(19) OFICINA ESPAÑOLA DE PATENTES Y MARCAS

ESPAÑA

(11) Número de publicación: **2 556 113**

(51) Int. CI.:

C12Q 1/68 (2006.01)

(12) TRADUCCIÓN DE PATENTE EUROPEA T3

(96) Fecha de presentación y número de la solicitud europea: **03.07.2006 E 06795293 (7)**

(97) Fecha y número de publicación de la concesión europea: **16.09.2015 EP 1899488**

(54) Título: **Ensayo de captura en chip (4C) de conformación cromosómica**

(30) Prioridad:

04.07.2005 GB 0513676
17.03.2006 GB 0605449

(45) Fecha de publicación y mención en BOPI de la traducción de la patente:
13.01.2016

(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:

GROSVELD, FRANK y
DE LAAT, WOUTER

(74) Agente/Representante:

ISERN JARA, Jorge

Observaciones :

Véase nota informativa (Remarks) en el folleto original publicado por la Oficina Europea de Patentes

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

Ensayo de captura en chip (4C) de conformación cromosómica

Campo de la invención

La presente invención se refiere al análisis de la frecuencia de interacción de dos o más secuencias nucleotídicas en el espacio nuclear.

Antecedentes de la invención

Los estudios sobre la arquitectura nuclear de mamífero aspiran a comprender el modo en que 2 indicadores de ADN se pliegan en un núcleo de 10 μm transversalmente, permitiendo al mismo tiempo la expresión precisa de los genes que especifican el tipo celular, y el modo en que esto se propaga fielmente durante cada ciclo celular. El progreso en este campo ha provenido en gran medida de los estudios de microscopia, que revelaron que los genomas se disponen de forma no aleatoria en el espacio nuclear. Por ejemplo, la heterocromatina empaquetada de forma densa está separada de eucromatina más abierta y los cromosomas ocupan distintos territorios en el espacio nuclear. Existe una intrincada relación entre el posicionamiento nuclear y la actividad transcripcional. Aunque la transcripción sucede en todo el interior nuclear, los genes activos que se agrupan en cromosomas se localizan preferentemente en el borde o exterior de su territorio cromosómico. Los genes individuales pueden migrar tras cambios en su estado de transcripción, medido frente a referencias nucleares relativamente grandes tales como territorios cromosómicos, centrómeros o la periferia nuclear. Además, los genes transcritos de forma activa decenas de megabases separándose el cromosoma pueden juntarse en el núcleo, como se ha demostrado recientemente por hibridación in situ de fluorescencia (FISH) para el locus de β-globina y unos pocos genes seleccionados diferentes. Además de la transcripción, la organización genómica está asociada con la coordinación de la replicación, recombinación y la probabilidad de que los loci se transloquen (lo que puede conducir a neoplasias) y el ajuste y reajuste de programas epigenéticos. En base a estas observaciones se cree que la organización arquitectónica del ADN en el núcleo celular es un factor clave para la función genómica.

Se han desarrollado diferentes ensayos para permitir una comprensión en la organización espacial de los loci genómicos in vivo. Se ha desarrollado un ensayo, llamado ARN-TRAP (Carter et al. (2002) Nat. Genet. 32, 623) que implica dirigir la peroxidasa de rábano rusticano (HRP) a transcritos de ARN naciente, seguido por cuantificación de la deposición de biotina catalizada por HRP en la cromatina adyacente.

Otro ensayo que se ha desarrollado se llama tecnología de captura de conformación cromosómica (3C), que proporciona una herramienta para estudiar la organización estructural de una región genómica. La tecnología 3C implica análisis por PCR cuantitativa de frecuencias de entrecruzamiento entre dos fragmentos dados de restricción de ADN, que da una medida de su proximidad en el espacio nuclear (véase la Figura 1). Originalmente desarrollado para analizar la conformación de cromosomas en 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 génicos intrincados de mamífero (véase, por ejemplo, Tolhuis et al., 2002; Palstra et al., 2003; y Drissen et al., 2004). En resumen, la tecnología 3C implica entrecruzamiento *in vivo* con formaldehído de células y digestión nuclear de cromatina con una enzima de restricción, seguido por ligamiento de fragmentos de ADN que se entrecruzaron en un complejo. Los productos de ligamiento después se cuantifican por PCR. La etapa de amplificación por PCR requiere el conocimiento de la información de secuencia para cada uno de los fragmentos de ADN que tienen que amplificarse. Por tanto, la tecnología 3C proporciona una medida de frecuencia de interacción entre fragmentos seleccionados de ADN.

Existe una necesidad importante de tecnología de alto rendimiento que pueda explorar de forma sistemática el genoma completo de un modo objetivo para el loci de ADN que contactan entre sí en el espacio nuclear.

La presente invención busca proporcionar mejoras en la tecnología 3C.

Sumario de la invención

La tecnología 3C aplicada actualmente solamente permite el análisis de una cantidad limitada de interacciones ADN-ADN seleccionadas a causa de las limitaciones en la etapa de amplificación por PCR, que requiere el conocimiento de información específica de secuencia para cada fragmento a analizar. Además, la selección de fragmentos de restricción como candidatos para interacciones de ADN de intervalo largo requiere una cantidad sustancial de conocimiento previo (por ejemplo, la localización de sitios hipersensibles) del locus de interés, que habitualmente no está disponible. Dada la relevancia funcional de muchas interacciones ADN-ADN de intervalo largo descritas hasta ahora, la capacidad de explorar aleatoriamente elementos de ADN que formen bucles con una secuencia de interés - tales como un promotor génico, potenciador, aislante, silenciador, origen de replicación o MAR/SAR - o una región genómica de interés - tal como una región densa en genes o pobre en genes o elemento repetitivo - puede facilitar enormemente el mapeado de secuencias implicadas en una red de regulación.

La presente invención se refiere a tecnología 4C (es decir, captura y caracterización de cromatina co-localizada), que proporciona el análisis de alto rendimiento de la frecuencia de interacción de dos o más secuencias nucleotídicas en el espacio nuclear.

La tecnología 4C (captura y caracterización de cromatina co-localizada) es una versión modificada de la tecnología 3C que permite una búsqueda objetiva del genoma completo de fragmentos de ADN que interaccionan con un locus de elección. En resumen, el análisis 3C se realiza de forma habitual, pero omitiendo la etapa de PCR. El molde 3C contiene un señuelo (por ejemplo, un fragmento de restricción de elección que abarca un gen de interés) ligado a muchas secuencias nucleotídicas diferentes de interés (que representan el entorno genómico de este gen). El molde se escinde por otra enzima de restricción secundaria, y se liga. De forma ventajosa, la una o más secuencias nucleotídicas de interés que se ligan a la secuencia nucleotídica diana se amplifican usando al menos uno (preferiblemente, al menos dos) cebador oligonucleotídico, donde el al menos un cebador hibrida con una secuencia de ADN que flanquea la secuencia de nucleótidos de interés. Típicamente, esto produce un patrón de fragmentos de PCR que es altamente reproducible entre reacciones independientes de amplificación y específico para un tejido dado. En una realización, *Hind*III y *Dpn*II, se usan como enzima de restricción primaria y secundaria. A continuación, los fragmentos amplificados pueden marcarse y opcionalmente hibridarse en una serie, típicamente frente a una muestra de control que contiene ADN genómico digerido con la misma combinación de enzimas de restricción.

En una realización preferida de la presente invención, los fragmentos ligados que se escinden por una enzima de restricción secundaria posteriormente se vuelven a ligar para formar círculos pequeños de ADN.

La tecnología 3C por tanto se ha modificado de modo que todas las secuencias nucleotídicas de interés que interaccionan con una secuencia nucleotídica diana se amplifiquen. De forma práctica, 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 usando uno o más cebadores oligonucleotídicos que hibridan con una secuencia de ADN que flanquea las secuencias nucleotídicas de interés. De forma ventajosa, 4C no está influenciada hacia el diseño de cebadores de PCR que se incluyen en la etapa de amplificación de PCR y por lo tanto puede usarse para buscar en el genoma completo elementos de ADN de interacción.

Aspectos resumidos de la presente invención

La invención reivindicada puede resumirse en los siguientes párrafos numerados:

1. Un método para analizar la frecuencia de interacción de una secuencia nucleotídica diana con una o más secuencias nucleotídicas de interés (por ejemplo, uno o más loci 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 nucleotídicas entrecruzadas;

(d) revertir el entrecruzamiento;

(e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria;

(f) ligar una o más secuencias de ADN de composición conocida de nucleótidos con el sitio o sitios de digestión con enzima de restricción secundaria que flanquean la una o más secuencias nucleotídicas de interés;

(g) amplificar la una o más secuencias nucleotídicas de interés usando al menos dos cebadores oligonucleotídicos, donde cada cebador hibrida con las secuencias de ADN que flanquean las secuencias nucleotídicas de interés;

(h) hibridar la secuencia o secuencias amplificadas en una serie; y

(i) determinar la frecuencia de interacción entre las secuencias de ADN.

2. El método de acuerdo con el párrafo 1, donde la reacción de ligamiento en la etapa (f) provoca la formación de círculos de ADN.

3. El método de acuerdo con el párrafo 1 o párrafo 2, donde la secuencia nucleotídica diana se selecciona entre el grupo que consiste en un reordenamiento genómico, promotor, un potenciador, un silenciador, un aislante, una región de adhesión a matriz, una región de control de locus, una unidad de transcripción, un origen de replicación, un punto caliente de recombinación, un punto de interrupción de translocación, un centrómero, un telómero, una región densa en genes, una región pobre en genes, un elemento repetitivo y un sitio de integración (viral).

4. El método de acuerdo con cualquiera de los párrafos precedentes, donde la secuencia nucleotídica diana es una secuencia nucleotídica que está asociada con o causa una enfermedad, o está localizada hasta o más de 15 Mb en un molde de ADN lineal desde un locus que está asociado con o causa una enfermedad.

5. El método de acuerdo con cualquiera de los párrafos precedentes, donde la secuencia nucleotídica diana se selecciona entre el grupo que consiste en: AML1, MLL, MYC, BCL, BCR, ABL1, IGH, LYL1, TAL1, TAL2, LMO2, TCRα/δ, TCRβ y HOX u otros loci asociados con enfermedad.

6. El método de acuerdo con cualquiera de los párrafos precedentes, donde la enzima de restricción primaria es una enzima de restricción que reconoce un sitio de reconocimiento de 6-8 pb.

7. El método de acuerdo con el párrafo 6, donde la enzima de restricción primaria se selecciona entre el grupo que consiste en: BglII, HindIII, EcoRI, BamHI, SpeI, PstI y NdeI.

8. El método de acuerdo con cualquiera de los párrafos precedentes, donde la enzima de restricción secundaria es una enzima de restricción que reconoce un sitio de reconocimiento de secuencia nucleotídica de 4 o 5 pb.

9. El método de acuerdo con cualquiera de los párrafos precedentes, donde el sitio de reconocimiento por enzima de restricción secundaria está localizado a más de aproximadamente 350 pb desde el sitio de restricción primario en la secuencia nucleotídica diana.

10. El método de acuerdo con cualquiera de los párrafos precedentes, donde la secuencia nucleotídica está marcada.

11. Un método para analizar la frecuencia de interacción de una secuencia nucleotídica diana con una o más secuencias nucleotídicas (por ejemplo, uno o más loci 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 nucleotídicas entrecruzadas;

(d) revertir el entrecruzamiento;

(e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria;

(f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios de digestión con enzima de restricción secundaria disponibles que flanquean la una o más secuencias nucleotídicas de interés que circulariza las secuencias nucleotídicas;

(g) amplificar la una o más secuencias nucleotídicas que están ligadas a la secuencia nucleotídica diana usando al menos dos cebadores oligonucleotídicos, donde cada cebador hibrida con las secuencias de ADN que flanquean las secuencias nucleotídicas de interés;

(h) hibridar opcionalmente las secuencias amplificadas con una serie; y

(i) determinar la frecuencia de interacción entre las secuencias de ADN.

12. Un método para preparar una secuencia nucleotídica circularizada 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 nucleotídicas entrecruzadas;

(d) revertir el entrecruzamiento;

(e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria; y

(f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios de digestión con enzima de restricción secundaria disponibles que flanquean la una o más secuencias nucleotídicas de interés que circulariza las secuencias nucleotídicas.

13. Un método para identificar una o más interacciones ADN-ADN que son indicativas de un estado patológico particular que comprende la etapa de realizar las etapas (a)-(i) de los párrafos 1-11, donde en la etapa (a) se proporciona una muestra de ADN reticulado de células enfermas y no enfermas, y donde una diferencia entre la frecuencia de interacción entre las secuencias de ADN de las células enfermas y no enfermas indica que la interacción ADN-ADN es indicativa de un estado patológico particular.

14. Un método de diagnóstico de una enfermedad o síndrome causado por o asociado con un cambio en las interacciones ADN-ADN que comprende la etapa de realizar las etapas (a)-(i) cualesquiera de los párrafos 1-11, donde la etapa (a) comprende proporcionar una muestra de ADN entrecruzado de un sujeto; y donde la etapa (i) comprende comparar la frecuencia de interacción entre las secuencias de ADN con la de un control no afectado; donde una diferencia entre el valor obtenido del control y el valor obtenido del sujeto es indicativa de que el sujeto está padeciendo la enfermedad o síndrome o es indicativa de que el sujeto padecerá la enfermedad o síndrome.

15. El método de acuerdo con el párrafo 14, donde una transición de frecuencias de interacción bajas a altas para una región cromosómica es indicativa de la localización de un punto de interrupción.

16. El método de acuerdo con 14, donde un patrón invertido de frecuencias de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una inversión.

17. El método de acuerdo con el párrafo 14, donde una reducción en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control, en combinación con un aumento en la frecuencia de interacción ADN-ADN para regiones cromosómicas más distantes, es indicativa de una deleción.

18. El método de acuerdo con el párrafo 14, donde un aumento en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una duplicación o inserción.

19. El método del párrafo 18 donde existe una disminución en las frecuencias de interacción ADN-ADN para una región cromosómica entre regiones más distantes de ADN en la muestra del sujeto en comparación con el control.

20. Un método de acuerdo con cualquiera de los párrafos 14-19, donde se usa cariotipado espectral y/o FISH antes de realizar dicho método.

21. El método de acuerdo con cualquiera de los párrafos 14-20, donde la enfermedad es una enfermedad genética.

22. El método de acuerdo con cualquiera de los párrafos 14-21, donde la enfermedad es cáncer.

23. Un método de diagnóstico de una enfermedad o síndrome causado por o asociado con un cambio en las interacciones ADN-ADN que comprende la etapa de: realizar las etapas (a)-(i) de cualquiera de los párrafos 1-11, donde la etapa (a) comprende proporcionar una muestra de ADN entrecruzado de un sujeto; y donde dicho método comprende la etapa adicional de: (j) identificar uno o más loci que han experimentado un reordenamiento genómico que está asociado con una enfermedad.

24. 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) de cualquiera de los párrafos 1-11, donde la etapa (a) comprende proporcionar ADN entrecruzado de la muestra;

donde 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.

25. Un método para detectar la localización de un punto de interrupción equilibrado y/o desequilibrado (por ejemplo, una translocación) que comprende la etapa de:

(a) realizar las etapas (a) a (i) de cualquiera de los párrafos 1-11; y

(b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control;

donde una transición de frecuencia de interacción ADN-ADN baja a alta para una región cromosómica en la muestra en comparación con el control es indicativa de la localización de un punto de interrupción.

26. Un método para detectar la localización de una inversión equilibrada y/o desequilibrada que comprende las etapas de:

(a) realizar las etapas (a) a (i) de cualquiera de los párrafos 1-11; y

(b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control;

donde un patrón invertido de frecuencias de interacción ADN-ADN para una región cromosómica para la muestra en comparación con el control es indicativo de una inversión.

27. Un método para detectar la localización de una deleción que comprende las etapas de:

(a) realizar las etapas (a) a (i) de cualquiera de los párrafos 1-11; y

(b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control;

donde una reducción en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra en comparación con el control es indicativa de deleción.

28. Un método para detectar la localización de una duplicación o inserción que comprende las etapas de:

(a) realizar las etapas (a) a (i) de cualquiera de los párrafos 1-11; y

(b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control;

donde un aumento en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una duplicación o inserción.

Se presentan aspectos de la presente invención en las reivindicaciones adjuntas.

