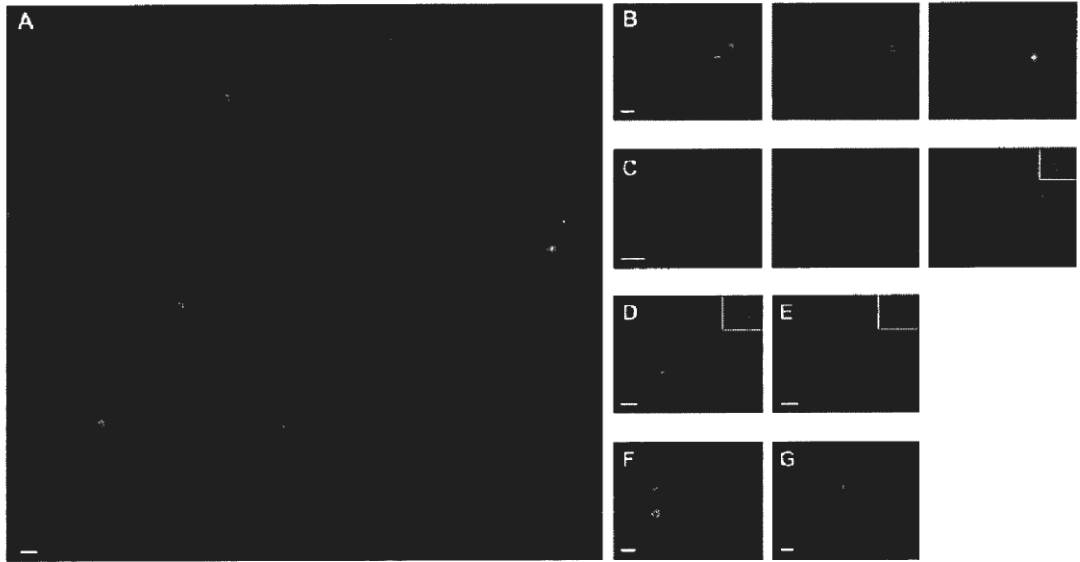
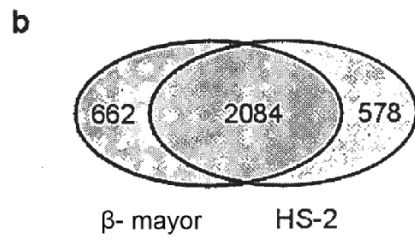
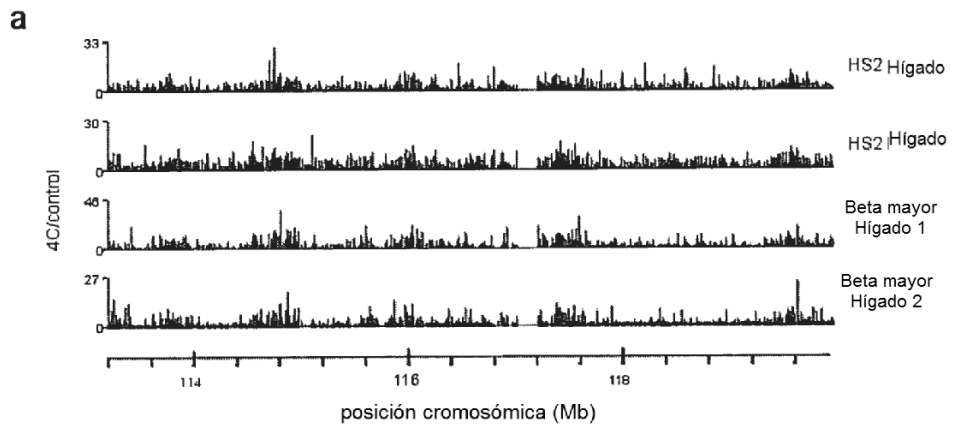


FIGURA 16



**FIGURA 17**

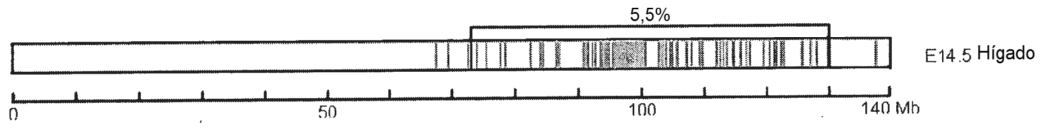


FIGURA 18

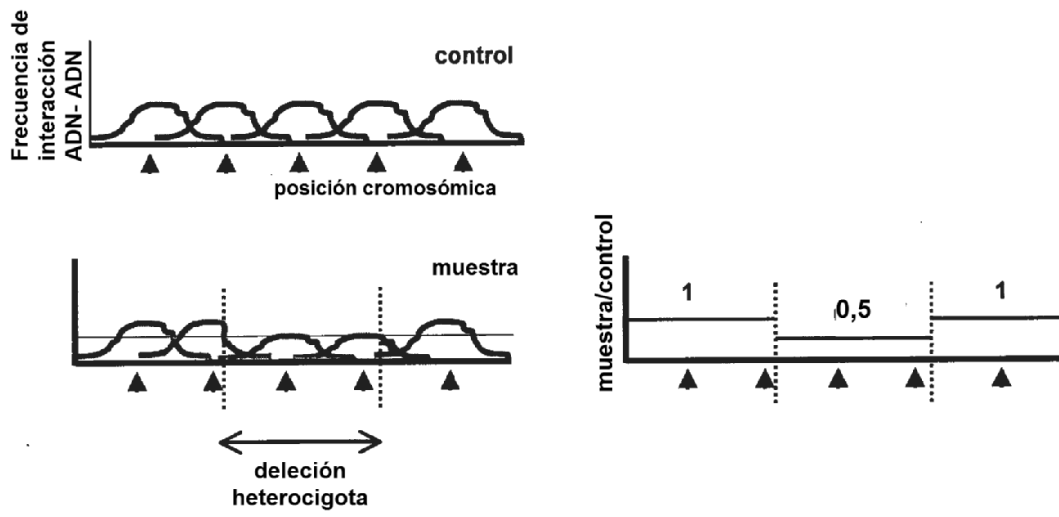
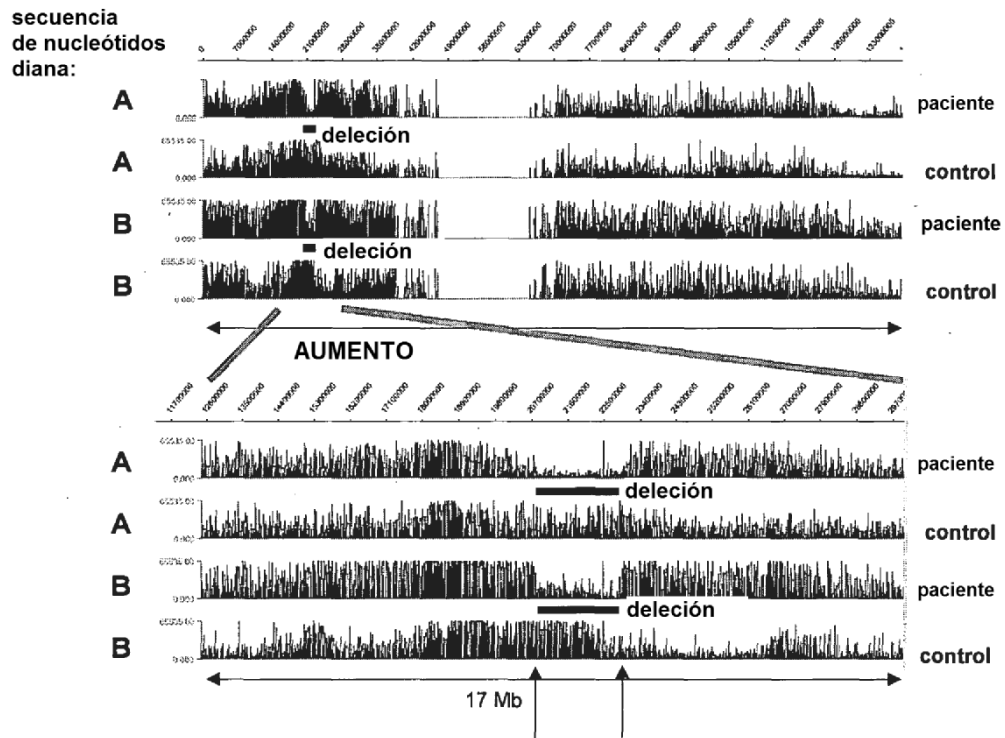