En un primer aspecto, se proporciona un método para analizar la frecuencia de interacción de una secuencia nucleotídica diana con una o más secuencias nucleotídicas de interés (por ejemplo, uno o más loci 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 nucleotídicas entrecruzadas; (d) revertir el entrecruzamiento; (e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria; (f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios de digestión con enzima de restricción secundaria disponibles que flanquean la una o más secuencias nucleotídicas de interés; (g) amplificar la una o más secuencias nucleotídicas de interés usando al menos dos cebadores oligonucleotídicos, donde cada cebador hibrida con las secuencias de ADN que flanquean las secuencias nucleotídicas de interés; (h) hibridar la secuencia o secuencias amplificadas con una serie; y (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 nucleotídica diana con una o más secuencias nucleotídicas (por ejemplo, uno o más loci 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 nucleotídicas entrecruzadas; (d) revertir el entrecruzamiento; (e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria; (f) circularizar las secuencias nucleotídicas; (g) amplificar la una o más secuencias nucleotídicas que están ligadas con la secuencia nucleotídica diana; (h) opcionalmente hibridar las secuencias amplificadas con una serie; y (i) determinar la frecuencia de interacción entre las secuencias de ADN.

También se describe en este documento una secuencia nucleotídica circularizada que comprende una primera y una segunda secuencia nucleotídica, donde cada extremo de la primera y segunda secuencia nucleotídica están separadas por diferentes sitios de reconocimiento por enzimas de restricción, y donde dicha primera secuencia nucleotídica es una secuencia nucleotídica diana y dicha segunda secuencia nucleotídica se puede obtener por entrecruzamiento de ADN genómico.

En un cuarto aspecto se proporciona un método para preparar una secuencia nucleotídica circularizada que comprende las etapas de: (a) proporcionar una muestra de1 ADN entrecruzado; (b) digerir el ADN entrecruzado con una enzima de restricción primaria; (c) ligar las secuencias nucleotídicas entrecruzadas; (d) revertir el entrecruzamiento; (e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria y; (f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios de digestión con enzima de

restricción secundaria disponibles que flanquean la una o más secuencias nucleotídicas de interés que circulariza las secuencias nucleotídicas.

También se describe en este documento un método para analizar la frecuencia de interacción de una secuencia nucleotídica diana con una o más secuencias nucleotídicas (por ejemplo, uno o más loci genómicos) que comprende el uso de la secuencia nucleotídica circularizada.

También se describe en este documento una serie de sondas inmovilizadas en un soporte que comprende una o más sondas que hibridan o son capaces de hibridar con la secuencia nucleotídica circularizada.

También se describe en este documento un conjunto de sondas complementarias en secuencia a la secuencia de ácido nucleico adyacente a cada uno de los sitios de reconocimiento con enzima de restricción primaria de una enzima de restricción primaria en ADN genómico.

También se describe en este documento un proceso para preparar un conjunto de sondas que comprende las etapas de: (a) identificar cada uno de los sitios de reconocimiento con enzima de restricción primaria para una enzima de restricción primaria en ADN genómico; (b) diseñar sondas que sean capaces de hibridar con la secuencia adyacente a cada uno de los sitios de reconocimiento con enzima de restricción primaria en el ADN genómico; (c) sintetizar las sondas; y (d) combinar las sondas juntas para formar un conjuntos de sondas o sustancialmente un conjunto de sondas.

También se describe en este documento un conjunto de sondas o sustancialmente un conjunto de sondas obtenida u obtenible mediante el proceso descrito en este documento.

También se describe en este documento una serie que comprende la serie de sondas o sustancialmente el conjunto de sondas descrito en este documento.

También se describe en este documento una serie que comprende el conjunto de sondas de acuerdo con lo descrito en este documento.

También se descrine en este documento un proceso para preparar una serie que comprende la etapa de inmovilizar en un soporte sólido sustancialmente la serie de sondas o sustancialmente el conjunto de sondas descrito es este documento.

También se describe en este documento un proceso para preparar una serie que comprende la etapa de inmovilizar en un soporte sólido la serie de sodas o el conjunto de sondas descrito en este documento.

También se describe en este documento una serie obtenida u obtenible mediante el método descrito en este documento.

En un decimoquinto aspecto se proporciona un método para identificar una o más interacciones ADN-ADN que son indicativas de un estado patológico particular que comprende la etapa de realizar las etapas (a)-(i) del primer y segundo aspecto de la presente invención, donde en la etapa (a) se proporciona una muestra de ADN entrecruzado de una célula enferma y no enferma, y donde una diferencia entre la frecuencia de interacción entre las secuencias de ADN de las células enfermas y no enfermas indica que la interacción ADN-ADN es indicativa de un estado patológico particular.

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

En un decimoséptimo aspecto se proporciona un método de diagnóstico de una enfermedad o síndrome causado por o asociado con un cambio en una interacción ADN-ADN que comprende la etapa de: realizar las etapas (a)-(i) del primer y segundo aspecto de la presente invención, donde la etapa (a) comprende proporcionar una muestra de ADN entrecruzado de un sujeto; y donde dicho método comprende la etapa adicional de: (j) identificar uno o más loci que han experimentado un reordenamiento genómico que está asociado con una enfermedad.

En un decimoctavo 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 primer y segundo aspecto de la presente invención, donde la etapa (a) comprende proporcionar ADN entrecruzado de la muestra; donde 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.

En un decimonoveno aspecto se proporciona un método para detectar la localización de un punto de interrupción equilibrado y/o desequilibrado (por ejemplo, una translocación) que comprende la etapa de: (a) realizar las etapas (a) a (i) del primer y segundo aspecto de la presente invención; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; donde una transición de frecuencia de interacción ADN-ADN baja a alta para una región cromosómica en la muestra en comparación con el control es indicativa de la localización de un punto de interrupción.

En un vigésimo aspecto se proporciona un método para detectar la localización de una inversión equilibrada y/o desequilibrada que comprende las etapas de: (a) realizar las etapas (a) a (i) del primer y segundo aspecto de la presente invención; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; donde un patrón invertido de frecuencias de interacción ADN-ADN para una región cromosómica para la muestra en comparación con el control es indicativo de una inversión.

En un vigésimo primer aspecto se proporciona un método para detectar la localización de una deleción que comprende las etapas de: (a) realizar las etapas (a) a (i) del primer y segundo aspecto de la presente invención; (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; donde una reducción en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra en comparación con el control es indicativa de deleción.

En un vigésimo segundo aspecto se proporciona un método para detectar la localización de una duplicación que comprende las etapas de: (a) realizar las etapas (a) a (i) del primer y segundo aspecto de la presente invención; y (b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control; donde un aumento o una disminución en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativa de una duplicación o inserción.

También se describe en este documento un agente obtenido u obtenible mediante el método de ensayo descrito en este documento.

También se describe en este documento el uso de la secuencia nucleotídica circularizada para identificar una o más interacciones ADN-ADN en una muestra.

También se describe en este documento el uso de una secuencia nucleotídica circularizada para el diagnóstico o pronóstico de una enfermedad o síndrome causado por o asociado con un cambio en una interacción ADN-ADN.

También se describe en este documento el uso de la serie de sondas o del conjunto de sondas descrito en este documento para identificar una o más interacciones ADN-ADN en una muestra.

También se describe en este documento el uso de la serie de sondas o del conjunto de sondas descrito en este documento para el diagnóstico o pronóstico de una enfermedad o síndrome causado por o asociado con un cambio en una interacción ADN-ADN.

También se describe en este documento el uso de la serie descrita en este documento para identificar una o más interacciones ADN-ADN en una muestra.

También se describe en este documento el uso de la serie descrita en este documento para el diagnóstico o pronóstico de una enfermedad o síndrome causado por o asociado con un cambio en una interacción ADN-ADN.

También se describe en este documento un método, una serie de sondas, un conjunto de sondas, un proceso, una serie, un método de ensayo, un agente, o un uso sustancialmente como se describe en este documento y con referencia a cualquiera de los Ejemplos o Figuras.

Realizaciones preferidas

Preferiblemente, la reacción de ligamiento en la etapa (f) provoca la formación de círculos de ADN.

Preferiblemente, la secuencia nucleotídica diana se selecciona entre el grupo que consiste en un reordenamiento genómico, promotor, un potenciador, un silenciador, un aislante, una región de adhesión a la matriz, una región de control del locus, una unidad de transcripción, un origen de replicación, un punto caliente de recombinación, un punto de interrupción de translocación, un centrómero, un telómero, una región densa en genes, una región pobre en genes, un elemento repetitivo y un sitio de integración (viral).

Preferiblemente, la secuencia nucleotídica diana es una secuencia nucleotídica que está asociada con o causa una enfermedad, o está localizada hasta o más de 15 Mb en un molde de ADN lineal desde un locus que está asociado con o causa una enfermedad.

Preferiblemente, la secuencia nucleotídica diana se selecciona entre el grupo que consiste en: *AML1, MLL, MYC, BCL, BCR*, *ABL1*, *IGH*, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCRαlδ*, *TCRβ* y *HOX* u otros loci asociados con enfermedad. Dichos nucleótidos diana se describen en "Catalogue of Unbalanced Chromosome Aberrations in Man" 2ª edición. Albert Schinzel. Berlín: Walter de Gruyter, 2001. ISBN 3-11-011607-3.

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

Preferiblemente, la enzima de restricción primaria se selecciona entre el grupo que consiste en *Bgl*II, *Hind*III, *Eco*RI, *Bam*HI, *Spe*I, *Pst*I y *Nde*I.

Preferiblemente, la enzima de restricción secundaria es una enzima de restricción que reconoce un sitio de reconocimiento de secuencia de nucleótidos de 4 a 5 pb.

Preferiblemente, el sitio de reconocimiento por enzima de restricción secundaria está localizado a más de aproximadamente 350 pb desde el sitio de restricción primario en la secuencia nucleotídica diana. Preferiblemente, la secuencia nucleotídica está marcada.

Por motivos de referencia solamente, se describen sondas y series para su uso con los métodos de la invención.

Preferiblemente, las sondas descritas en este documento son complementarias en secuencia a la secuencia de ácido nucleico adyacente a cada lado de cada uno de los sitios de reconocimiento por enzima de restricción primaria de una enzima de restricción primaria en ADN genómico.

Preferiblemente, las sondas descritas en este documento son complementarias en secuencia a la secuencia de ácido nucleico que está menos de 300 pb desde cada uno de los sitios de reconocimiento con enzima de restricción primaria de una enzima de restricción primaria en ADN genómico.

Preferiblemente, las sondas descritas en este documento son complementarias a la secuencia que está menos de 300 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria de una enzima de restricción primaria en ADN genómico.

Preferiblemente, las sondas descritas en este documento son complementarias a la secuencia que está entre 200 y 300 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria de una enzima de restricción primaria en ADN genómico.

Preferiblemente, las sondas descritas en este documento son complementarias a la secuencia que está entre 100 y 200 pb o 0 a 100 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria de una enzima de restricción primaria en ADN genómico.

Preferiblemente, dos o más sondas descritas en este documento son capaces de hibridar con la secuencia adyacente a cada sitio de reconocimiento por enzima de restricción primaria de una enzima de restricción primaria en el ADN genómico.

Preferiblemente, las sondas descritas en este documento solapan o solapan parcialmente.

Preferiblemente, el solapamiento es de menos de 10 nucleótidos.

Preferiblemente, la secuencia de sonda descrita en este documento corresponde a todo o parte de la secuencia entre cada uno de los sitios de reconocimiento con enzima de restricción primaria de una enzima de restricción primaria y cada uno de los primeros sitios de reconocimiento con enzima de restricción secundaria adyacentes de una enzima de restricción secundaria.

Preferiblemente, cada sonda descrita en este documento es al menos un oligómero de 25 monómeros.

Preferiblemente, cada sonda descrita en este documento es un oligómero de 25-60 monómeros.

Preferiblemente, las sondas descritas en este documento son productos de amplificación por PCR.

Preferiblemente, la sonda descrita en este documento comprende aproximadamente 300.000-400.000 sondas.

Preferiblemente, la serie descrita en este documento comprende aproximadamente 385.000 o más sondas, preferiblemente, aproximadamente 750.000 sondas, más preferiblemente, 6 x 750.000 sondas.

Preferiblemente, la serie descrita en este documento comprende o consiste en una representación del genoma completo de una especie dada a resolución inferior.

Preferiblemente, una de cada 2, 3, 4, 5, 6, 7, 8, 9 o 10 sondas ordenadas en un molde de cromosoma lineal está contenida en la serie descrita en este documento.

Preferiblemente, una transición de frecuencias de interacción bajas a altas es indicativa de la localización de un punto de interrupción equilibrado y/o desequilibrado.

Preferiblemente, 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 desequilibrada.

Preferiblemente, 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 desequilibrada.

Preferiblemente, 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.

Preferiblemente, se usa cariotipado espectral y/o FISH antes de realizar dicho método.

Preferiblemente, la enfermedad es una enfermedad genética.

Preferiblemente, la enfermedad es cáncer.

Preferiblemente, las dos o más secuencias amplificadas están marcadas de forma diferencial.

Preferiblemente, las dos o más secuencias amplificadas están marcadas de forma idéntica cuando las secuencias residen en diferentes cromosomas.

Preferiblemente, las dos o más secuencias amplificadas están marcadas de forma idéntica cuando las secuencias residen en el mismo cromosoma a una distancia que está suficientemente lejos para un solapamiento mínimo entre las señales de interacción ADN-ADN.

Preferiblemente, el diagnóstico o pronóstico es de diagnóstico o pronóstico prenatal.

Ventajas

La presente invención tiene varias ventajas. Estas ventajas serán evidentes en la siguiente descripción.

A modo de ejemplo, la presente invención es ventajosa ya que proporciona *inter alia* secuencias nucleotídicas comercialmente útiles, y procesos.

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

A modo de ejemplo adicional, la presente invención es ventajosa ya que usando tecnología 3C convencional, cada interacción ADN-ADN individual debe analizarse por una reacción de PCR única que contenga un par único de cebadores. El análisis de alto rendimiento por lo tanto es solamente posible si la PCR está automatizada, pero los costes de tantos cebadores serán demasiado elevados. Por consiguiente, el análisis de alto rendimiento (genoma completo) de interacciones ADN-ADN no es viable con tecnología 3C convencional. Por el contrario, la presente invención ahora permite 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 resolución del análisis.

A modo de ejemplo adicional, la presente invención es ventajosa ya que usando tecnología 3C convencional, la exploración está influenciada hacia aquellas secuencias de ADN para las cuales se diseñaron cebadores oligonucleotídicos, dispuestos e incluidos en el análisis. La elección de dichos cebadores oligonucleotídicos se basa típicamente en el conocimiento respecto a la posición de, por ejemplo, potenciadores (distantes) y/u otros elementos reguladores/sitios hipersensibles que se cree que entrecruzarán con la secuencia nucleotídica que se está investigando. Por tanto, la 3C convencional está influenciada hacia el diseño de cebadores de PCR que se incluyen en la etapa de amplificación de PCR, mientras que 4C no está influenciada y puede usarse para buscar en el genoma completo elementos de interacción ADN. Esto es porque la amplificación de secuencias entrecruzadas en 4C no se basa en el conocimiento predicho de secuencias que entrecruzan con la secuencias nucleotídica que se está investigando. En su lugar, en una realización de 4C, las secuencias que entrecruzan con la primera secuencia nucleotídica (diana) pueden amplificarse usando cebadores de PCR que hibridan con esa secuencias nucleotídica.

Por tanto, la presente invención permite una exploración no influenciada del genoma completo para interacciones ADN-ADN.

A modo de ejemplo adicional, la presente invención es ventajosa porque usando tecnología 3C convencional solamente se permite la amplificación selectiva de una única interacción ADN-ADN. Esta no es informativa cuando hibrida con una serie. La tecnología se ha mejorado de modo que todos los fragmentos que interaccionan con una primera secuencia nucleotídica (diana) ahora se amplifican, por ejemplo, se amplifican de forma selectiva.

A modo de ejemplo adicional, la presente invención es ventajosa porque la tecnología 4C puede usarse para detectar aberraciones genéticas equilibradas o desequilibradas - tales como todos los tipos de translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos - en ácido nucleico, por ejemplo, cromosomas. La tecnología 4C (que mide la proximidad de fragmentos de ADN) puede incluso determinar una predisposición del sujeto a adquirir ciertas translocaciones, deleciones, inversiones, duplicaciones u otros reordenamientos genómicos (por ejemplo, translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos equilibrados o desequilibrados). Una ventaja sobre las estrategias actuales es que no se requiere conocer la posicione exacta del cambio porque la resolución de la tecnología 4C es tal que puede usarse para detectar reordenamientos incluso cuando el "señuelo-4C" (definido por los sitios de reconocimiento con enzima de restricción primaria y secundaria que se analizan) está localizado lejos (por ejemplo, hasta una megabase o incluso más) del cambio. Otra ventaja es que la tecnología 4C permite el mapeado preciso de cambios ya que puede usarse para definir los dos sitios de restricción (primarios) entre los que ocurren cambios. Otra ventaja es que las células no tienen que cultivarse antes de su fijación. Por tanto, por ejemplo también pueden analizarse tumores sólidos para reordenamientos genómicos.

A modo de 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 usarse no solamente en el diagnóstico de enfermedad sino también en el pronóstico de enfermedad.

A modo de ejemplo adicional, el diseño de serie de acuerdo con la presente invención es particularmente ventajoso en comparación con las series de agrupamiento genómico existentes - tales como las series de agrupamiento genómico de Nimblegen - ya que el diseño permite la representación de una parte mucho más grande del genoma por una única serie. A modo de ejemplo, para una enzima de restricción que reconoce una secuencia hexa-nucleotídica, aproximadamente 3 series con aproximadamente 385.000 sondas cada una será suficiente para cubrir, por ejemplo, el genoma completo humano o de ratón. Para una enzima de restricción que reconoce más de 6 pb, puede usarse una única serie de aproximadamente 385.000 sondas para cubrir, por ejemplo, el genoma completo humano o de ratón. Las ventajas del diseño de serie son que: (1) cada sonda es informativa ya que analiza un evento de ligamiento independiente, facilitando enormemente la interpretación de los resultados; y (2) puede depositarse puntualmente una gran representación del genoma en una única serie que es rentable.

La tecnología 4C puede usarse ventajosamente para el mapeado fino de reordenamientos mal caracterizados originalmente detectados por enfoques citogenéticos (microscopia óptica, FISH, SKY, etc.).

La tecnología 4C puede usarse ventajosamente para la exploración simultánea en una única serie para combinaciones de reordenamientos que han sucedido cerca de múltiples loci.

Breve descripción de las figuras

Figura 1
El principio de tecnología 3C.

Figura 2
(a) El principio de una realización de tecnología 4C. El análisis 3C se realiza de forma virtual, con por ejemplo HindIII (H) como enzima de restricción. Después de revertir los entrecruzamientos, la mezcla de ADN contendrá una primera secuencia nucleotídica (diana) ligada a muchos fragmentos diferentes. Estos fragmentos se amplificarán y marcarán usando métodos de amplificación - tales como PCR inversa - en, por ejemplo, *Dpn*II, círculos, usando cebadores específicos de la primera secuencia nucleotídica (diana). Los productos de amplificación marcados pueden hibridarse a las series como se describe en este documento. *Hind*III y *Dpn*II, se dan como ejemplos, pero también pueden usarse otras combinaciones de enzimas de restricción - tales como cortadoras de 6 u 8 - y 4 o 5. (b) Resultados de PCR separados por electroforesis en gel de dos muestras independientes de hígado fetal (L1, L2) y cerebro (B1, B2). (c) Representación esquemática de la localización de las sondas de microserie. Las sondas se diseñaron con 100 pb de sitios HindIII. Por tanto, cada sonda analiza un posible patrón de ligamiento.

Figura 3
La tecnología 4C detecta el entorno genómico de β-globina (cromosoma 7). Se muestran proporciones no procesadas (señales 4C para HS2 de β-globina divididas por señales obtenidas para la muestra de control) para

sondas localizadas en regiones genómicas de ~35 Mb en el cromosoma 10, 11, 12, 14, 15, 7 y 8 de ratón (de arriba abajo; las regiones mostradas están a distancia idéntica de cada correspondiente centrómero). Obsérvese el gran grupo de señales fuertes alrededor del señuelo (globina) en el cromosoma 7 (fila 6), que demuestra que la tecnología 4C detecta fragmentos genómicos cercanos en el molde de cromosoma lineal (de acuerdo con el hecho de que las frecuencias de interacción son inversamente proporcionales a la separación del sitio genómico). Obsérvese que la región ligada en cis alrededor del señuelo que muestra altas intensidades de señal es grande (>5 Mb), lo que implica, por ejemplo, que pueden detectarse translocaciones incluso con señuelos de más de 1 Mb desde el punto de interrupción.

Figura 4
La tecnología 4C detecta el entorno genómico de Rad23A (cromosoma 8). Se muestran proporciones no procesadas (señales 4C para Rat23A divididas por la señal obtenida para la muestra de control) para sondas localizadas en regiones genómicas de ~15 Mb o más en el cromosoma 10, 11, 12, 14, 15, 7 y 8 de ratón (de arriba abajo; las regiones mostradas están a distancia idéntica de cada correspondiente centrómero). Obsérvese el gran grupo de señales fuertes alrededor del señuelo (Rad23A) en el cromosoma 8 (fila 7), lo que demuestra que la tecnología 4C detecta fragmentos genómicos cercanos en el molde de cromosoma lineal (de acuerdo con el hecho de que las frecuencias de interacción son inversamente proporcionales a la separación del sitio genómico). Obsérvese que la región ligada en cis alrededor del señuelo que muestra altas intensidades de señal es grande (>5 Mb), lo que implica, por ejemplo, que pueden detectarse translocaciones incluso con señuelos de más de 1 Mb desde el punto de interrupción.

Figura 5
Las interacciones 4C de β-globina en el cromosoma 7 (~135 Mb) para un tejido en transcripción (hígado fetal) y un tejido no en transcripción (cerebro fetal) (analizado por un enfoque de media de ejecución). Obsérvese que interacciones de rango largo con β-globina difieren entre tejidos (probablemente dependientes del estado de transcripción del gen). Independientemente del tejido las señales 4C fuertes demarcan una región grande (>5 Mb) alrededor del señuelo.

Figura 6
Uros y Eraf interaccionan con β-globina en células de hígado fetal. El enfoque 4C revela que dos genes, Eraf y Uros, interaccionan sobre >30 Mb con el locus de β-globina localizado a ~30 Mb. Estas dos interacciones se descubrieron previamente por una tecnología diferente (Hibridación In Situ de Fluorescencia) como se describe en Osborne et al., Nature Genetics 36, 1065 (2004). Este ejemplo muestra que interacciones de rango largo detectadas por tecnología 4C pueden verificarse por FISH y reflejar verdaderamente la proximidad nuclear.

Figura 7
La tecnología 4C identifica de forma precisa transiciones entre regiones genómicas no relacionadas que están ligadas en cis. Para estos experimentos se usaron ratones transgénicos que contienen un casete de región de control del locus de β-globina humana (LCR) (~20 kb) insertado (mediante 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 homocigóticos para esta inserción. Se usó un fragmento HindIII dentro del casete de integración (HS2) como "señuelo-4C". Los datos muestran que la tecnología 4C define de forma precisa ambos extremos del casete transgénico (fila inferior: solamente sondas en el LCR humano (~20 kb) dan señales 4C y no sondas en el resto de la secuencia de β-globina humana de ~380 kb) y revela claramente la posición de integración en el cromosoma 8 de ratón (panel superior: compara señales en el cromosoma 8 (para la posición de integración, véase la flecha) con señales en otros 6 cromosomas de ratón) (se representan cromosomas completos). Este ejemplo muestra que la tecnología 4C puede usarse para detectar la posición genómica de fragmentos de ADN integrados de forma ectópica (virus, transgén, etc.). Muestra que pueden identificarse de forma precisa transiciones entre regiones genómicas no relacionadas que están ligadas en cis, lo que puede usarse para identificar puntos de interrupción genómica y patrones de translocación.

Figura 8
La tecnología 4C produce datos reproducibles ya que el perfil para HS2 y β-globina son muy similares. Se realizaron cuatro experimentos 4C biológicamente independientes en hígados fetales E14.5, usando el gen β-mayor de β-globina (2 filas superiores) o HS2 de β-globina (dos filas inferiores) como señuelos. Estos señuelos están ~40 kb separados en el molde de cromosoma lineal pero demostraron previamente estar cercanos en el 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 20-20 Mb desde el locus de β-globina. Los datos muestran alta reproducibilidad entre experimentos independientes y demuestran que dos fragmentos cercanos en el espacio nuclear compartes patrones de interacción localizados en cualquier otra parte en el genoma.

Figura 9
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 un paciente con translocación (A;B) (parte inferior). Las intensidades de señal que representan frecuencias de interacción ADN-ADN (eje Y) se representan para sondas ordenadas en moldes de cromosoma lineal (eje X). En células normales, se detectan interacciones ADN-ADN frecuentes en el cromosoma A alrededor de la secuencia X. En células de paciente, se observa una reducción del 50% en las frecuencias de interacción para sondas en el cromosoma A localizadas en el otro lado del punto de interrupción (BP) (compárese la curva gris (paciente) con la línea negra (persona sana)). Además, la translocación pone parte de cromosoma B en cercana proximidad física a la secuencia X, y ahora se observan interacciones ADN-ADN frecuentes para esta región en el cromosoma B. La transición abrupta de frecuencias de interacción bajas a altas en este cromosoma marca la localización de su punto de interrupción.

Figura 10
Puede detectarse una o más inversiones (equilibradas) por tecnología 4C. Se observan patrones invertidos de frecuencias de interacción ADN-ADN (medidas por tecnología 4C como intensidades de señal de hibridación) en sujetos enfermos (curva sólida) en comparación con sujetos no enfermos (curva punteada), lo que revela la presencia y tamaño de la inversión.

Figura 11
Detección de una o más deleciones heterocigóticas por tecnología 4C. Sondas con frecuencias de interacción ADN-ADN reducidas (medidas por tecnología 4C como intensidades de señal de hibridación) en sujetos enfermos (curva gris) en comparación con sujetos no enfermos (curva negra), revelan la posición y tamaño de la región delecionada. Las señales de hibridación residuales en la región delecionada del sujeto enfermo provienen del alelo intacto (deleción heterocigótica). La deleción está acompañada típicamente por un aumento en las intensidades de señal para sondas localizadas directamente más allá de la región delecionada (obsérvese que la curva gris está por encima de la curva blanca a mano derecha de la deleción), ya que estas regiones están en cercana proximidad a la secuencia 4C (señuelo).

Figura 12
Duplicación detectada por tecnología 4C. Sondas con señales de hibridación aumentadas en un paciente (curva sólida) en comparación con un sujeto normal (curva punteada) indican la posición y tamaño de duplicación. La duplicación detectada por tecnología 4C está típicamente acompañada por señales de hibridación disminuidas en sujetos enfermos frente a no enfermos para sondas más allá de la región duplicada (la duplicación aumenta su separación de sitio genómico desde la secuencia 4C).

Figura 13
Interacciones de rango largo con β-globina reveladas por tecnología 4C. **a**, Proporciones no procesadas de 4C sobre señales de hibridación de control, que revelan interacciones de HS2 de β-globina con el cromosoma 7 y dos cromosomas no relacionados (8 y 14). **b-c**, Datos no procesados para dos muestras independientes de hígado fetal (parte superior, en rojo) y cerebro fetal (parte inferior, en azul) representados a lo largo de dos regiones diferentes de 1-2 Mb en el cromosoma 7. Se observan grupos altamente reproducibles de interacciones en las dos muestras de hígado fetal (**b**) o las dos muestras de cerebro (**c**). **d-e**, Datos de media de ejecución para las mismas regiones. La tasa de descubrimiento de falso se estableció al 5% (línea punteada). **f**, Representación esquemática de regiones de interacción con β-globina activa (hígado fetal, parte superior) e inactiva (cerebro fetal, parte inferior) en el cromosoma 7.

Figura 14
La β-globina activa e inactiva interacciona con regiones cromosómicas activas e inactivas, respectivamente. **a**, Comparación entre interacciones de rango largo de β-globina en hígado fetal (media de ejecución 4C, parte superior), análisis de expresión en microserie en hígado fetal (escalo log, parte central) y la localización de genes (parte inferior) representada a lo largo de una región de 4 Mb que contienen el gen Uros (~30 Mb desde la β-globina), que muestra que la β-globina activa interacciona de forma preferente con otros genes transcritos de forma activa. **b**, La misma comparación en cerebro fetal alrededor de un grupo génico OR localizado ~38 Mb desde la globina, que muestra que la β-globina inactiva interacciona de forma preferente con regiones inactivas. **c**, Caracterización de regiones que interaccionan con β-globina en hígado fetal (izquierda) y cerebro (derecha) en términos de contenido y actividad génica.

Figura 15
Rad23A expresado de forma ubicua interacciona con regiones activas muy similares en hígado fetal y cerebro. **a**, Representación esquemática de regiones en el cromosoma 8 que interaccionan con Rad23A activo en hígado fetal (parte superior, rojo) y cerebro (parte inferior, azul). **b**, Comparación entre interacciones de rango largo de Rad23A (media de ejecución 4C) y análisis de expresión en microserie (escala log) en hígado fetal (dos paneles superiores), interacciones de rango largo de Rad23A (media de ejecución 4C) y análisis de expresión en microserie (escala log) en cerebro fetal (panel 3 y 4) y la localización de genes (panel inferior) representada 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 términos de contenido y actividad génica.

Figura 16
Crio-FISH confirma que la tecnología 4C identifica realmente regiones de interacción. **a**, ejemplo de parte de una crio-sección (200 nm) que muestra más de 10 núcleos, algunos de los cuales contienen el locus de β-globina (verde) y/o Uros (rojo). Debido al seccionamiento, muchos núcleos no contienen señales para estos dos loci. **b-d**, ejemplos de señales completa (b) y parcialmente (c) solapantes y señales de contacto (d), que se valoraron todas como positivas para interacción. **e-g**, Ejemplos que contiene alelos no de contacto (e-f) y un núcleo que contiene solamente β-globina (g), que se valoraron todos como negativos para interacción. **h-i**, Representación esquemática de resultados de crio-FISH. Los porcentajes de interacción con β-globina (h) y Rad23A (i) se indican encima de los cromosomas para regiones identificadas positivamente (punta de flecha roja) e identificadas negativamente (punta de flecha azul) por tecnología 4C. Se usaron los mismos BAC para los dos tejidos. Se indican frecuencias de interacción medidas por crio-FISH entre dos grupos génicos OR distantes en hígado fetal y cerebro debajo de los cromosomas.

Figura 17
El análisis 4C de HS2 y β-mayor dan resultados altamente similares. (a) (Datos 4C son procesar de cuatro muestras independientes de hígado E14.5 muestran un patrón muy similar de interacción con HS2 (parte superior) y β-mayor (parte inferior). (b) Existe un gran solapamiento entre sondas valoradas positivas para interacción en el experimento de HS2 y sondas que valoraron positivas para interacción en el experimento de β-mayor.

Figura 18
Una comparación entre interacciones en cis y en trans. (a) Datos 4C no procesados de dos experimentos independientes que muestran interacciones de β-globina con una región identificada positivamente en cis (cromosoma 7, parte superior) y una región en trans que contiene el locus α-globina (chr.11, parte inferior). (b) Datos 4C no procesados de dos experimentos independientes que muestran interacciones de Rad23A con una región identificada positivamente en cis (cromosoma 8, parte superior) y una región en trans que aparecía en la parte superior cuando se clasificaba de acuerdo a la media de ejecución más alta. Ninguna de las regiones en trans cumplía las condiciones rigurosas que permitían la identificación de regiones de larga interacción en cis.

Figura 19
Regiones que interaccionan con β-globina también contactan frecuentemente entre sí. Dos regiones (separadas casi 60 Mb), que contienen genes transcritos de forma activa e identificados por tecnología 4C por interaccionar con β-globina en hígado fetal, mostraron frecuencias de co-localización por crio-FISH del 5,5%, que fue significativamente mayor que las frecuencias de co-localización de fondo.

Descripción detallada de la invención

Tecnología 3C

El método 3C se ha descrito en 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, 3C se realiza digiriendo ADN entrecruzado con una enzima de restricción primaria seguido por ligamiento ha concentraciones de ADN muy bajas. En estas condiciones, el ligamiento de fragmentos entrecruzados, que es intramolecular, se favorece fuertemente sobre el ligamiento de fragmentos aleatorios, que es intermolecular. El entrecruzamiento después se revierte y se detecta productos de ligamiento individuales y se cuantifican por la reacción en cadena de la polimerasa (PCR) usando cebadores específicos de locus. La frecuencia de entrecruzamiento (X) de dos loci específicos se determina por reacciones de PCR cuantitativa usando moldes de control y entrecruzados, y X se expresa como la relación de la cantidad del producto obtenido con el molde entrecruzado y con el molde de control.