FIGURA 19



Secuencia sobre el punto de corte :

AAGGAAACCCCATGCCATAAGACGTCACTAATTTCTGAACCTTGTTTTTTTTTTTTTTTTTTT  
 CAAGTAGTTCTCATCTAAGTAGTTGTTTTTGTTCATGAGAAAATCAGATATGTTGCTAAAAATTCACAA  
 CTATTGCAAGAAAAATAAAAGAC

FIGURA 20

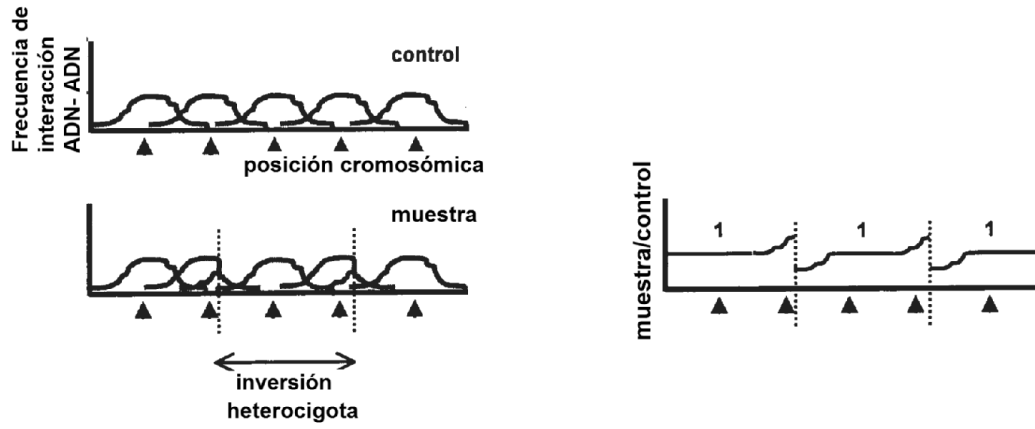


FIGURA 21

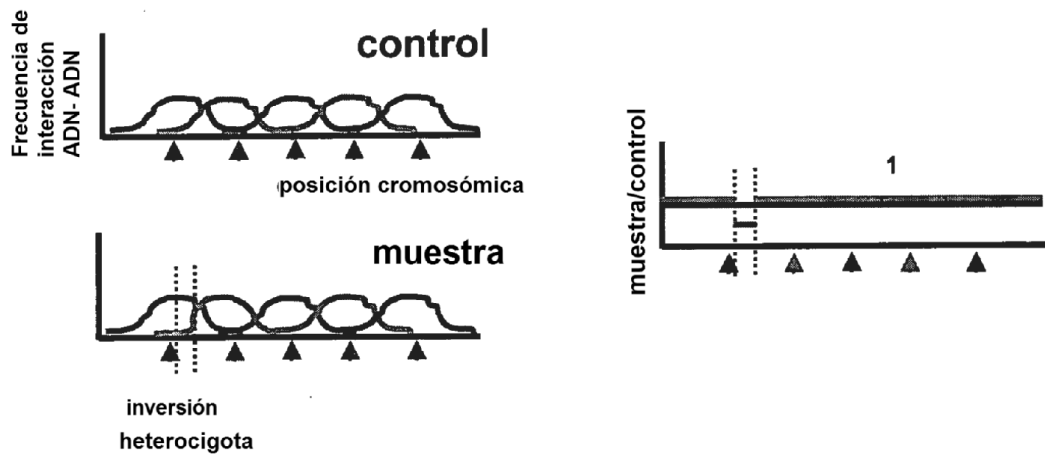
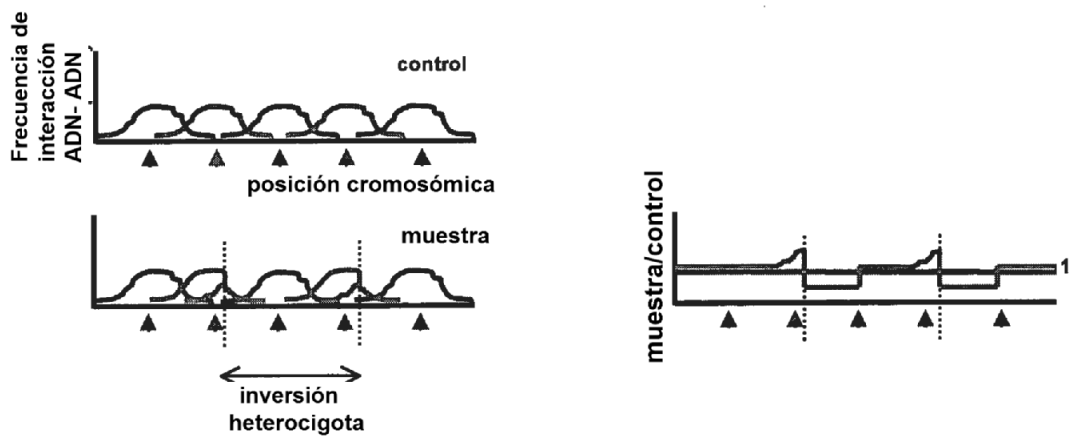