De acuerdo con la presente invención, se prepara un molde 3C usando los métodos descritos por Splinter et al., (2004) Methods Enzymol. 375, 493-507 (es decir, fijación con formaldehído, digestión con enzima de restricción (primaria), re-ligamiento de fragmentos de ADN entrecruzados y purificación de ADN). En resumen, se fija una muestra - tal como células, tejido o núcleos - usando un agente de entrecruzamiento - tal como formaldehído. La digestión con enzima de restricción primaria después se realiza de modo que se digiera el ADN en el contexto del núcleo entrecruzado. El ligamiento intramolecular después se realiza a bajas concentraciones de ADN (por ejemplo, aproximadamente 3,7 ng/µl), que favorece el ligamiento entre fragmentos de ADN entrecruzado (es decir, ligamiento intramolecular) sobre el ligamiento entre fragmentos de ADN no entrecruzado (es decir, ligamiento intermolecular o aleatorio). A continuación, los entrecruzamientos se revierten y puede purificarse el ADN. El molde 3C que se produce contiene fragmentos de restricción que se ligan porque estaban originalmente cercanos en el espacio nuclear.

Como se usa 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 nucleotídica (diana) y la secuencia nucleotídica que se ha ligado. Por consiguiente, el sitio de reconocimiento primario está localizado entre la primera secuencia nucleotídica (diana) y la secuencia nucleotídica ligada (es decir, la segunda secuencia ligada).

Secuencia nucleotídica

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

La secuencia nucleotídica puede ser ADN o ARN de origen genómico, sintético o recombinante, por ejemplo, ADNc. Por ejemplo, pueden prepararse secuencias nucleotídicas recombinantes usando técnicas de clonación por PCR. Esto implicará preparar un par de cebadores que flanquean una región de la secuencia que se desea clonar, poner los cebadores en contacto con ARNm o ADNc obtenido de, por ejemplo, un mamífero (por ejemplo, célula animal o humana) o célula no de mamífero, realizar una reacción en cadena de la polimerasa (PCR) en condiciones que conseguirán 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 adecuados de reconocimiento por enzimas de restricción de modo que el ADN amplificado pueda clonarse en un vector de clonación adecuado.

La secuencia nucleotídica puede ser bicatenario o monocatenaria representando la hebra con sentido o antisentido o combinaciones de las mismas.

Para algunos aspectos, se prefiere que la secuencia nucleotídica sea ADN monocatenario - tal como cebadores y sondas monocatenarias.

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

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

Para algunos aspectos, se prefiere que la secuencia nucleotídica sea ADN cromosómico.

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

Los sitios de reconocimiento por enzima de restricción primaria y secundaria serán diferentes entre sí y típicamente existirán solamente una vez en la secuencia nucleotídica.

En un aspecto, se proporciona una secuencia nucleotídica circularizada que comprende una primera secuencia nucleotídica y una segunda secuencia nucleotídica (por ejemplo, ligada) separada (por ejemplo, dividida o separada) por un sitio de reconocimiento por enzima de restricción primaria y secundaria, donde dicha primera secuencia nucleotídica es una secuencia nucleotídica diana y dicha segunda secuencia nucleotídica se puede obtener por entrecruzamiento de ADN genómico (por ejemplo, *in vivo* o *in vitro*). Los sitios de reconocimiento por enzima de restricción primaria y secundaria serán diferentes entre sí y típicamente existirán solamente una vez en la secuencia nucleotídica.

También se describe en este documento una secuencia nucleotídica circularizada que comprende una primera secuencia nucleotídica y una segunda secuencia nucleotídica (por ejemplo, ligada) separada (por ejemplo, dividida o separada) por un sitio de reconocimiento por enzima de restricción primaria y secundaria, donde dicha primera secuencia nucleotídica es una secuencia nucleotídica diana y dichas primera y segunda secuencias nucleotídicas se pueden obtener por un proceso que comprende las etapas de: (a) entrecruzamiento de ADN genómico (por ejemplo, *in vivo* o *in vitro*); (b) digestión del ADN entrecruzado con una enzima de restricción primaria; (c) ligamiento de las secuencias nucleotídicas entrecruzadas; (d) reversión del entrecruzamiento; y (e) digestión de las secuencias nucleotídicas con una enzima de restricción secundaria para circularizar las secuencias nucleotídicas.

Preferiblemente, la segunda secuencia nucleotídica interseca (por ejemplo biseca) la primera secuencia nucleotídica (diana). Por consiguiente, la secuencia nucleotídica comprende la segunda secuencia nucleotídica, que separa la primara secuencia nucleotídica (diana) en dos partes o fragmentos - tal como aproximadamente dos partes o fragmentos de igual tamaño. Típicamente, las partes o fragmentos serán de al menos aproximadamente 16 nucleótidos de longitud.

Primera secuencia nucleotídica

La primera secuencia nucleotídica es una secuencia nucleotídica diana.

Como se usa en este documento, la expresión "secuencia nucleotídica diana" se refiere a la secuencia que se usa como una secuencia señuelo para identificar la una o más secuencias con la que entrecruza (por ejemplo, una o más secuencias nucleotídicas de interés y una o más secuencias de composición de secuencia nucleotídica desconocida).

La secuencia nucleotídica diana es de secuencia conocida.
El entrecruzamiento es indicativo de que la secuencia nucleotídica diana y la secuencia entrecruzada con la misma estaban originalmente cerca en el espacio nuclear. Determinando la frecuencia por la cual las secuencias están cerca entre sí, es posible entender, por ejemplo, la conformación de cromosomas y regiones cromosómicas en el contexto espacial del núcleo (por ejemplo, *in vivo* o *in vitro*). Además, es posible entender las intrincadas organizaciones estructurales dentro del genoma, por ejemplo, cuando potenciadores y otros elementos reguladores de la transcripción comunican con promotores distantes localizados en *cis* o incluso en *trans*. Además, incluso es posible entender el posicionamiento de una región genómica dada respecto a secuencias nucleotídicas presentes en el mismo cromosoma (en *cis*) así como a secuencias nucleotídicas en otros cromosomas (en *trans*). Por tanto, es posible mapear secuencias nucleotídicas en diferentes cromosomas que comparten frecuentemente sitios en el espacio nuclear. Además, incluso es posible detectar aberraciones genéticas equilibradas y/o desequilibradas - tales como translocaciones equilibradas y/o desequilibradas, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos (por ejemplo, deleciones o translocaciones en uno o más cromosomas). A este respecto, las aberraciones genéticas provocan cambios en las interacciones ADN-ADN en la posición en que ha sucedido el cambio, que pueden detectarse.

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

En una realización, la primera secuencia nucleotídica (diana) será mayor de aproximadamente 350 pb de longitud ya que se elige una enzima de restricción secundaria que corta la primera secuencia nucleotídica (diana) a aproximadamente 350 pb o más desde el sitio de restricción primario. Esto puede minimizar una predisposición en la formación de círculos debido a las restricciones topológicas (Rippe et al. (2001) Trends in Biochem. Sciences 26, 733-40).

Adecuadamente, la primera secuencia nucleotídica (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 usados para amplificar la segunda secuencia nucleotídica son de aproximadamente 16 bases cada uno.

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

Como se ha mencionado anteriormente, la primera secuencia nucleotídica (diana) puede comprender completa o parcialmente (por ejemplo, un fragmento), o estar cerca de (por ejemplo, en la proximidad de) una secuencia nucleotídica en 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 (secuencia nucleotídica diana) puede por lo tanto ser una secuencia nucleotídica (por ejemplo, un gen o un locus), adyacente a (en el molde de ADN físico) o en la región genómica en que cambios se han asociado con o correlacionado a una enfermedad - tal como una enfermedad genética o congénita. En otras palabras, la primera secuencia nucleotídica (diana) puede ser o puede elegirse en base a 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 los mismos. Ejemplos no limitantes de dichos genes/loci son *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, loci de inmunoglobulina, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCRα/δ*, *TCRβ*, *HOX* y otros loci en diversas 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.lmdatabases.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.n1:8080/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 modo que no existen nucleótidos intermedios entre dos secuencias adyacentes.

En otra realización, el término "adyacente" en el contexto de la secuencia de ácido nucleico y el sitio de reconocimiento por enzima de restricción primaria significa "directamente adyacente" de modo que no existen nucleótidos intermedios entre la secuencia de ácido nucleico y el sitio de reconocimiento por enzima de restricción primaria.

Segunda secuencia nucleotídica

La segunda secuencia nucleotídica se puede obtener, se obtiene, se identifica, o se puede identificar por entrecruzamiento de ADN genómico (por ejemplo, *in vivo* o *in vitro*).

La segunda secuencia nucleotídica (por ejemplo, secuencia nucleotídica de interés) llega a ligarse a la primera secuencia nucleotídica (diana) después de tratar una muestra con un agente de entrecruzamiento y digerir/ligar los fragmentos de ADN entrecruzado. Dichas secuencias se entrecruzan con la primera secuencia nucleotídica (diana) porque estaban originalmente cerca en el espacio nuclear y se ligan con la primera secuencia nucleotídica (diana) porque las condiciones de ligamiento favorecen el ligamiento entre fragmentos de ADN entrecruzados (intramoleculares) sobre eventos de ligamiento aleatorio.

Las enfermedades basadas en alteraciones - tales como translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos - generalmente están causadas por interacciones ADN-ADN aberrantes. La tecnología 4C mide las frecuencias de interacción ADN-ADN, que son principalmente una función de la separación de sitios genómicos, es decir, las frecuencias de interacción ADN-ADN son inversamente proporcionales a la distancia lineal (en kilobases) entre dos loci de ADN presentes en el mismo molde de ADN físico (Dekker et al., 2002). Por tanto, una o más alteraciones que crean nuevos moldes y/o diferentes físicamente de ADN, están acompañadas por interacciones ADN-ADN alteradas y esto puede medirse por tecnología 4C.

Adecuadamente, la segunda secuencia nucleotídica es de al menos 40 pares de bases.

Los agentes de entrecruzamiento - tales como formaldehído - pueden usarse para entrecruzar proteínas con otras proteínas adyacentes y ácido nucleico. Por tanto, dos o más secuencias nucleotídicas pueden entrecruzarse solamente mediante proteínas unidas a (una de) estas secuencias nucleotídicas. También pueden usarse agentes de entrecruzamiento diferentes a formaldehído de acuerdo con la presente invención, incluyendo aquellos agentes de entrecruzamiento que entrecruzan directamente secuencias nucleotídicas. Ejemplos de agentes que entrecruzan ADN incluyen, aunque sin limitación, luz UV, mitomicina C, mostaza de nitrógeno, melfalán, 1,3-butadieno diepóxido, *cis* diaminedicloroplatino(II) y ciclofosfamida.

Adecuadamente, el agente de entrecruzamiento formará entrecruzamientos que unen distancias relativamente cortas - tales como aproximadamente 2 Å - seleccionado de ese modo interacciones íntimas que pueden revertirse.

El entrecruzamiento puede realizarse por, por ejemplo, incubación de las células en formaldehído al 2% a temperatura ambiente - tal como incubación de 1×10^7 células en 10 ml de DMEM-FCS al 10% suplementado con formaldehído al 2% durante 10 min a temperatura ambiente.

Enzima de restricción primaria

Como se usa en este documento, la expresión "enzima de restricción primaria" se refiere a una primera enzima de restricción que se usa para digerir el ADN entrecruzado.

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

La enzima de restricción primaria puede seleccionarse entre enzimas de restricción que reconocen secuencias de al menos 6 pb o más de ADN.

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

Las enzimas de restricción que reconocen una secuencia de más de 6 pb de ADN incluyen, aunque 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.

Para algunos aspectos de la presente invención, en el caso de enzimas de restricción que reconocen secuencias de 6 pb, se prefieren *Bgl*II, *Hind*III o *Eco*RI.

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

Enzima de restricción secundaria

Como se usa en este documento, la expresión "enzima de restricción secundaria" se refiere a una segunda enzima de restricción que se usa después de la digestión con enzima de restricción primaria, ligamiento del ADN entrecruzado, reversión del entrecruzamiento y (opcional) purificación del ADN. En una realización, la enzima de restricción secundaria se usa para proporcionar extremos definidos de ADN en la secuencias nucleotídicas de interés, que permite el ligamiento de secuencias de composición nucleotídica conocida con los sitios de reconocimiento por enzima de restricción secundaria que flanquean las secuencias nucleotídicas de interés.

En una realización, el ligamiento de secuencias de composición nucleotídica conocida con los sitios de reconocimiento por enzima de restricción secundaria que flanquean (por ejemplo, están a cada lado o extremo de) las secuencias nucleotídicas de interés implica el ligamiento en condiciones diluidas para favorecer el ligamiento intramolecular entre los sitios de reconocimiento por enzima de restricción secundaria que flanquean secuencias nucleotídicas diana y las secuencias nucleotídicas ligadas de interés. Esto provoca de forma eficaz la formación de círculos de ADN en que secuencias nucleotídicas diana conocidas flanquean secuencias desconocidas de interés.

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

En una realización, la enzima de restricción secundaria se elige de modo que no haya sitios para enzima de restricción secundaria en aproximadamente 350 pb (por ejemplo, 350-400 pb) desde el sitio de restricción primario.

En otra realización, la enzima de restricción secundaria se elige de modo que el mismo sitio para enzima de restricción secundaria esté probablemente localizado en la secuencia nucleotídica ligada (es decir, la secuencia entrecruza ligada). Como los extremos de la primera secuencia nucleotídica (diana) y la secuencia nucleotídica ligada pueden ser extremos cohesivos (o romos) compatibles, las secuencias pueden incluso ligarse para circularizar el ADN. Por consiguiente, la etapa de digestión está seguida por ligamiento en condiciones diluidas que favorecen interacciones intramoleculares y circularización opcional del ADN mediante los extremos compatibles.

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

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

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

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 secuencias de ADN de composición de secuencia nucleotídica conocida con el sitio de digestión por enzima de restricción secundaria de la una o más secuencias que están ligadas a la secuencia nucleotídica diana.

Enzima de restricción terciaria

Como se usa en este documento, la expresión "enzima de restricción terciaria" se refiere a una tercera enzima de restricción que puede usarse opcionalmente después de la etapa con enzima de restricción secundaria para linealizar el ADN circularizado antes de la amplificación.

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

Preferiblemente, la enzima de restricción terciaria digiere la primera secuencia nucleotídica (diana) entre los sitios de reconocimiento por 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 nucleotídica (diana) demasiado cerca de los sitios de reconocimiento por enzima de restricción primaria y secundaria de modo que los cebadores de amplificación ya no pueden hibridar. Por consiguiente, se prefiere que el sitio de reconocimiento por enzima de restricción terciaria esté localizado al menos a la misma distancia desde los sitios de reconocimiento por enzima de restricción primaria y secundaria que la longitud del cebador a usar de modo que el cebador o cebadores de amplificación aún puedan hibridar.

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

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

Sitio de reconocimiento

Las endonucleasas de restricción son enzimas que escinden la estructura azúcar-fosfato del ADN. En ámbitos más prácticos, una enzima de restricción dada corta ambas hebras de ADN dúplex con un tramo de justo unas pocas bases. Los sustratos para enzimas de restricción son secuencias de ADN bicatenario llamadas sitios/secuencias de reconocimiento.

La longitud de los sitios de reconocimiento de restricción varía, dependiendo de la enzima de restricción que se usa. La longitud de la secuencia de reconocimiento dictamina cómo de frecuentemente la enzima cortará una secuencia de ADN.

A modo de ejemplo, varias enzimas de restricción reconocen una secuencia de 4 pb de ADN. Las secuencias y la enzima que reconoce la secuencia de 4 pb de ADN incluyen, aunque sin limitación, AATT (TspEI), ACGT (MaeII), AGCT (AluI), CATG (NlaIII), CCGG (HpaII), CGCG (FnuDII), CTAG (MaeI), GATC (DpnI, DpnII, Sau3AI y MboI), GCGC (HhaI), GGCC (HaeIII), GTAC (RsaI), TCGA (TaqI), TGCA (CviRI), TTAA (MseI), CCCG (Sth132I), CCGC (AciI) y CCTC (MnII).

A modo de ejemplo adicional, varias enzimas de restricción reconocen una secuencia de 6 pb de ADN. Las secuencias y la enzima que reconoce la secuencia de 6 pb de ADN incluyen, aunque 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 (SpII), CTCGAG (XhoI), CTGCAG (PstI), CTTAAG (AflII), GAATTC (EcoRI), GACGTC (AatII), GAGCTC (SacI), GATATC (EcoRV), GCATGC (SphI), GCCGGC (NaeI), GCGCGC (BsePI), GCTAGC (NheI), GGATCC (BamHI), GGCGCC (NarI), GGGCCC (ApaI), GGTACC (KpnI), GTATAC (SnaI), GTCGAC (SalI), GTGCAC (ApaLI), GTTAAC (HpaI), TACGTA (SnaBI), TCATGA (BspHI), TCCGGA (BspMII), TCGCGA (NruI), TCTAGA (XbaI), TGATCA (BclI), TGCGCA (MstI), TGGCCA (BalI), TGTACA (Bsp1407I), TTATAA (PsiI), TTCGAA (AsuII) y TTTAAA (AhaIII).

A modo de ejemplo adicional, varias enzimas de restricción reconocen una secuencia de 7 pb de ADN. Las secuencias y la enzima que reconoce la secuencia de 7 pb de ADN incluyen, aunque sin limitación, CCTNAGG (SauI), GCTNAGC (EspI), GGTNACC BstEII y TCCNGGA PfoI.

A modo de ejemplo adicional, varias enzimas de restricción reconocen una secuencia de 8 pb de ADN. Las secuencias y la enzima que reconoce la secuencia de 8 pb de ADN incluyen, aunque sin limitación, ATTTAAAT (SwaI), CCTGCAGG (Sse8387I), CGCCGGCG (Sse232I), CGTCGACG (SgrDI), GCCCGGGC (SrfI), GCGATCGC (SgfI), GCGGCCGC (NotI), GGCCGGCC (FseI), GGCGCGCC (AscI), GTTTAAAC (PmeI) y TTAATTAA (PacI).

Estas varias enzimas contienen la secuencia CG que puede metilarse *in vivo*. Varias enzimas de restricción son sensibles a esta metilación y no escindirán la secuencia metilada, por ejemplo, HpaII no escindirá la secuencia $CC^{m}GG$ mientras que su isoesquizómero MspI es insensible a esta modificación y escindirá la secuencia metilada. Por consiguiente, en algunos casos, no se usan las enzimas sensibles a metilación eucariota.

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

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

Preferiblemente, se elige una enzima de restricción secundaria que corta la primera secuencia nucleotídica (diana) a más de aproximadamente 350 pb (por ejemplo, 350-400 pb) desde el sitio de restricción primario. De forma ventajosa, esto minimiza la predisposición en la formación de círculos debido a restricciones topológicas (Rippe et al. (2001) Trends in Biochem. Sciences 26, 733-40).

Preferiblemente, la enzima de restricción secundaria es una cortadora frecuente que reconoce un sitio de reconocimiento por enzima de restricción de 4 o 5 pb. Por tanto, es posible obtener los fragmentos de restricción más pequeños para eficacias iguales de amplificación 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 por enzima secundaria en la primera secuencia nucleotídica (diana) localizado a más de aproximadamente 350-400 pb desde el sitio de restricción primario y otro sitio de reconocimiento por enzima secundaria localizado en la secuencia nucleotídica que se ha ligado (es decir, en la segunda secuencia nucleotídica).

Preferiblemente, la etapa de digestión con enzima de restricción secundaria se realiza durante más de 1 hora hasta durante una noche y seguido por inactivación por calor de la enzima.

Preferiblemente, el ADN en esta mezcla de reacción se purifica usando métodos/kits convencionales que son conocidos en la técnica.

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

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

Preferiblemente, la reacción de ligamiento se realiza a una concentración de ADN de aproximadamente 1-5 ng/µl.

Preferiblemente, la reacción de ligamiento se realiza durante más de 1 hora (por ejemplo, 2, 3, 4 o más horas) a aproximadamente 16-25ºC.

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 nucleotídica (diana), el sitio de reconocimiento por enzima de restricción primaria y los sitios de reconocimiento por enzima de restricción secundaria definirán los extremos de la primera secuencia nucleotídica (diana) y la secuencia nucleotídica ligada (es decir, la segunda secuencia nucleotídica). Por consiguiente, la primera secuencia nucleotídica (diana) y la secuencia nucleotídica ligada están separadas (por ejemplo, divididas) por el sitio de reconocimiento por enzima de restricción primaria y el sitio de reconocimiento por enzima de restricción secundaria.

Amplificación

Puede realizarse una o más reacciones de amplificación para amplificar los moldes de ADN 4C.

La amplificación de ADN puede realizarse usando varios métodos diferentes que son conocidos en la técnica. Por ejemplo, el ADN puede amplificarse usando la reacción en cadena de la polimerasa (Saiki et al., 1988); la PCR mediada por ligamiento, la amplificación por 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 auto-sostenida (Fahy, Kwoh y Gingeras, 1991) y amplificación por desplazamiento de hebra (Walker et al., 1992).

Preferiblemente, el ADN se amplifica usando PCR. "PCR" se refiere al método de K. B. Mullis patentes de Estados Unidos Nº 4.683.195, 4.683.202, y 4.965.188 que describen un método para aumentar la concentración de un segmento de una secuencia nucleotídica en una mezcla de ADN genómico sin clonación o purificación.

En una realización, se usa 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 usa la reacción en cadena de la polimerasa (PCR), pero tiene los cebadores orientados en la dirección inversa de la orientación usual. El molde para los cebadores inversos es un fragmento de restricción que se ha ligado sobre sí 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 usando una enzima de restricción terciaria. Preferiblemente, se usa una enzima de restricción terciaria que es una cortadora de 6 pb o más. Preferiblemente, la enzima de restricción terciaria corta la primera secuencia nucleotídica (diana) entre los sitios de enzimas de restricción primaria y secundaria.

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

Para la etapa de amplificación, se usan al menos dos cebadores oligonucleotídicos en que cada cebador hibrida con una secuencia de ADN que flanquea las secuencias nucleotídicas de interés.

En una realización preferida, se usan al menos dos cebadores oligonucleotídicos en que cada cebador hibrida con la secuencia diana que flanquea las secuencias nucleotídicas de interés.

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

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

Como se usa en este documento, el término "cebador" se refiere a un oligonucleótido, sea de origen natural como en una digestión de restricción purificada o producido sintéticamente, que es capaz de actuar como punto de inicio de la síntesis cuando se pone en condiciones en que se induce la síntesis de un producto de extensión de cebador que es complementario a una hebra de ácido nucleico (es decir, en presencia de nucleótidos y un agente de inducción tal como una ADN polimerasa y a una temperatura y pH adecuados). El cebador es preferiblemente monocatenario para una eficacia máxima en la amplificación, pero puede ser bicatenario. Si es bicatenario, el cebador se trata primero para separar sus hebras antes de usarse para preparar productos de extensión. Preferiblemente, el cebador es un oligodesoxirribonucleótido. El cebador debe ser suficientemente largo para cebar la síntesis de productos de extensión en presencia del agente inductor. Las longitudes exactas de los cebadores dependerán de muchos factores, incluyendo la temperatura, fuente de cebador y el uso del método.

Adecuadamente, los cebadores serán de al menos 15, preferiblemente al menos 20, por ejemplo al menos 25, 30 o 40 nucleótidos de longitud. Preferiblemente, los cebadores de amplificación son de 16 a 30 nucleótidos de longitud.

Preferiblemente, los cebadores se diseñan para estar lo más cerca posible a los sitios de reconocimiento por enzima de restricción primaria y secundaria que separan la primera secuencia nucleotídica (diana) y la segunda secuencia nucleotídica. Los cebadores pueden diseñarse de modo que estén en 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 nucleótido(s) desde los sitios de reconocimiento por enzima de restricción primaria y secundaria.

Adecuadamente, los cebadores de amplificación se diseñan de modo que sus extremos 3' estén enfocados hacia el exterior hacia los sitios de reconocimiento por enzima de restricción primaria y secundaria de modo que la extensión proceda inmediatamente a través de los sitios de restricción en la segunda secuencia nucleotídica.

Si el método de amplificación que se usa es PCR inversa, entonces se prefiere que las reacciones de amplificación se realicen en aproximadamente 100-400 ng de ADN de molde 4C (por aproximadamente 50 µl de mezcla de reacción de PCR) u otras cantidades de ADN para las que reacciones replicadas de PCR dan resultados reproducibles (véase la Figura 1) e incluyen una cantidad máxima de eventos de ligamiento por reacción de PCR.

Preferiblemente, la reacción de amplificación por PCR inversa se realiza usando el sistema de PCR de molde largo Expand (Roche), usando Tampón 1 de acuerdo con las instrucciones del fabricante.

Muestra

El término "muestra", como se usa en este documento, tiene su significado natural. La muestra puede ser cualquier entidad física que comprende ADN que es o es capaz de entrecruzarse. La muestra puede ser o puede derivar de material biológico.

La muestra puede ser o puede derivar de una de más entidades - tal como una o más células, uno o más núcleos, o una o más muestras tisulares. Las entidades pueden ser o pueden derivarse de cualquier entidad en que esté presente ADN - tal como cromatina. La muestra puede ser o puede derivar de una o más células aisladas o una o más muestras tisulares de tejido aisladas, o uno o más núcleos aislados.

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

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

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

La muestra puede ser o puede derivar de un sujeto que tiene que ensayarse para la probabilidad de que padezca una enfermedad en el futuro.

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

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

Marcador

Preferiblemente, las secuencias nucleotídicas (por ejemplo, moldes de ADN 4C amplificados, cebadores o sondas, etc.) se marcan para ayudar a sus aplicaciones corriente abajo - tal como hibridación de series. A modo de ejemplo, los moldes de ADN 4C pueden marcarse usando cebado aleatorio o traslado de mella.

Puede usarse una amplia diversidad de marcadores (por ejemplo, indicadores) para marcar las secuencias nucleotídicas descritas en este documento, particularmente durante la etapa de amplificación. Los marcadores adecuados incluyen radionúclidos, enzimas, agentes fluorescentes, quimio luminiscentes, o cromogénicos así como sustratos cofactores, inhibidores, partículas magnéticas y similares. Las patentes que muestran 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.

Marcadores adicionales incluyen, aunque sin limitación β-galactosidasa, invertasa, proteína fluorescente verde, luciferasa, cloranfenicol, acetiltransferasa, β-glucuronidasa, exoglucanasa y glucoamilasa. También pueden usarse marcadores fluorescentes, así como reactivos fluorescentes sintetizados específicamente con propiedades químicas particulares. Está disponible una amplia diversidad de modos para medir la fluorescencia. Por ejemplo, algunos marcadores fluorescentes muestran un cambio en los espectros de excitación o emisión, algunos muestran transferencia de energía de resonancia donde un indicador fluorescente pierde fluorescencia, mientras que un segundo gana en fluorescencia, algunos muestran una pérdida (inactivación) o aparición de fluorescencia, mientras que algunos presentan movimientos de rotación.

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

Serie

Como se describe en este documento los moldes de ADN 4C preparados de acuerdo con los métodos descritos en este documento pueden hibridarse con una serie. Por consiguiente, puede usarse la tecnología de series (por ejemplo, microseries) para identificar secuencias nucleotídicas - tales como fragmentos genómicos - que comparten frecuentemente un sitio nuclear con una primera secuencia nucleotídica (diana).

Pueden usarse series existentes - tales como series de expresión y genómicas - como se describe en este documento.

Por motivos de referencia, una "serie" es una colección creada de forma intencionada de ácidos nucleicos que pueden prepararse de forma sintética o bio sintética y explorarse para la actividad biológica en una diversidad de formatos diferentes (por ejemplo, bibliotecas de moléculas solubles; y bibliotecas de oligos aferrados a perlas de resina, de sílices u otros soportes sólidos). Adicionalmente, el término "serie" incluye aquellas bibliotecas de ácidos nucleicos que pueden prepararse por aplicación puntual de ácidos nucleicos de esencialmente cualquier longitud (por ejemplo, de 1 a aproximadamente 1000 monómeros nucleotídicos de longitud) en un sustrato.

La tecnología de series y las diversas técnicas y aplicaciones asociadas con la misma se describe en líneas generales 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, Gnomes y Chips. En DNA Microarrays: A Practical Approach (ed. M. Schena), Oxford University Press, Oxford, RU, 1999), The Chipping Forecast (Nature Genetics número especial; Suplemento de enero de 1999), Mark Schena (Ed.), Microarray Biochip Technology, (Eaton Publishing Company), Cortes, 2000, The Scientist 14[17]:25, Gwynne y Page, Microarray analysis: the next revolution en molecular biology, Science, 6 de agosto de 1999; y Eakins y Chu, 1999, Trends en Biotechnology, 17, 217-218.

La tecnología de series supera las desventajas con los métodos tradicionales en biología molecular, que generalmente funcionan sobre una base de "un gen en un experimento", que produce bajo rendimiento la incapacidad de apreciar la "escena completa" de la función génica. Actualmente, las aplicaciones principales para la tecnología de series incluyen la identificación de secuencia (gen/mutación génica) y la determinación del nivel de expresión (abundancia) de genes. El perfilado de la expresión génica puede hacer uso de la tecnología de series, opcionalmente en combinación con técnicas de proteómica (Celis et al, 2000, FEBS Lett, 480(1):2-16; Lockhart y Winzeler, 2000, Nature 405(6788):827-836; Khan et al., 1999, 20(2):223-9). Otras aplicaciones de la tecnología de series también son conocidas en la técnica; por ejemplo, el descubrimiento de genes, la 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), el análisis de SNP (Wang et al, 1998, Science, 280(5366):1077-82), el descubrimiento de fármacos, la farmacogenómica, diagnóstico de enfermedades (por ejemplo, utilizando dispositivos microfluidics: Chemical & Engineering News, 22 de febrero de 1999, 77(8):27-36), toxicología (Rockett y Dix (2000), Xenobiotica, 30(2): 155-77; Afshari et al., 1999, Cancer Res1;59(19):4759-60) y toxicogenómica (un híbrido de genómica funcional y toxicología molecular).

En general, cualquier biblioteca puede disponerse de cualquier modo ordenadamente en una serie, separando espacialmente los miembros de la biblioteca. Ejemplos de bibliotecas adecuadas para la formación de series incluyen bibliotecas de ácido nucleico (incluyendo bibliotecas de ADN, ADNc oligonucleótidos, etc), bibliotecas de péptidos, polipéptidos y proteínas, tales como bibliotecas que comprenden cualquier molécula, tal como bibliotecas de ligando, entre otras.

Las muestras (por ejemplo, miembros de una biblioteca) generalmente se fijan o inmovilizan en una fase sólida, preferiblemente un sustrato sólido, para limitar la difusión y mezcla de las muestras. Preferiblemente, pueden preparase bibliotecas de ADN uniendo ligandos. En particular, las bibliotecas pueden inmovilizarse en una fase sólida sustancialmente plana, incluyendo membranas y sustratos no porosos tales como plástico y vidrio. Además, las muestras se disponen preferiblemente de tal modo que se facilite el indexado (es decir, referencia o acceso a una muestra particular). Típicamente, las muestras se aplican como manchas en una formación de rejilla. Los sistemas de ensayo comunes pueden adaptarse para este propósito. Por ejemplo, una serie puede inmovilizarse en la superficie de una microplaca, con múltiples muestras en un pocillo, o con una única muestra en cada pocillo. Además, el sustrato sólido puede ser una membrana, tal como una membrana de nitrocelulosa o nylon (por ejemplo, membranas usadas en experimentos de transferencia). Sustratos alternativos incluyen vidrio, o sustratos basados en sílice. Por tanto, las muestras se inmovilizan mediante cualquier método conocido en la técnica, por ejemplo, por interacciones de carga, o por acumulamiento químico a las paredes o fondo de los pocillos, o la superficie de la membrana. Pueden usarse otros medios de disposición y fijación, por ejemplo, pipeteo, contacto por caída, medios piezoeléctricos, tecnología de inyección de tinta y chorro de tinta, aplicación electrostática, etc. En el caso de chips basados en sílice, puede utilizarse fotolitografía para disponer y fijar las muestras sobre el chip.