FIGURA 22





**FIGURA 23**

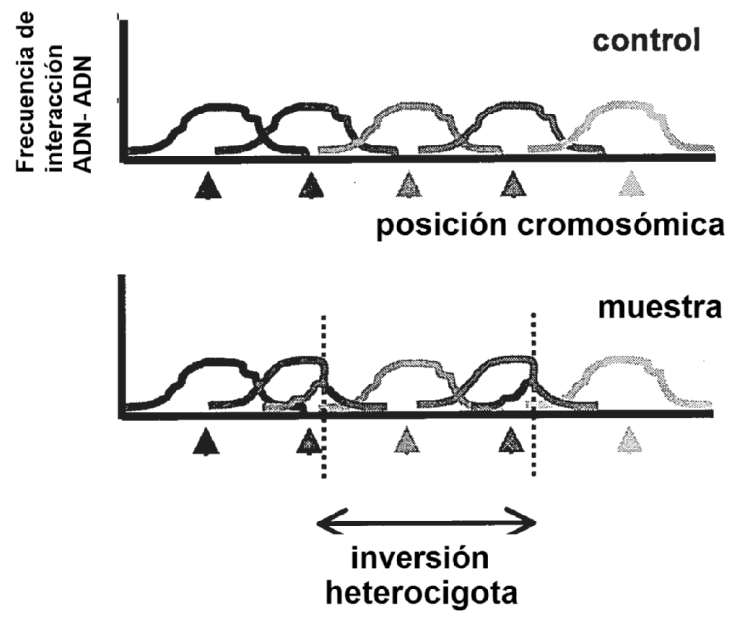


FIGURA 24

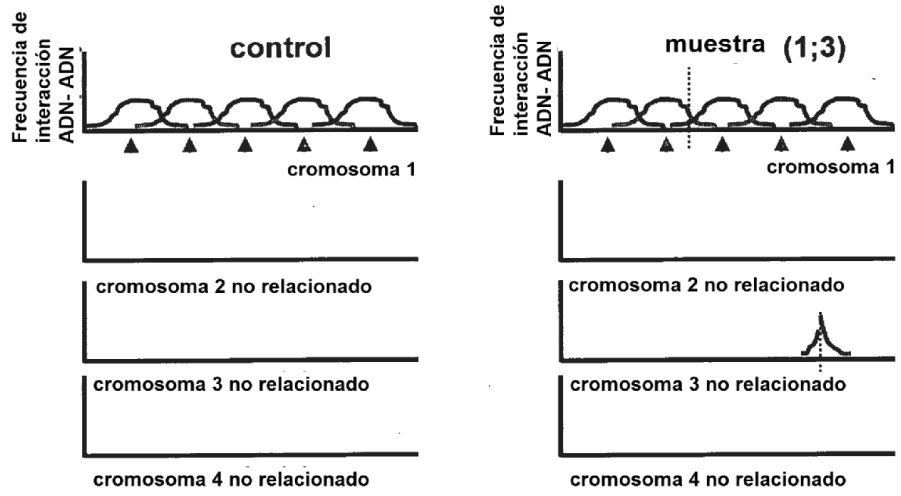


FIGURA 25

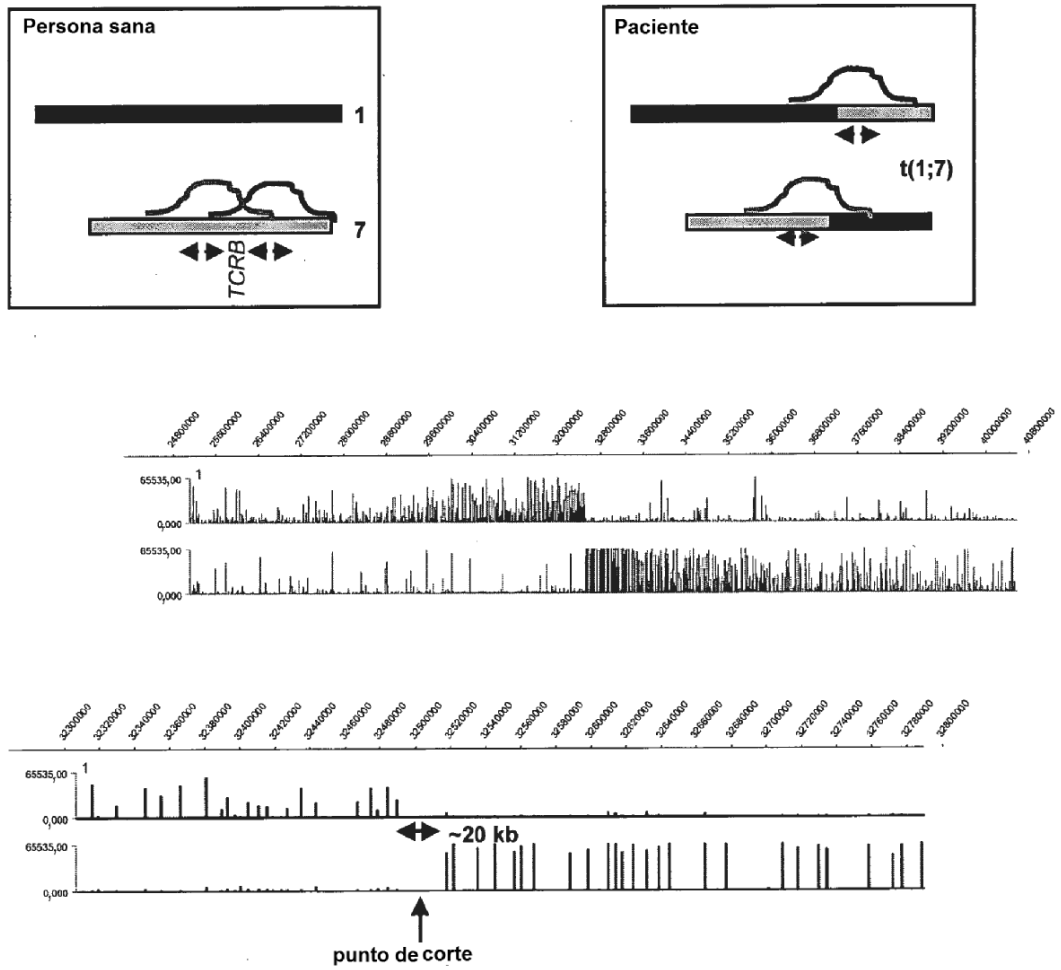


FIGURA 26

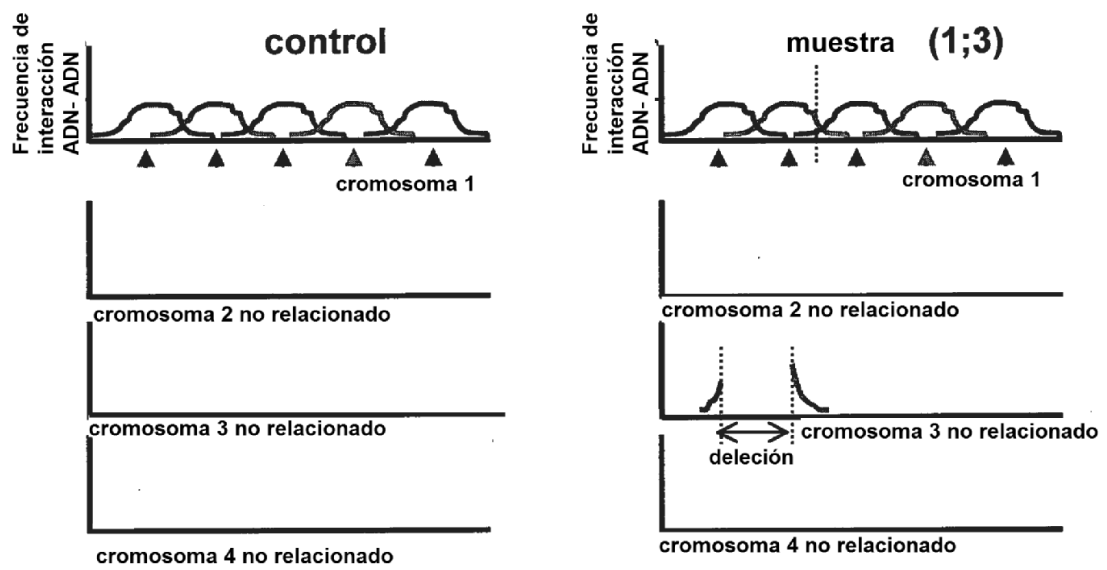
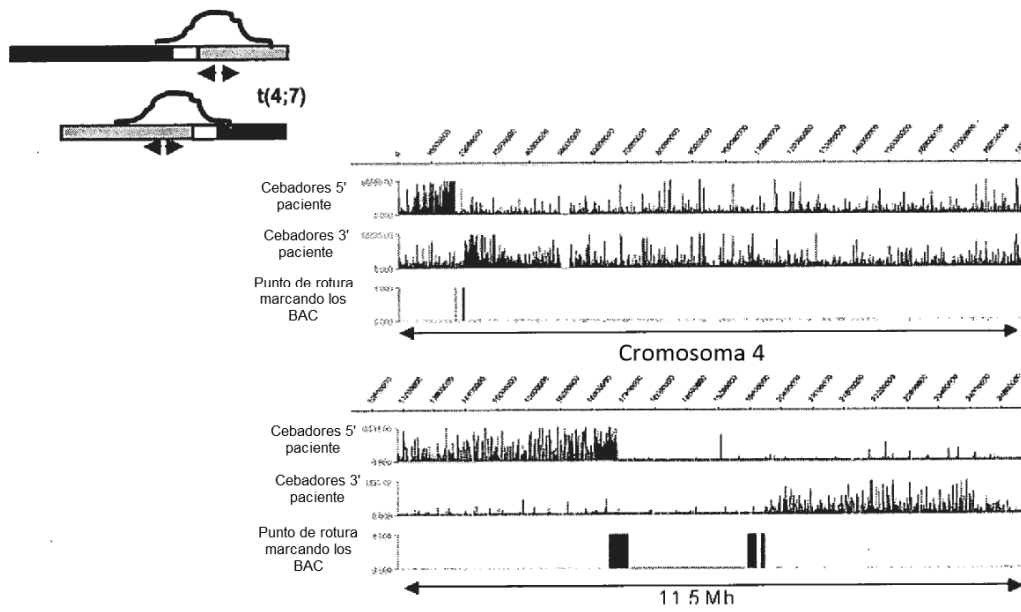