Las muestras pueden disponerse por aplicación "puntual" sobre el sustrato sólido; esto puede hacerse a mano o haciendo uso de robótica para depositar la muestra. En general, las series pueden describirse como macro series o microseries, siendo la diferencia el tamaño de las manchas de muestra. Las macro series típicamente contienen tamaños de mancha de muestra de aproximadamente 300 micrómetros o mayores y pueden extraerse imágenes fácilmente mediante escáneres existentes de gel y transferencia. Los tamaños de mancha de muestra en microseries son típicamente de menos de 200 micrómetros de diámetro y estas series habitualmente contienen miles de manchas. Por tanto las microseries pueden requerir robótica y equipo de imágenes especializados, que puede que haya que fabricarlos a medida. La instrumentación se describe en líneas generales en una revisión de Cortese, 2000, The Scientist 14[11]:26.

Las técnicas para producir bibliotecas inmovilizadas de moléculas de ADN se han descrito en la técnica. Generalmente, los métodos de la técnica más anterior describían el modo de sintetizar bibliotecas de moléculas de ácido nucleico monocatenario, usando por ejemplo técnicas de filtrado para crear diversas permutaciones de

secuencias en las diversas posiciones concretas en el sustrato sólido. La patente de Estados Unidos Nº 5.837.832 describe un método mejorado para producir series de ADN inmovilizadas en sustratos de silicio basados en tecnología de integración a escala muy grande. En particular la patente de Estados Unidos Nº 5.837.832 describe una estrategia llamada "agrupamiento" para sintetizar conjuntos específicos de sondas en localizaciones espacialmente definidas sobre un sustrato que puede usarse para producir las bibliotecas de ADN inmovilizadas descritas en este documento. La patente de Estados Unidos Nº 5.837.832 también proporciona referencias para técnicas previas que también pueden usarse.
Las series descritas en este documento también pueden construirse usando química de foto deposición.

Las series de péptidos (o peptidomiméticos) también pueden sintetizarse sobre una superficie de un modo que coloca cada distinto miembro de la biblioteca (por ejemplo, secuencia peptídica única) en una localización predefinida concreta en la serie. La identidad de cada miembro de la biblioteca se determina por su localización espacial en la serie. Se determinan las localizaciones en la serie donde suceden interacciones de unión entre una molécula predeterminada (por ejemplo, una diana o sonda) y miembros reactivos de la biblioteca, identificando de ese modo las secuencias de los miembros reactivos de la biblioteca en base a la localización espacial. Estos métodos se describen en la patente de Estados Unidos Nº 5.143.854; documento WO90/15070 y documento WO92/10092; Fodor et al. (1991) Science, 251: 767; Dower y Fodor (1991) Ant. Rep. Med. Chem., 26: 271.

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

Ejemplos específicos de series de ADN son los siguientes (para referencia):

Formato I: se inmoviliza ADNc de sonda (500-5.000 bases de longitud) en una superficie sólida tal como vidrio usando deposición puntual con robot y se expone a un conjunto de dianas por separado o en una mezcla. Este método se considera ampliamente como desarrollado en Stanford University (Ekins y Chu, 1999, Trends in Biotechnology, 1999, 17, 217-218).

Formato II: una serie de oligonucleótidos (oligos de 20-25 monómeros, preferiblemente, oligos de 40-60 monómeros) o sondas de ácido peptidonucleico (PNA) se sintetizan in situ (sobre chip) o por síntesis convencional seguida por inmovilización sobre chip. La serie se expone a ADN de muestra marcado, se hibrida, y se determina la entidad/abundancia de secuencias complementarias. Dicho chip de ADN se vende por Affymetrix, Inc., con la marca GeneChip®. Agilent y Nimblegen también proporcionan series adecuadas (por ejemplo, series de agrupamiento genómico).

Ejemplos de algunos formatos de microserie disponibles en el mercado se exponen en la siguiente Tabla 1 (véase también Marshall y Hodgson, 1998, Nature Biotechnology, 16(1), 27-31).

Tabla 1: Ejemplos de formatos de microserie de hibridación disponibles

Empresa	Nombre del producto	Método de formación de serie	Etapa de hibridación	Lectura
Affymetrix.		*In situ* (sobre chip)	10.000-260.000 oligo	Fluorescencia
Inc., Santa Clara, California	GeneChip®	Síntesis fotolitográfica de oligos de ~20-25 monómeros en obleas de silicio, que se pican en virutas de 1,25 cm2 o 5,25 cm2	Presenta sondas con fragmentos marcados de 30-40 nucleótidos de ADNc de muestra o ARN antisentido	
Brax, Cambridge, RU		Oligo sintético corto, sintetizado fuera de chip	1000 oligos en un "chip universal" sondeado con ácido nucleico marcado	Espectrometría de masas
Gene Logic, Inc., Columbia, Maryland	READS™			
Genometrix Inc., The Woodlands, Texas	Universal Arrays™			

Empresa	Nombre del producto	Método de formación de serie	Etapa de hibridación	Lectura
GENSET, París, Francia				
Hyseq Inc., Sunnyvale, California	HyChip™	Muestras de ADN de 500-2000 nt impresas en membranas de 0,6 cm2 (HyGnostics) o ~18 cm2 (Gene Discovery)	64 machas de ADNc de muestra sondeadas con 8.000 oligos de 7 monómeros (HyGnostics) o <=55,000 manchas de ADNc de muestra sondeadas con 300 oligo de 7 monómeros (Gene Discovery)	Radioisótopo
		Oligos de 5 monómeros fabricados impresos como series de 1,15 cm2 en vidrio (HyChip)	1024 manchas de oligo universal sondeadas con ADNc de muestra de 10 kb, oligo de 5 monómeros marcado, y ligasa	Fluorescencia
Incyte Pharmaceuticals Inc., Palo Alto, California	GEM	Impresión piezoeléctrica para aplicar puntualmente fragmentos de PCR y síntesis sobre chip de oligos	<=1000 (finalmente 10.000) oligo/manchas de fragmento de PCR sondeadas con ARN marcado	Fluorescencia y radioisótopo
Molecular Dynamics, Inc., Sunnyvale, California	Storm® FluorImager ®	ADNc de 500-5000 nt impresos por estilográfica en ~10 cm2 en portaobjetos de vidrio	~10.000 manchas de ADNc sondeadas con ADNc de muestra marcados de 200-400 nt	Fluorescencia
Nanogen, San Diego, California	Semiconductor Microchip	Oligos de ~20 monómeros prefabricados, capturados en manchas electroactivas en obleas de silicio, que se pican en virutas de <=1 cm2	25, 64, 400 (y finalmente 10.000) manchas de oligo polarizadas para potenciar la hibridación con ADNc de muestra marcados de 200-400 nt	Fluorescencia
Protogene Laboratories. Palo Alto, California		Síntesis sobre chip de oligos de 40-50 monómeros en viruta de vidrio de 9 cm2 mediante impresión en una serie de tensión superficial	<=8.000 manchas de oligo sondeadas con ácidos nucleicos de muestra marcados de 200-400 nt	Fluorescencia
Sequenom. Hamburgo, Alemania, y San Diego, California	MassArray SpectroChip	Impresión off-set de serie; oligos de aproximadamente 20-25 monómeros	250 localizaciones por SpectroChip interrogado por desorción laser y espectrometría de masas	Espectrometría de masas
Synteni. Inc., Fremont, California	UniGEM™	ADNc de 500-5.000 nt impresos por cabezal en chip de vidrio de ~4 cm^2	<=10.000 manchas de ADNc sondeadas con ADNc de muestra marcadas de 200-400 nt	Fluorescencia

Empresa	Nombre del producto	Método de formación de serie	Etapa de hibridación	Lectura
Nimblegen Systems Inc., Madison	*Homo sapiens* Whole-Genome 60mer Microarray	38.000 transcritos con 5 sondas por gen 17,4 mm x 13 mm		Plataforma de exploración de 5 micrómetros
The German Cancer Institute, Heidelberg, Alemania		Macrochip de PNA prototipo con síntesis sobre chip de sondas usando química f-moc o t-moc	Aproximadamente 1.000 manchas en un chip de 8 x 12 cm	Fluorescencia/espectrometría de masas

Para generar datos a partir de ensayos basados en serie, se detecta una señal que identifica la presencia o ausencia de hibridación entre una sonda y una secuencia nucleotídica. También se describen en este documento técnicas de marcaje directo e indirecto. Por ejemplo, el marcaje directo incorpora colorantes fluorescentes directamente en la secuencias nucleotídicas que hibridan con las sondas asociadas en la serie (por ejemplo, se incorporan colorantes en la secuencia nucleotídica por síntesis enzimática en presencia de nucleótidos o cebados de PCR marcados). Los esquemas de marcaje directo producen fuertes señales de hibridación, típicamente usando familias de colorantes fluorescentes con estructuras químicas y características similares, y son simples de implementar. En realizaciones preferidas que comprenden marcaje directo de ácidos nucleicos, se utilizan cianina o análogos Alexa en análisis comparativos de series de múltiple-flúor. En otras realizaciones, pueden utilizarse esquemas de marcaje indirecto para incorporar epítopos en los ácidos nucleicos antes de o después de la hibridación con las sondas de microserie. 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, proporcionando de ese modo una señal fluorescente en virtud de la conjugación de la molécula decolorante con el epítopo de la especie hibridada).

El análisis de datos también es una parte importante de un experimento que implica series. Los datos sin procesar de un experimento de serie típicamente son imágenes, que tienen que transformarse en matrices-tablas donde las filas representan por ejemplo genes, las columnas representan por ejemplo diversas muestras tales como tejidos o condiciones experimentales, y los números en cada celda por ejemplo caracterizan la expresión de una secuencia particular (preferiblemente, una segunda secuencia que se ha ligado con la primera secuencia nucleotídica (diana)) en la muestra particular. Estas matrices tienen que analizarse adicionalmente, si tiene que extraerse algún conocimiento acerca de los procesos biológicos subyacentes. Los métodos de análisis de datos (incluyendo análisis supervisados y no supervisados de datos así como enfoques bio informáticos) se describen en Brazma y Vilo J (2000) FEBS Lett 480(1):17-24.

Como se escribe en este documento, la una o más secuencias nucleotídicas (por ejemplo, el molde de ADN) que se marcan y posteriormente se hibridan con una serie comprenden una secuencia nucleotídica que está enriquecida para pequeños tramos de secuencias con una firma distinta, es decir, que abarcan la secuencia nucleotídica entre el sitio de reconocimiento por enzima de restricción primaria que se ligó durante el procedimiento 3C con la primera secuencia nucleotídica (diana), y sus respectivos sitios de reconocimiento por enzima de restricción secundaria adyacentes.

Una única serie puede comprender múltiples (por ejemplo, dos o más) secuencias señuelo.

Sondas - para referencia

Como se usa en este documento, el término "sonda" se refiere a una molécula (por ejemplo, un oligonucleótido, sea de origen natural como en una digestión de restricción purificada o producido sintéticamente, de forma recombinante o por amplificación por PCR), que es capaz de hibridar con otra molécula de interés (por ejemplo, otro oligonucleótido). Cuando las sondas son oligonucleótidos pueden ser mono catenarias o bi catenarias. Las sondas son útiles en la detección, identificación y aislamiento de dianas particulares (por ejemplo, secuencias génicas). Como se describe en este documento, se contempla que sondas descritas en este documento pueden marcarse con un marcador de modo que sean detectables en cualquier sistema de detección, incluyendo, aunque sin limitación sistemas enzimáticos (por ejemplo, ELISA, así como ensayos histoquímicos basados en enzimas), fluorescentes, radiactivos, y luminiscentes.

Con respecto a series y microseries, el término "sonda" se usa para hacer referencia a cualquier material hibridable que se fije a la serie con el fin de detectar una secuencia nucleotídica que haya hibridado con dicha sonda. Preferiblemente, estas sondas son de 25-60 monómeros o más largas.

Se describen estrategias para el diseño de sondas en el documento WO95/11995, documento EP 717.113 y documento WO97/29212.

Como 4C permite una búsqueda en el genoma completo sin influencias para interacciones, es ventajoso preparar una serie con sondas que interroguen cualquier posible (por ejemplo, único/no repetitivo) sitio de reconocimiento por enzima de restricción primaria en el genoma. Por tanto, el diseño de series solamente depende de la elección de la enzima de restricción primaria y no de las secuencias nucleotídicas primera o secundaria reales.
Aunque pueden usarse series existentes, se prefiere usar configuraciones descritas en este documento.

En una configuración, se diseñan una o más sondas en la serie de modo que puedan hibridar cerca de los sitios que se digieren por la enzima de restricción primaria. Más preferiblemente, la sonda o sondas están en aproximadamente 20 pb del sitio de reconocimiento por enzima de restricción primaria. Más preferiblemente, la sonda o sondas están en aproximadamente 50 pb del sitio de reconocimiento por enzima de restricción primaria.

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

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

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

En otra configuración, las sondas en la serie se diseñan de modo que puedan hibridar en cada lado de los sitios que se digieren por la enzima de restricción primaria. Adecuadamente, puede usarse una única sonda en cada lado del sitio de reconocimiento por enzima de restricción primaria.

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

En otra configuración más, pueden diseñarse dos o más sondas (por ejemplo, 3, 4, 5, 6, 7 u 8 o más) cerca de cada sitio de reconocimiento por enzima de restricción primaria independientemente del sitio de reconocimiento por enzima de restricción secundaria más cercano. En esta configuración todas las sondas aún deberán estar cerca de los sitios de reconocimiento por enzima de restricción primaria (preferiblemente en 300 pb del sitio de restricción).

De forma ventajosa, el último diseño y también el diseño que usa una sonda por (lado de) sitio de reconocimiento por enzima de restricción primaria, permite el uso de diferentes enzimas de restricción secundarias en combinación con una enzima de restricción primaria dada.

De forma ventajosa, el uso de múltiples (por ejemplo, 3, 4, 5, 6, 7 u 8 o más) sondas por sitio de reconocimiento por enzima de restricción primaria puede minimizar el problema de obtener resultados falsos negativos debido a un mal rendimiento de las sondas individuales. Además, también puede aumentar la fiabilidad de los datos obtenidos con un único experimento de chip y reducir la cantidad de series necesarias para extraer conclusiones estadísticamente válidas.

Las sondas para su uso en la serie pueden ser de más de 40 nucleótidos de longitud y pueden ser isotérmicas.

Preferiblemente, se excluyen sondas que contienen secuencias repetitivas de ADN.

Se espera, que sondas de diagnóstico para los sitios de restricción que flanquean directamente o están cerca de la primera secuencia nucleotídica den señales de hibridación muy fuertes y también puedan excluirse del diseño de sonda.

La serie puede cubrir cualquier genoma incluyendo genomas de mamífero (por ejemplo, ser humano, ratón (por ejemplo, cromosoma 7)), de vertebrado (por ejemplo, pez cebra), o de no vertebrado (por ejemplo, bacteria, levadura, hongo o insecto (por ejemplo, Drosophila)).

También se describen en este documento series que contienen 2-6 sondas alrededor de cada sitio de restricción primario único y lo más cerca posible al sitio de digestión con la enzima de restricción.

Preferiblemente la distancia máxima desde el sitio de digestión con la enzima de restricción es aproximadamente 300 pb.

También se describen en este documento series para enzimas de restricción - tales como *Hind*III, *EcoR*I, *Bgl*II y *Not*I - que cubren los genomas de mamífero o no de mamífero, que se proporcionan. De forma ventajosa, el diseño de las series descritas en este documento esquivan la necesidad de rediseñar series para cada secuencia diana, con la condición de que el análisis se realice en la misma especie.

Conjuntos de sondas - para referencia

Como se usa en este documento, la expresión "conjunto de sondas" se refiere a un juego o colección de sondas que hibridan con cada uno de los sitios de reconocimiento por enzima de restricción primaria para una enzima de restricción primaria en un genoma.

Por consiguiente, se describe en este documento un conjunto de sondas complementarias en secuencia a la secuencia de ácido nucleico adyacente a cada uno de los sitios de reconocimiento por enzima de restricción primaria para una enzima de restricción primaria en ADN genómico.

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 por enzima de restricción primaria en ADN genómico. El conjunto de sondas puede ser complementario en secuencia a un lado (por ejemplo cualquiera) o ambos lados del sitio de reconocimiento por 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 por enzima de restricción primaria en el ADN genómico.

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

También es posible definir una ventana de aproximadamente 100 pb desde el sitio de reconocimiento por enzima de restricción primaria para identificar sondas óptimas cerca de cada sitio de restricción.

El conjunto de sondas descrito en este documento puede ser complementario a la secuencia que está a menos de 300 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico, complementario a la secuencia que está entre 200 y 300 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico y/o complementario a la secuencia que está entre 100 y 200 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico.

El conjunto de sondas descrito en este documento puede ser complementario a la secuencia que está de 0 a 300 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico, complementario a la secuencia que está entre 0 a 200 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico y/o complementario a la secuencia que está entre 0 a 100 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico (por ejemplo, aproximadamente 10, 20, 30, 40, 50, 60, 70, 80 o 90 pb desde cada uno de los sitios de reconocimiento por enzima de restricción primaria en ADN genómico).

Pueden diseñarse incluso dos o más sondas que son capaces de hibridar con la secuencia adyacente a cada sitio de reconocimiento por enzima de restricción primaria en el ADN genómico.

Las sondas descritas en este documento pueden solapar o solapar parcialmente. Si las sondas solapan entonces el solapamiento es preferiblemente, de menos de 10 nucleótidos.

También pueden usarse fragmentos de PCR que representan los primeros 1-300 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 por enzima de restricción primaria.

También pueden usarse fragmentos de PCR como sondas que corresponden exactamente con cada sitio genómico que está flanqueado por el sitio de reconocimiento por enzima de restricción primaria y el primer sitio de reconocimiento por enzima de restricción secundaria adyacente. Por consiguiente, las secuencias de sonda

descritas en este documento pueden corresponder con todo o parte de la secuencia entre cada uno de los sitios de reconocimiento por enzima de restricción primaria y cada uno de los primeros sitios de reconocimiento por enzima de restricción secundaria adyacentes.

Típicamente, las sondas, serie de sondas o conjunto de sondas descritos en este documento se inmovilizarán en un soporte. Los soportes (por ejemplo, soportes sólidos) pueden fabricarse de una diversidad de materiales - tales como vidrio, sílice, plástico, nylon o nitrocelulosa. Los soportes son preferiblemente rígidos y tienen una superficie plana. Los soportes típicamente tienen de aproximadamente 1-10.000.000 regiones abordables espacialmente concretas, o celdas. Los soportes que tienen aproximadamente 10-1.000.000 o aproximadamente 100-100.000 o aproximadamente 1000-100.000 células son habituales. La densidad de celdas es típicamente al menos aproximadamente 1000, 10.000, 100.000 o 1.000.000 celdas en un centímetro cuadrado. En algunos soportes, todas las celdas están ocupadas por mezclas combinadas de sondas o un conjunto de sondas. En otros soportes, algunas celdas están ocupadas por mezclas combinadas de sondas o un conjunto de sondas, y otras celdas están ocupadas, al menos hasta el grado de pureza obtenible por métodos de síntesis, por un único tipo de oligonucleótido.

Preferiblemente, la serie descrita en este documento comprende más de una sonda por sitio de reconocimiento por enzima de restricción primaria, que en el caso de una enzima de restricción de corte en 6 pb existe, por ejemplo, aproximadamente 750.000 veces por genoma humano o de ratón.

Para una enzima de restricción que reconoce una secuencia de reconocimiento de >6 pb, puede usarse una única serie de aproximadamente 2 x 750.000 sondas para cubrir, por ejemplo, el genoma humano o de ratón completo, con 1 sonda en cada lado de cada sitio de restricción.

En un diseño de serie preferido, la cantidad total de moléculas de sonda de una secuencia nucleotídica dada presente en la serie está en gran exceso a los fragmentos homólogos presentes en la muestra 4C a hibridar con dicha serie. Dada la naturaleza de la tecnología 4C, los fragmentos que representan regiones genómicas cercanas a la secuencia nucleotídica analizada en el molde de cromatina lineal estará en gran exceso en la muestra de hibridación 4C (como se describe en la Figura 2). Para obtener información cuantitativa acerca de las eficacias de hibridación de dichos fragmentos abundantes, puede ser necesario reducir la cantidad de muestra a hibridar y/o aumentar la cantidad de moléculas de una sonda de secuencia oligonucleotídica dada en la serie.

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

La serie puede comprender aproximadamente 3 × 750.000 sondas, 4 × 750.000 sondas, 5 × 750.000 sondas, o preferiblemente, 6 × 750.000 sondas. Más preferiblemente, la serie comprende 6 × 750.000 sondas con 2, 3, 4, 5, 6, 7 o 8 o más sondas en cada lado de cada sitio de restricción. Más preferiblemente, la serie comprende 6 × 750.000 sondas con 3 sondas en cada lado de cada sitio de restricción.

Las series de sondas o conjuntos de sondas descritos en este documento pueden sintetizarse de un modo por etapas en un soporte o pueden fijarse 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 implica el uso de luz para dirigir la síntesis de sondas oligonucleotídicas en series de tamaño reducido de alta densidad. Se describen algoritmos para diseñar filtros para reducir la cantidad de ciclos de síntesis en los documentos US 5.571.639 y US. 5.593.839. Las series también pueden sintetizarse de un modo combinado suministrando monómeros a celdas de un soporte mediante trayectorias de flujo mecánicamente restringidas, como se describe en el documento EP 624.059. Las series también pueden sintetizarse depositando puntualmente reactivos en un soporte usando una impresora de inyección de tinta (véase, por ejemplo, el documento EP 728.520).

Como se usa en este documento las expresiones "sustancialmente un conjunto de sondas", "sustancialmente la serie de sondas" significan que el conjunto o la serie de sondas comprende al menos aproximadamente el 50, 60, 70, 80, 90, 95, 96, 97, 98 o 99% del conjunto o serie entera o completa de sondas. Preferiblemente, el conjunto o la serie de sondas es un conjunto entero o completo de sondas (es decir, 100%).

La serie descrita en este documento preferiblemente comprende una sonda única solamente por lado de cada sitio de reconocimiento por enzima de restricción primaria que está presente en un genoma dado. Si esta cantidad de sondas excede la cantidad de sondas que puede contenerse por una única serie, la serie puede contener preferiblemente aún una representación del genoma completo de una especie dada, pero a una resolución inferior, con por ejemplo una de cada 2, 3, 4, 5, 6, 7, 8, 9, 10, 10^2, 10^3 o 10^4 etc. sondas ordenadas en el molde de cromosoma lineal presente en la serie. Dichas series que cubren el genoma humano completo, u otro genoma, a

resolución sub-óptima pueden preferirse sobre series de alta resolución que cubren parte del mismo genoma, por ejemplo en casos donde tienen que encontrarse compañeros de translocación.

Preferiblemente, la representación del genoma completo de una especie dada a resolución inferior se obtiene mediante sondas en la serie que representan cada una un único fragmento de restricción obtenido después de digestión con una enzima de restricción primaria. Preferiblemente, esto se obtiene ignorando cada segunda, tercera, cuarta, quinta, sexta, séptima, octava, novena, décima, vigésima, trigésima, cuadragésima, quincuagésima, sexagésima, septuagésima, octogésima, nonagésima o centésima etc. sonda que hibrida con el mismo fragmento de restricción.

Preferiblemente, la representación del genoma completo de una especia dada a resolución inferior comprende sondas que se distribuyen equitativamente a lo largo de los moldes de cromosoma lineales. Preferiblemente, esto es obtiene ignorando una o más sondas en estas regiones genómicas que muestras la mayor densidad de sonda.

Hibridación

El término "hibridación" como se usa en este documento incluirá "el proceso por el cual una hebra de ácido nucleico se une con una hebra complementaria a través de apareamiento de bases" así como el proceso de amplificación realizado en, por ejemplo, tecnologías de reacción en cadena de la polimerasa (PCR).

Las secuencias nucleotídicas con capacidad de hibridación selectiva generalmente serán al menos un 75%, preferiblemente al menos un 85 o 90% y más preferiblemente al menos un 95% o 98% homólogas a la correspondiente secuencia nucleotídica complementaria sobre una región de al menos 20, preferiblemente 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 solamente con una secuencia nucleotídica particular en condiciones rigurosas (por ejemplo, 65ºC y SSC 0,1x {SSC 1x = NaCl 0,15 M, Na-citrato 0,015 M pH 7,0}). Las condiciones rigurosas son condiciones en las que una sonda hibridará con su secuencia diana, pero no con otras secuencias. Las condiciones rigurosas son dependientes de secuencia y son diferentes en diferentes circunstancias. Secuencias más largas hibridan específicamente a temperaturas más altas. Generalmente, las condiciones rigurosas se seleccionan para que sean aproximadamente 5ºC inferiores que el punto de fusión térmica (Tm) para la secuencia específica a una fuerza iónica y pH definidos. La Tm es la temperatura (a fuerza iónica, pH, y composición de ácido nucleico definidos) a la cual el 50% de las sondas complementarias a una secuencia diana hibridan con la secuencia diana en equilibrio. (Como las secuencias diana están generalmente presentes en exceso, a Tm, el 50% de las sondas está ocupado en equilibrio). Típicamente, las condiciones rigurosas incluyen una concentración salina de al menos aproximadamente 0,01 to 1,0 M de concentración de iones Na (u otras sales) a pH 7,0 a 8,3 y la temperatura es al menos aproximadamente 30ºC para sondas cortas. Las condiciones rigurosas también pueden conseguirse con la adición de agentes desestabilizantes - tales como formamida o sales de tetraalquil amonio.

Como entenderán los expertos en la materia, puede usarse una hibridación de rigurosidad máxima para identificar o detectar secuencias nucleotídicas idénticas mientras que puede usarse hibridación de rigurosidad intermedia (o baja) para identificar o detectar secuencias polinucleotídicas similares o relacionadas.

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

Frecuencia de interacción

La cuantificación de las frecuencias de ligamiento de fragmentos de restricción da una medida de sus frecuencias de entrecruzamiento. Adecuadamente, esto puede conseguirse usando PCR como se usa en tecnología 3C convencional como se describe por Splinter *et al.* (2004) (*supra*). En resumen, la formación de productos de PCR puede medirse explorando las intensidades de señal después de la separación en geles de agarosa teñidos con bromuro de etidio, usando un generador de imágenes Typhoon 9200 (Molecular Dynamics, Sunnyvale, CA). Adecuadamente, se san varios controles para la correcta interpretación de datos como también se describe en Splinter *et al.* (2004) (supra).

Como la tecnología 4C descrita en este documento proporciona el análisis de alto rendimiento de la frecuencia de interacción de dos o más secuencias nucleotídicas en el espacio nuclear, se prefiere que las frecuencias de ligamiento de fragmentos de restricción se cuantifiquen usando las series descritas en este documento.

Para la cuantificación, las señales obtenidas para una muestra 4C pueden normalizarse a señales obtenidas para una muestra de control. La muestra 4C y la muestra o muestras de control se marcarán con marcadores diferentes y

discernibles (por ejemplo colorantes) y se hibridarán simultáneamente a la serie. La muestra o muestras de control típicamente contendrán fragmentos de ADN (es decir todas las potenciales segundas secuencias nucleotídicas que se han ligado a la primera secuencia nucleotídica (diana)) en cantidades equimolares y, para excluir la influencia en la eficacia de hibridación, deben ser similares en tamaño a la segunda o segundas secuencias nucleotídicas. Por tanto, el molde de control típicamente contendrá ADN genómico (del mismo fondo genético que el usado para obtener el molde 4C), digerido tanto con la enzima de restricción primaria como con la secundaria y marcado por el mismo método (por ejemplo, cebado aleatorio) como molde 4C. Dicho molde de control hace posible corregir las diferencias entre sondas en la eficacia de hibridación. La normalización de las señales de la serie 4C con las señales de la serie de control hace posible expresar los resultados en términos de enriquecimiento sobre eventos aleatorios.

El molde 4C marcado incluso puede hibridarse a una serie con o sin una muestra de control marcada de forma diferencial y con o sin uno o más moldes 4C diferentes marcados de forma diferencial. Otros moldes 4C pueden ser ajenos a este molde 4C, por ejemplo pueden obtenerse de tejido diferente y/u obtenerse con un conjunto diferente de cebadores de PCR inversa. Por ejemplo, el primer molde 4C puede ser un material del paciente y el segundo molde 4C puede obtenerse de un sujeto sano o una muestra de control.

Dados los sorprendentes patrones de hibridación que se esperan para reordenamientos genéticos, no siempre será necesario comparar sujetos enfermos con sujetos sanos. Por consiguiente, pueden hibridarse múltiples (por ejemplo, dos o más) moldes 4C, que interrogan cada uno un locus diferentes del mismo paciente o sujeto a una serie (por ejemplo una o más).

Los moldes 4C pueden marcarse de forma diferencial (por ejemplo con hibridación de dos colores o múltiples colores) y/o pueden marcarse de forma idéntica en caso de que dichos loci residan normalmente en diferentes cromosomas o en el mismo cromosoma a una distancia suficientemente lejana para solapamiento mínimo entre señales de interacción ADN-ADN. Como un ejemplo, puede procesarse el material de un sujeto con leucemia de células T para obtener moldes 4C para *TCR*α/δ (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 puede hibridarse a una sonda, que permitirá el análisis simultáneo en múltiples loci para un reordenamiento genómico asociado con la enfermedad.

Para la cuantificación de frecuencias de interacción, también pueden considerarse intensidades absolutas de señal o relaciones sobre la muestra de control. Además, pueden usarse señales de sondas adyacentes en el molde de cromosoma lineal para identificar regiones cromosómicas de interacción. Dicha información posicional se analiza preferiblemente ordenando las sondas en el molde de cromosoma lineal y analizando las intensidades absolutas de señal, o relaciones sobre señales de molde de control, por enfoques de ventana deslizante, usando por ejemplo enfoques de media de ejecución o mediana de ejecución.

Método de ensayo

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.

Como se usa en este documento, el término "modular" se refiere a prevenir, disminuir, suprimir, restaurar, elevar, aumentar o afectar de otro modo a la interacción ADN-ADN.

En algunos casos, puede ser deseable evaluar dos o más agentes juntos para su uso en la modulación de la interacción ADN-ADN. En estos casos, los ensayos pueden modificarse fácilmente añadiendo dicho o dichos agentes adicionales de forma simultánea con, o posteriormente a, el primer agente.

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

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

Los usos médicos de dichos agentes terapéuticos están dentro del alcance de la presente descripción así como los propios programas de desarrollo de fármacos y composiciones farmacéuticas que comprenden dichos agentes. Un programa de desarrollo de fármacos pude implicar, por ejemplo, obtener un agente identificado o identificable por los métodos descritos en este documento, opcionalmente modificarlo (por ejemplo, modificando su estructura y/o proporcionando una nueva composición que comprenda dicho resto) y realizar estudios adicionales (por ejemplo, estudios de toxicidad y/o estudios sobre la actividad, estructura o función). Pueden realizarse ensayos en animales no humanos y finalmente pueden realizarse sobre seres humanos. Dichos ensayos generalmente incluirán la determinación del efecto o efectos de diferentes niveles de dosificación. Los programas de desarrollo de fármacos pueden utilizar ordenadores para analizar restos identificados por exploración (por ejemplo, para predecir la estructura y/o función, para identificar posibles agonistas o antagonistas, para buscar otros restos que puedan tener estructuras o funciones similares, etc.).

Ensayo de diagnóstico

Actualmente, diversos reordenamientos genómicos siguen siendo difíciles de detectar por las técnicas moleculares-citogenéticas disponibles. Aunque la técnica de hibridación genómica comparativa en serie (serie-CGH) es una técnica recién desarrollada para la detección de amplificación cromosómica y/o deleciones con una resolución de 35-300 Kb, esta técnica no es adecuada para detectar translocaciones equilibradas e inversiones cromosómicas. Por otro lado, el cariotipado espectral (SKY) o cariotipado convencional a menudo se realiza sobre material del paciente para la detección de translocaciones cromosómicas así como cambios numéricos, pero la resolución para definir los puntos de interrupción de translocación es baja, habitualmente 10-50 Mb y 5-10 Mb, respectivamente. Por consiguiente, los resultados obtenidos por ambos métodos y especialmente por SKY conducirán a experimentos de validación que llevan mucho tiempo y son trabajosos como hibridación in situ de fluorescencia (FISH) y estrategias de clonación de punto de interrupción molecular.