FIGURA 27

Paciente



Secuencia del punto de corte nº 2 :

CTCCAATGTAAGTGTGGATTACACCTAAAAGAGCCAGAAAACACAGACTC  
 TCTGTGGAACCATGACACAACAGTGCTTGGTATTATTTTTTCCTAGTTAG

FIGURA 28

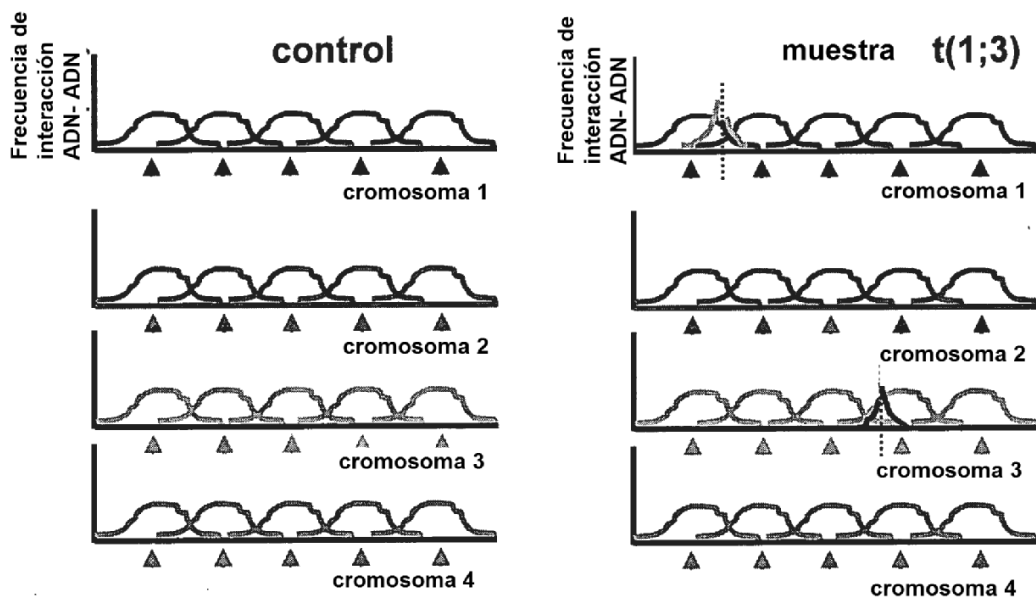
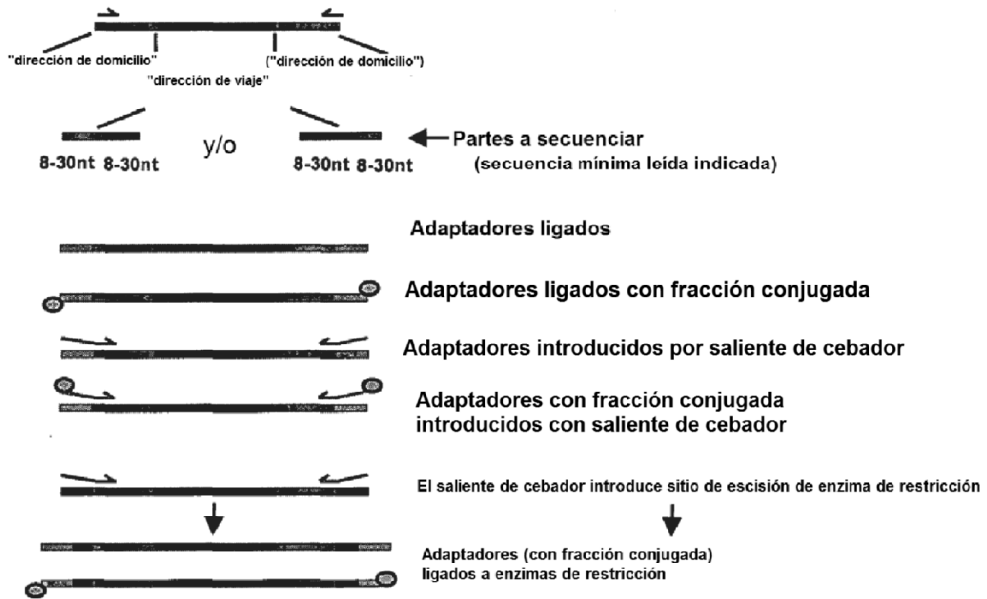
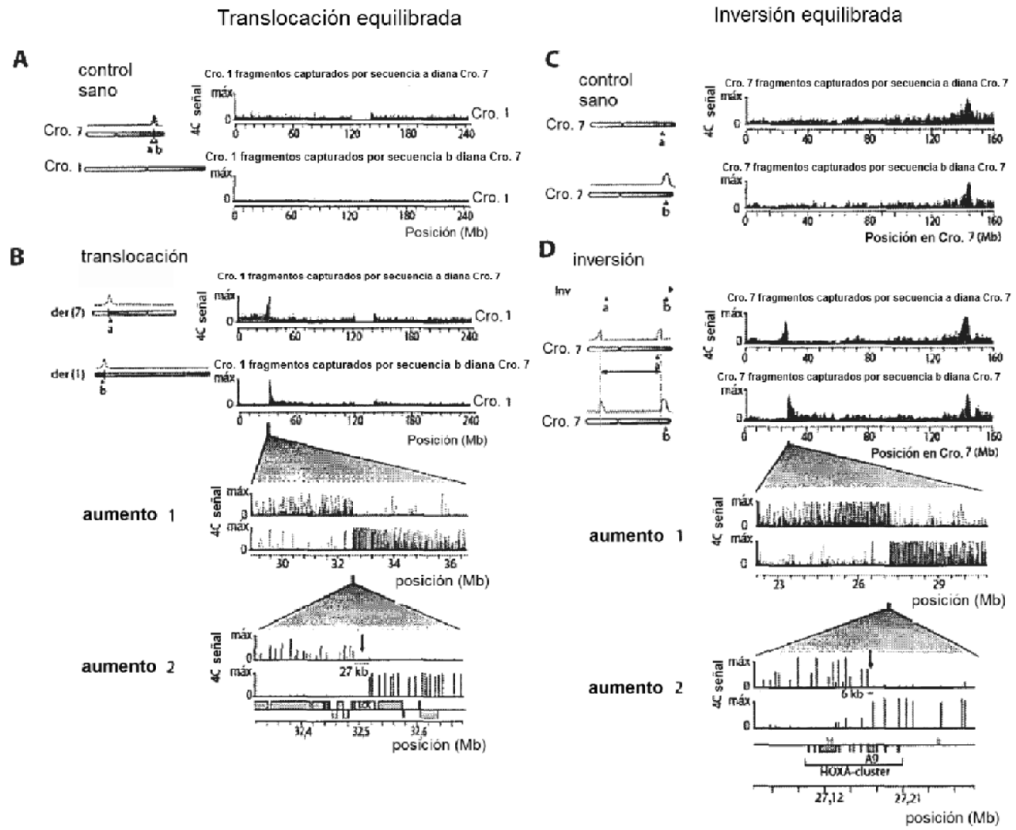


FIGURA 29



Nota: Los adaptadores de cada extremo del producto PCR pueden ser iguales o diferentes

FIGURA 30

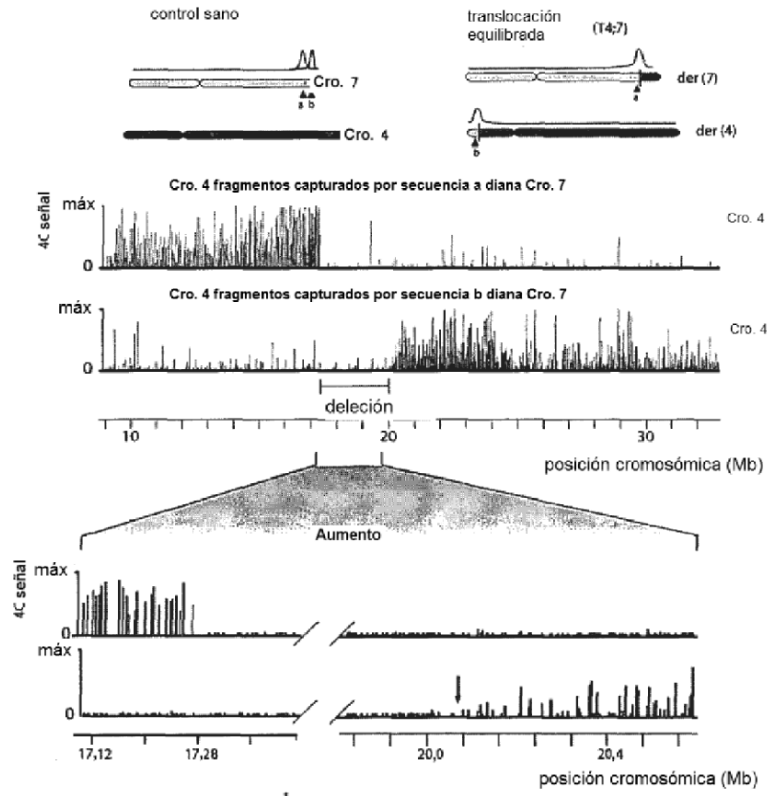




**FIGURA 31**

**A**

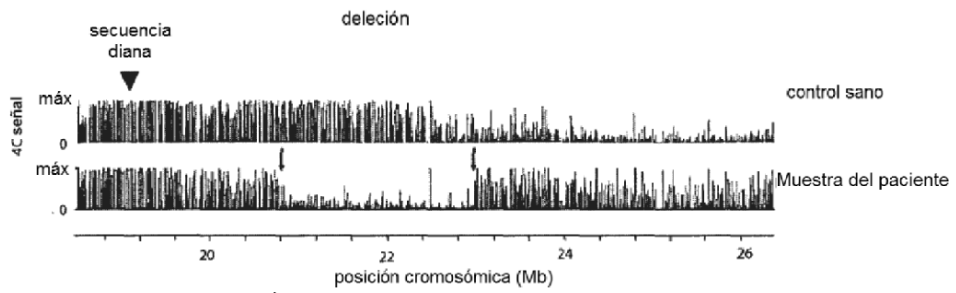
**translocación no equilibrada**



**B**

aagagccagaaaaacagactctctgtGGAACCATGACACAACAGTGCTTGGTATTATT der(4)

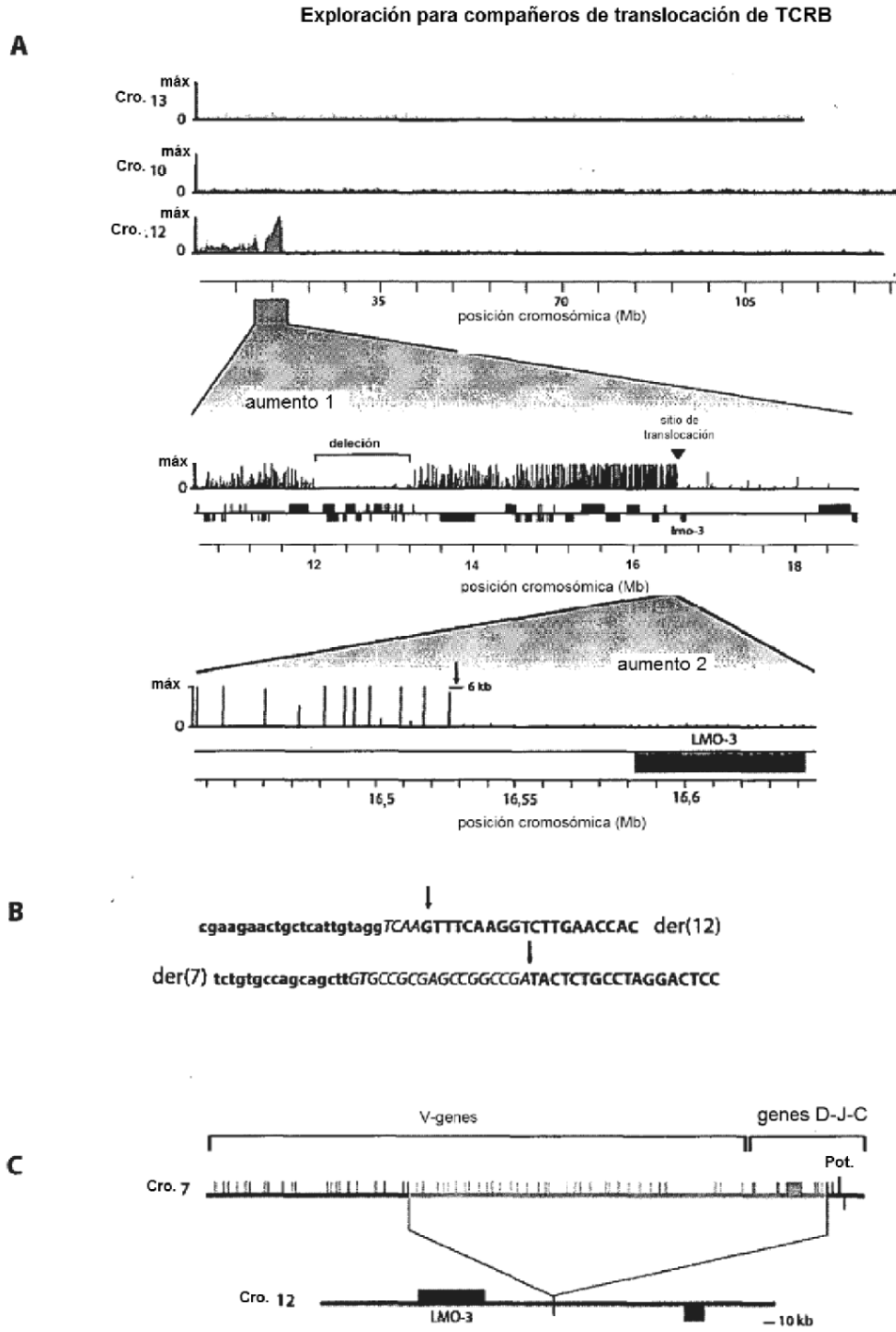
**C**



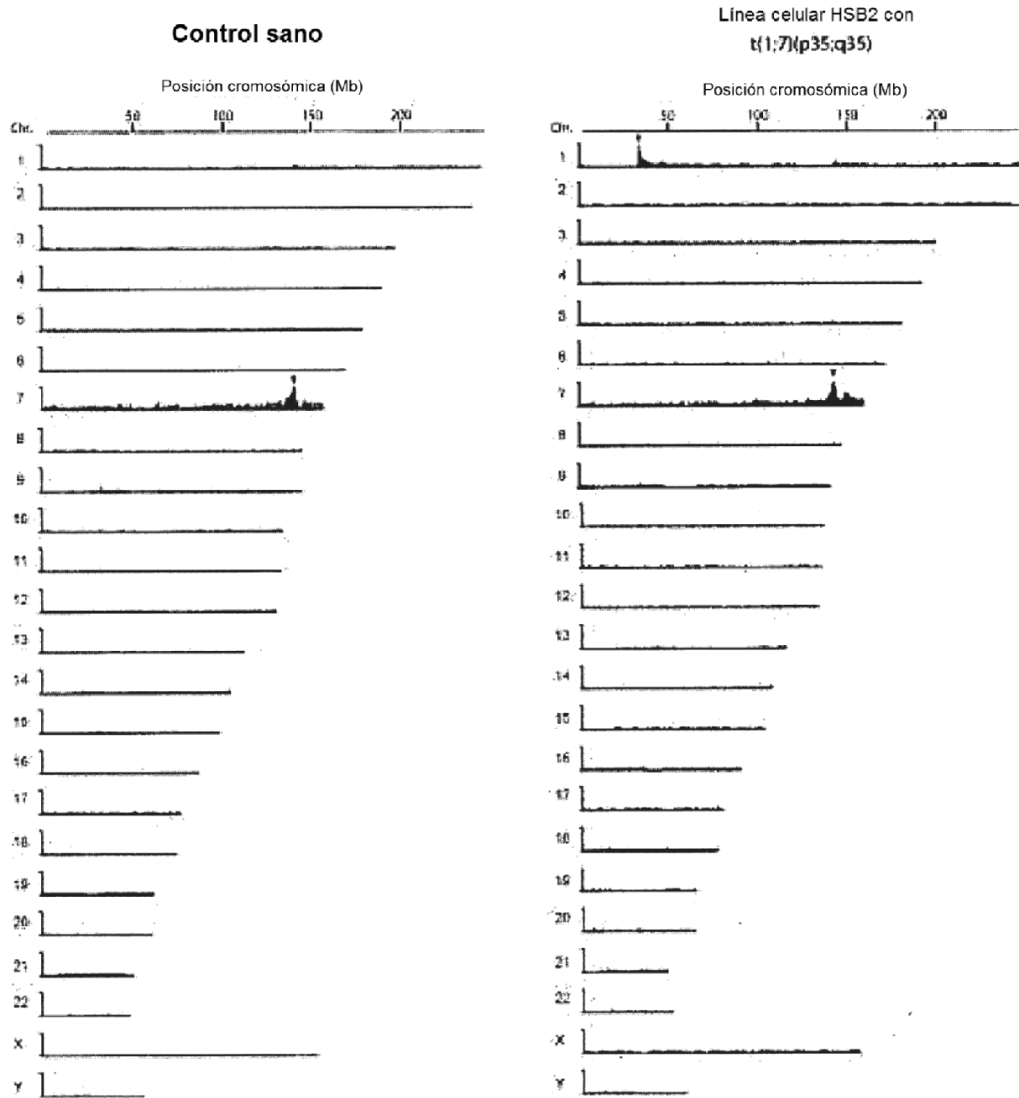
**D**

aaacccatgccataagagctactAATTTCTGAACCTGTGTTTTTTT

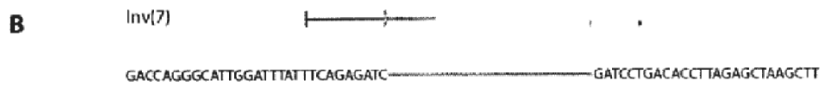
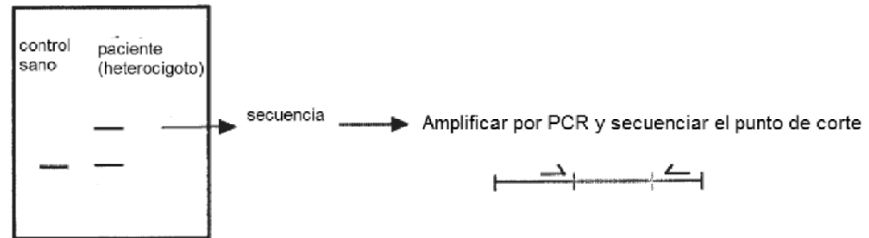
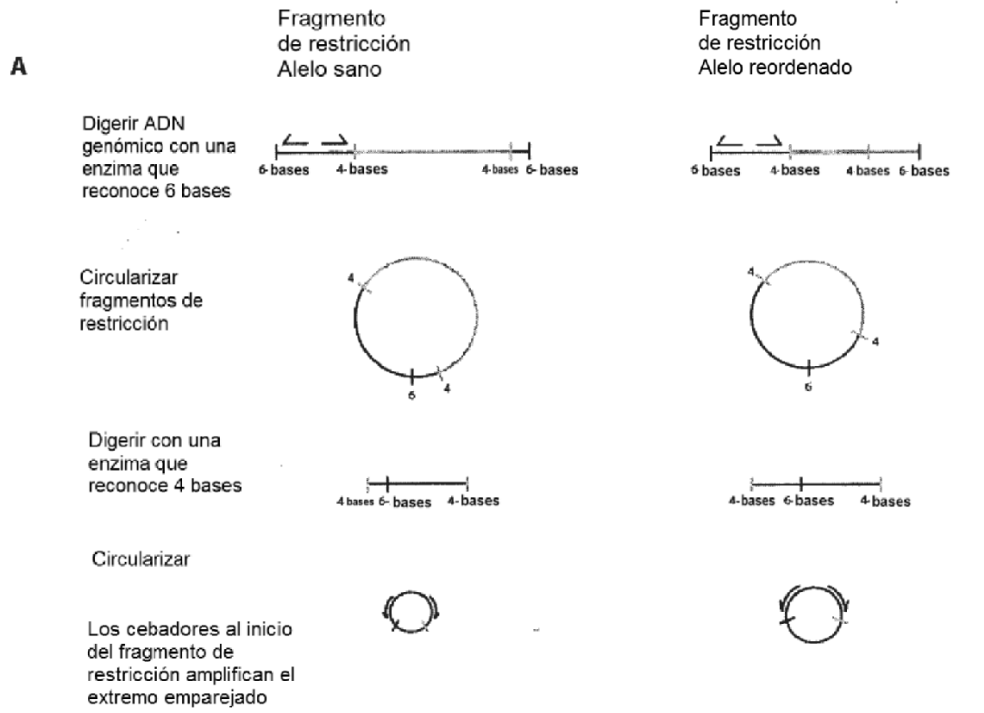
FIGURA 32



**FIGURA 33**



**FIGURA 34**



**FIGURA 35**

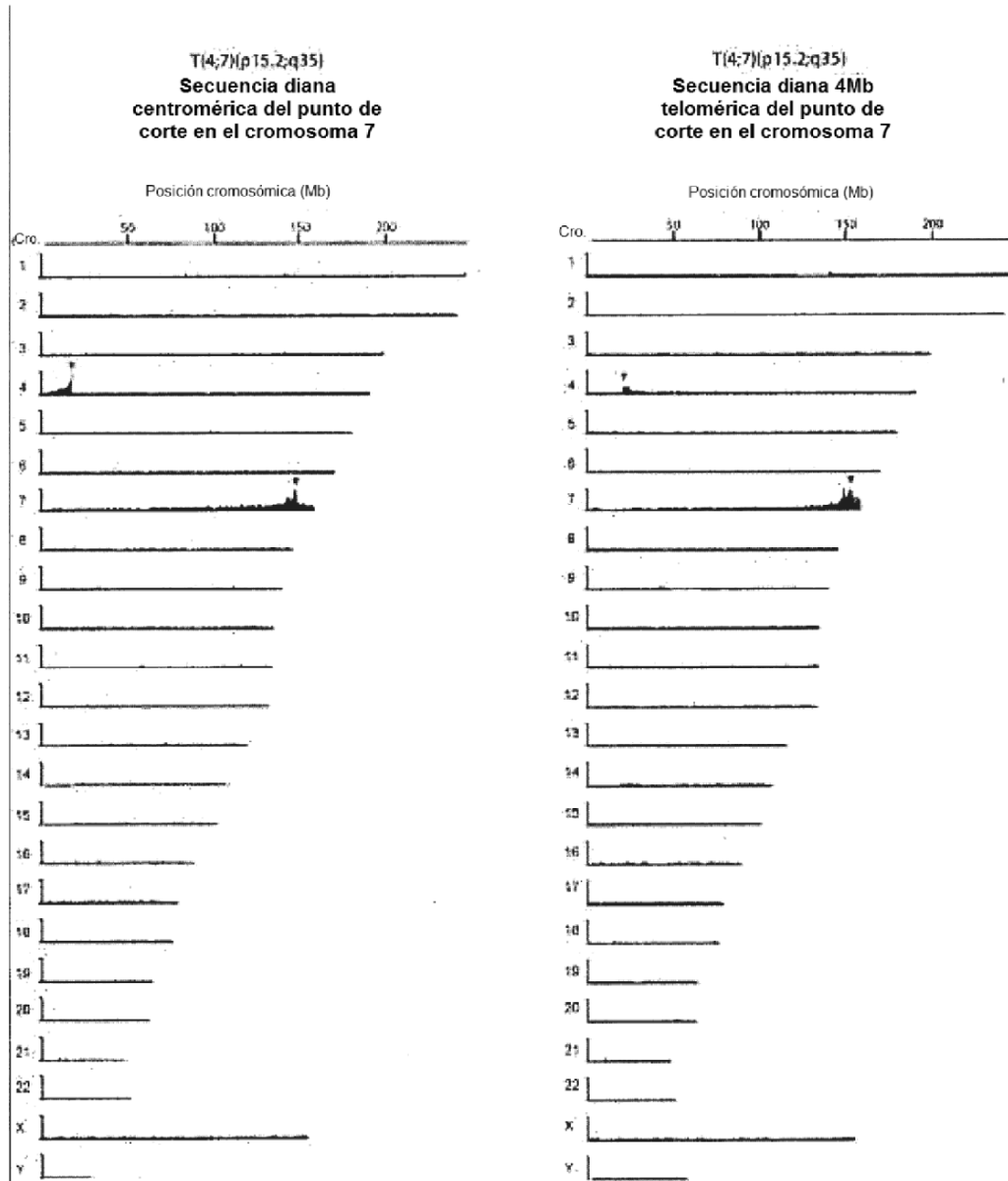
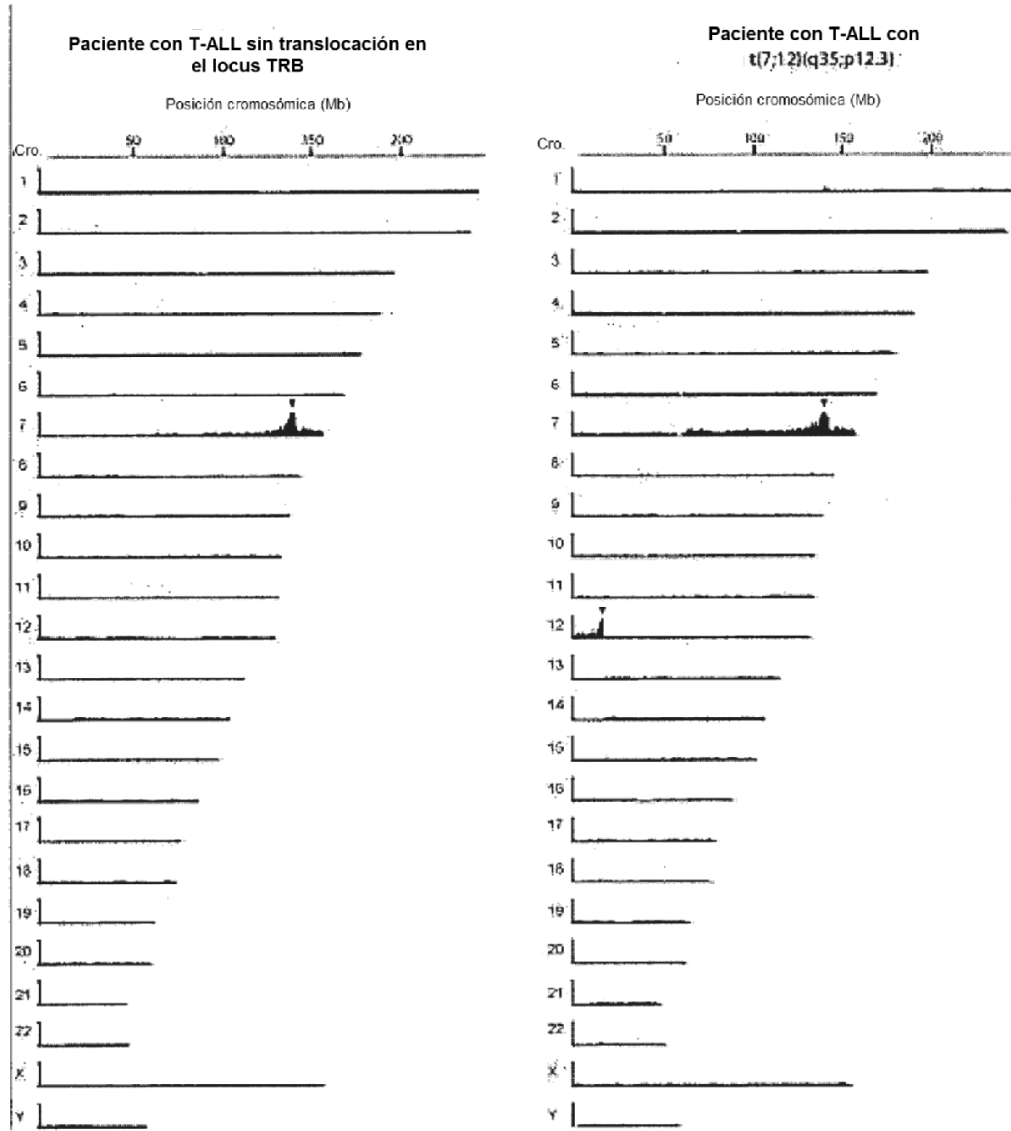


FIGURA 36



**FIGURA 37**

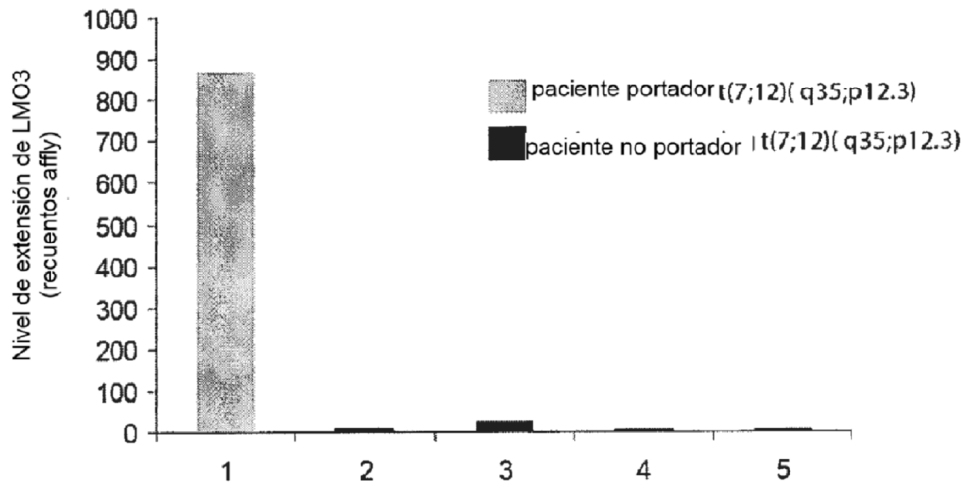
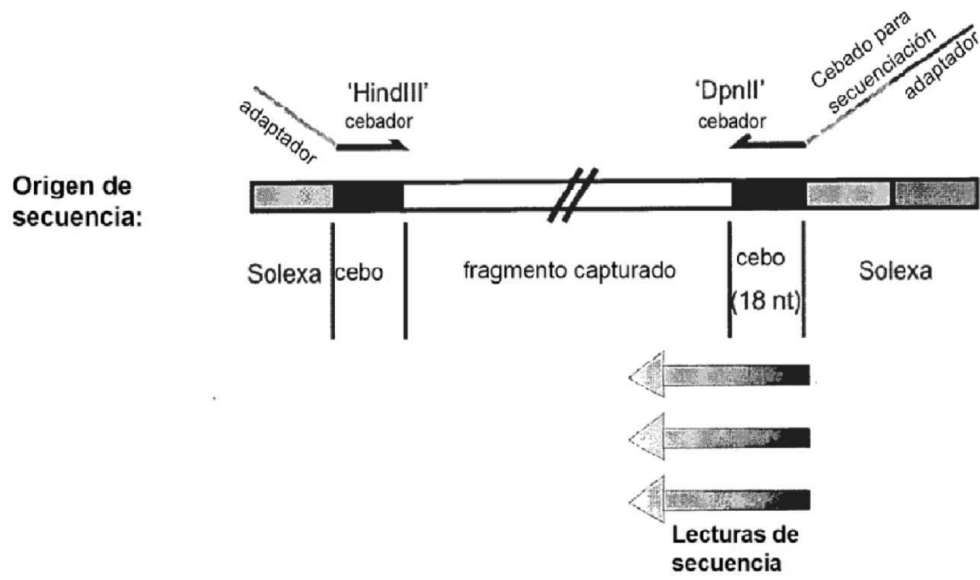


FIGURA 38



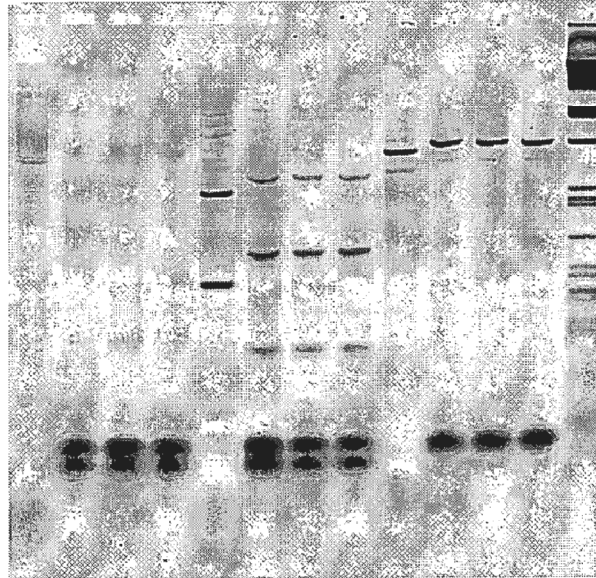


**FIGURA 39**

**Cebadores:**

	85 Mb				105 Mb				139 Mb				
posición:	-	+	+	+	-	+	+	+	-	+	+	+	
adaptador:	-	+	+	+	-	+	+	+	-	+	+	+	

Marcador



**FIGURA 40**

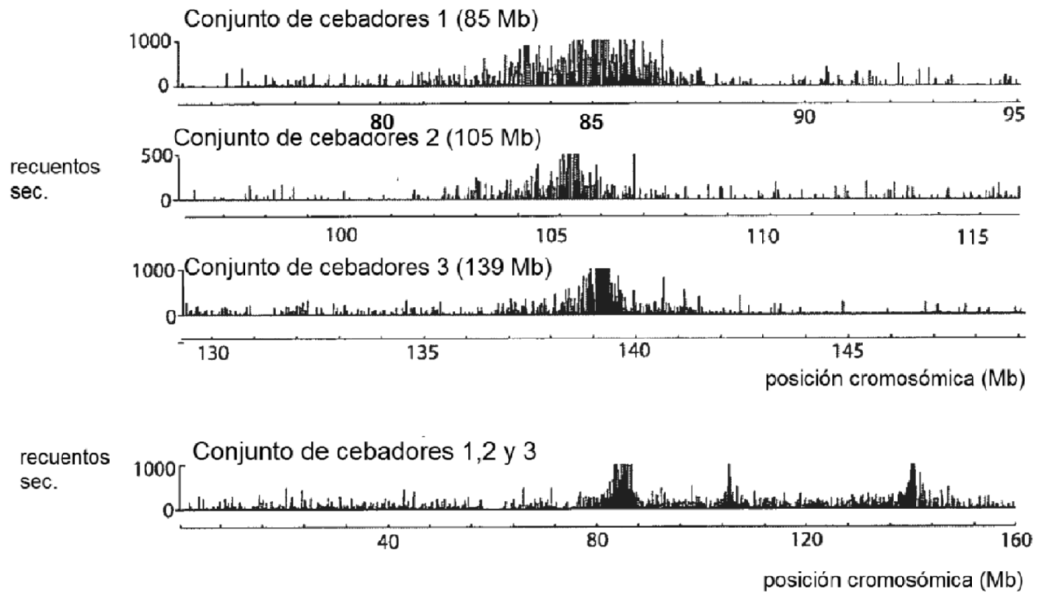


FIGURA 41