La tecnología 4C implica un procedimiento que puede detectar cualquier reordenamiento cromosómico en base a frecuencias de interacción cambiadas entre secuencias de ADN físicamente ligadas. La tecnología 4C por lo tanto es útil para la identificación de reordenamientos cromosómicos (recurrentes) para la mayoría de neoplasias/malformaciones congénitas múltiples o retardo mental en seres humanos. Una ventaja importante de la tecnología 4C es que permite el mapeado muy exacto del punto de interrupción en una región de solamente varios miles de pares de bases. Otra ventaja de la tecnología 4C es que no se requiere conocimiento previo sobre la posición exacta del punto de interrupción, ya que los puntos de interrupción serán detectables incluso cuando la secuencia 4C-señuelo está localizada 1-5 Mb desde el punto de interrupción. Esto también tiene la ventaja de que puede usarse la misma secuencia señuelo para la detección de reordenamientos cromosómicos específicos que cubren grandes áreas de punto de interrupción. El mapeado exacto de los reordenamientos genómicos por tecnología 4C facilitará enormemente la identificación de uno o más genes expresados de forma aberrante subyacentes a enfermedades o trastornos genéticos, que contribuirá de forma importante a una mejor comprensión de las correlaciones genotipo-fenotipo, ayudará a la toma de decisiones de tratamiento y añadirá información pronostica importante.

En una realización de la presente invención, para proporcionar una base para el diagnóstico de enfermedad, se establecen valores normales o convencionales a partir de un sujeto. Esto puede conseguirse ensayando muestra tomadas de sujetos normales - tales como animales o seres humanos. La frecuencia de la interacción ADN-ADN puede cuantificarse comparándola con una serie de diluciones de controles positivos. Entonces, pueden compararse valores convencionales obtenidos de muestras normales con valores obtenidos de muestras de sujetos afectados o potencialmente afectados por una enfermedad o un trastorno.

La desviación entre los valores convencionales y del sujeto establece la presencia de la patología.

Dichos ensayos de diagnóstico pueden adaptarse para evaluar la eficacia de un régimen de tratamiento terapéutico particular y puede usarse 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 enfermedad, puede establecerse un perfil normal o convencional para la interacción ADN-ADN. Los valores convencionales obtenidos de muestras normales pueden compararse con valores obtenidos de muestras de sujetos potencialmente afectados por un trastorno o enfermedad. La desviación entre los valores convencionales y del sujeto establece la presencia de la patología. Si se establece enfermedad, puede administrarse un agente terapéutico existente, y puede generarse el perfil o valores de tratamiento. Finalmente, el método puede repetirse en una base regular para evaluar si los valores progresan hacia o retornan al patrón normal o convencional. Pueden usarse perfiles de tratamientos sucesivos para mostrar la eficacia de tratamientos sobre un periodo de varios días o varios meses.

La tecnología 4C detecta de forma precisa al menos 5 Mb de ADN genómico ligado en *cis* a la secuencia nucleotídica que se analiza (véase la Figura 2-3 y 5). De ventajosa, la tecnología 4C puede usarse para detectar cualquier aberración genómica que esté acompañada por un cambio en la separación del sitio genómico entre secuencias reordenadas y una secuencia 4C (señuelo) de elección. Dicho cambio puede ser, por ejemplo, un aumento o disminución en la separación del sitio genómico o puede ser una sub-representación (como en deleciones) o sobre-representación (como en duplicaciones) de secuencias proximales (por ejemplo, hasta o más de 15 Mb) a la secuencia 4C (señuelo). Típicamente, dichas aberraciones o reordenamientos genómicos son una causa de o están asociados con enfermedades - tales como cáncer (por ejemplo, leucemia) y otras enfermedades genéticas o congénitas descritas en este documento.

Las aberraciones genéticas (por ejemplo, aberraciones genómicas o cromosómicas - tales como aberraciones genómicas o cromosómicas equilibradas y/o desequilibradas) incluyen, aunque sin limitación, reordenamientos, translocaciones, inversiones, inserciones, deleciones y otras mutaciones de ácido nucleico (por ejemplo, cromosomas) y también pierde o gana parte de cromosomas 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. De este modo, se retiran genes (o fragmentos de genes) del contexto fisiológico normal de un cromosoma particular y se localizan en un

cromosoma destinatario, adyacente a genes no relacionados o fragmentos de genes (a menudo oncogenes o protooncogenes).

Las neoplasias pueden incluir leucemias agudas, linfomas malignos y tumores sólidos. Ejemplos no limitantes de alteraciones son t(14;18) que sucede frecuentemente en NHL; t(12;21) que se encuentra frecuentemente en ALL-precursor-B infantil; y la presencia de aberraciones 11q23 (gen MLL (leucemia mieloide-linfoide o leucemia de linaje mixto)) en leucemias agudas.
El gen MLL en la región cromosómica IIq23 está implicado en varias translocaciones tanto en ALL como en leucemias mieloides agudas (AML). Hasta le fecha, 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 suceden en ALL, mientras que otros, 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 la región IIq23 suceden muy frecuentemente en leucemias agudas infantiles (aproximadamente un 60-70%), y a un grado mucho menor en leucemias en niños y adultos (cada una aproximadamente un 5%).

Los reordenamientos en neoplasias linfoides a menudo implican genes Ig o TCR. 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 que el gen MYC se acopla a segmentos génicos de cadena pesada de Ig (IGH), kappa de Ig (IGK), o lambda de 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 linfomas foliculares, uno de los tipos principales de NHL. En esta 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 sobre-expresión de la proteína BCL2, que desempeña un papel como factor de supervivencia en el control del crecimiento inhibiendo la muerte celular programada.

El gen BCL2 consta de tres exones, pero estos están dispersos sobre un área grande. De éstos, el último exón codifica una larga región no traducida 3' (3' UTR). Esta 3' UTR es una de las dos regiones en que muchos puntos de interrupción t(14;18) se agrupan y se llama la "región de punto de interrupción principal"; la otra región de punto de interrupción implicada en translocaciones t(14;18), se localiza 20-30 kb cadena abajo del locus BCL2 y se llama la "región de agrupamiento minoritaria". Una tercera área de punto de interrupción de BCL2, la VCR (región de agrupamiento variante), se localiza en el extremo 5' del locus BCL2 y está implicada entre otras cosas en translocaciones variantes, es decir, t(2;18) y t(18;22), en que los segmentos génicos IGK e IGL son los genes compañeros.

Por tanto, a modo de ejemplo, la tecnología 4C puede aplicarse a la exploración de material de paciente para aberraciones genéticas cerca o en loci que se eligieron en base a su asociación frecuente con un fenotipo clínico dado. Ejemplos no limitantes adicionales de dichos loci son *AML1, MLL, MYC, BCL, BCR, ABL1,* loci de inmunoglobulina, *LYL1*, *TAL1, TAL2, LMO2, TCRα/δ, TCRβ, HOX* y otros loci en diversas leucemias linfoblásticas.

De forma ventajosa, si se sospecha una aberración genética, puede aplicarse tecnología 4C como primera y única exploración para verificar y mapear la presencia de la aberración como se explica en este documento.

Detección de reordenamientos genómicos

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

Actualmente, los reordenamientos genómicos - tales como puntos de interrupción de translocación - son muy difíciles de detectar. Por ejemplo, las microseries de hibridación genómica comparativa (CGH) pueden detectar varios tipos de reordenamientos pero no logran detectar translocaciones. Si se sospecha una translocación en un paciente pero se desconocen los compañeros cromosómicos, puede realizarse cariotipado espectral (SKY) para descubrir los compañeros de translocación y obtener una estimación aproximada de las localizaciones de los puntos de interrupción. Sin embargo, la resolución es muy mala (habitualmente no mejor de ~50 Mb) y habitualmente se requiere un mapeado fino adicional (que a la vez requiere mucho tiempo y es costoso). Esto se hace normalmente usando Hibridación In Situ de Fluorescencia (FISH), que de nuevo proporciona resolución limitada. Usando FISH, los puntos de interrupción pueden localizarse a una región +/-50 kb a máxima resolución.

Las frecuencias de interacción ADN-ADN principalmente son una función de la separación del sitio genómico, es decir, las frecuencias de interacción ADN-ADN son inversamente proporcionales a la distancia lineal (en kilobases) entre dos loci de ADN presentes en el mismo molde físico de ADN (Dekker et al., 2002). Por tanto, una translocación, que crea uno o más nuevos moldes físicos de ADN, está acompañada por interacciones alteradas ADN-ADN cerca de los puntos de interrupción, y esto puede medirse por tecnología 4C. Las enfermedades basadas en translocaciones están típicamente causadas por interacciones ADN-ADN aberrantes, ya que la translocación es el resultado de la unión física (interacción) de brazos cromosómicos (ADN) rotos.

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

A modo de ejemplo, puede aplicarse tecnología 4C para explorar material del paciente para translocaciones cerca del loci que se eligieron en base a su asociación frecuente con un fenotipo clínico dado como se describe en este documento.

Si se sospecha una translocación en un paciente pero se desconocen los compañeros cromosómicos, puede realizarse un mapeado inicial usando métodos actualmente disponibles como cariotipado espectral (SKY). Esto puede identificar los compañeros de translocación y proporcionar una estimación muy aproximada de las localizaciones de punto de interrupción (habitualmente no mejor de ~50 Mb de resolución). Después puede aplicarse tecnología 4C, usando frecuencias "señuelo" en esta región localizada por ejemplo cada 2 Mb, 5 Mb, 10 Mb, 20 Mb (u otros intervalos como se describe en este documento) para mapear de forma fina el punto de interrupción e identificar por ejemplo el gen o genes que están mal expresados como consecuencia de la translocación.

Típicamente se identificará una translocación mediante una transición abrupta de frecuencias de interacción bajas a altas en un cromosoma diferente a uno que contenga la secuencia 4C-señuelo, o en otra parte en ese mismo cromosoma.

En una realización preferida, la muestra del sujeto está en un estado pre-maligno.

En una realización preferida, la muestra del sujeto consiste en amniocitos cultivados o no cultivados obtenidos por amniocentesis para diagnóstico prenatal.

En un diseño de serie preferido como se describe en este documento, las sondas presentes en una única serie representan el genoma completo de una especie dada a máxima resolución. Por tanto, en las series para detectar translocaciones y similares por tecnología 4C contienen sondas descritas en este documento complementarias a cada lado de cada sitio de reconocimiento por enzima de restricción primaria en el genoma de una especie dada (por ejemplo, ser humano).

En otro diseño preferido como se describe en este documento, las sondas presentes en una única serie representan el genoma completo de una especie dada, pero no a resolución máxima. Por tanto, las series para detectar translocaciones y similares por tecnología 4C contienen sondas descritas en este documento que son complementarias a solamente un lado de cada sitio de reconocimiento por enzima de restricción primaria en el genoma de una especie dada (por ejemplo, ser humano).

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

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

En otro diseño preferido como se describe en este documento, las sondas presentes en una única serie representan el genoma completo de una especie dada, pero no a resolución máxima. Por tanto, las series para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C contienen sondas descritas en este documento complementarias a 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 por enzima de restricción primaria ordenadas a lo largo del molde lineal del genoma de una especie dada (por ejemplo, ser humano). Las series para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C pueden contener sondas descritas en este documento, que representan el genoma completo, pero con una única sonda cada 100 kilobases. Las series para detectar translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C pueden contener sondas descritas en este documento que representan cada sitio de reconocimiento por enzima de restricción primaria único en el genoma que pueden representarse por una secuencia de sonda única.

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

En otro diseño de serie preferido como se describe en este documento, las sondas descritas en este documento en una única serie representan regiones genómicas de un tamaño dado - tal como aproximadamente 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3Mb, 4Mb, 5Mb, 6Mb, 7Mb, 8Mb, 9Mb o 10Mb - (por ejemplo, de aproximadamente 50kb-10Mb) alrededor de una selecciones de loci que se sabe que están implicados en translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos. Las selecciones pueden hacerse según criterios adecuados, por ejemplo, pueden representar solamente los loci que están implicados en un tipo dado de enfermedad

En otro diseño de serie preferido como se describe en este documento, las sondas descritas en este documento en una única serie 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, 30 Mb, 40 Mb, 50 Mb, 60 Mb, 70 Mb, 80 Mb, 90 Mb, o 100 Mb (por ejemplo, 100kb-10Mb) (parte de) un cromosoma o múltiples cromosomas, representando cada sonda múltiples veces (por ejemplo, 10, 100, 1000) para permitir mediciones cuantitativas de intensidades de señal de hibridación en cada secuencia de sonda.

En un diseño experimental preferido, la secuencia 4C (señuelo) está en aproximadamente 0kb, 10kb, 20kb, 30kb, 40kb, 50kb, 100kb, 200kb, 300kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3Mb, 4Mb, 5Mb, 6Mb, 7Mb, 8Mb, 9Mb 10Mb, 11Mb, 12Mb, 13Mb, 14Mb o 15Mb (por ejemplo, de aproximadamente 0-15Mb) o más desde la secuencia reordenada real (es decir, punto de interrupción en caso de una translocación).

En una hibridación preferida, se hibridan dos moldes 4C marcados de forma diferencial obtenidos con una secuencia (señuelo-4C) de un sujeto enfermo y no enfermo de forma simultánea en la misma serie. Las diferencias en las interacciones ADN-ADN permiten la detección del punto de interrupción en *cis* (en el mismo cromosoma que el señuelo-4C) y en *trans* (en el compañero de translocación).

En una hibridación preferida, se hibridan múltiples moldes 4C marcados de forma diferencial obtenidos con una secuencia (señuelo-4C) de sujetos enfermos y no enfermos simultáneamente en la misma serie. Las diferencias en las interacciones ADN-ADN permiten la detección del punto de interrupción en *cis* (en el mismo cromosoma que el señuelo-4C) y en *trans* (en el compañero de translocación).

De forma ventajosa, pueden utilizarse análisis multicolor, en lugar de color dual en microseries que permite la hibridación simultánea de más de dos muestras en una única serie. Por consiguiente, puede usarse hibridación multicolor en tecnología 4C.

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

En otra hibridación preferida, se hibridan dos moldes 4C marcados de forma diferencial del mismo sujeto no enfermo, obtenidos con dos secuencias diferentes (señuelos-4C) que representan cada uno otro posible compañero de translocación, simultáneamente en la misma serie. Grupos de fuertes señales de hibridación observadas en el molde lineal de cromosomas no relacionados con el cromosoma que porta la secuencia de interés (señuelo-4C) identificarán el cromosoma compañero de translocación y el punto de interrupción en el compañero de translocación.

En otra hibridación preferida, se hibridarán múltiples moldes 4C marcados de forma diferencial del mismo sujeto no enfermo, obtenidos con múltiples secuencias diferentes (señuelos-4C) que representan cada uno otro posible compañero de translocación, simultáneamente en la misma serie. Grupos de señales fuertes de hibridación observadas en el molde lineal de cromosomas no relacionados con el cromosoma que porta la secuencia de interés (señuelo-4C) identificarán el cromosoma compañero de translocación y su punto de interrupción para la secuencia de interés.

El material usado para la detección de translocaciones, deleciones, inversiones, duplicaciones y otros reordenamientos genómicos por tecnología 4C pueden obtenerse por entrecruzamiento (y procesado adicional, como se describe) de células vivas y/o células muestras y/o lisados nucleares y/o cromatina aislada etc. (como se describe en este documento) de sujetos enfermos y/o no enfermos.

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 cerca (por ejemplo, hasta aproximadamente 1-15 Mb o más) de la secuencia 4C (señuelo).

La detección de inversiones (equilibradas) se basa en la identificación de aquellas interacciones ADN-ADN que eran diferentes entre sujetos enfermos y no enfermos. Las inversiones cambiarán la posición relativa (en kilobases) en el molde físico de ADN de todas las secuencias (pero las localizadas de forma más central) de la región reordenada medida frente a una secuencia adyacente al mismo cromosoma que se toma como secuencia 4C (señuelo). Como las frecuencias de interacción ADN-ADN están inversamente relacionadas con la separación del sitio genómico, los sujetos enfermos darán patrones invertidos de intensidades de hibridación para 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 la posición y tamaño de inversiones (equilibradas).

De acuerdo con este aspecto de la presente invención, un diseño de serie especializado preferido como se describe en este documento comprende sondas en una única sonda que representan regiones genómicas de un tamaño dado - 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, 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 serie especializado preferido como se describe en este documento, las sondas en una única serie representan regiones genómicas de un tamaño dado (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 otro reordenamiento. Para análisis cuantitativo fiable de intensidades de señal, la cantidad de sonda presente en la serie está típicamente en gran exceso a la cantidad de fragmentos afines que se hibridan en la serie. Por lo tanto, puede ser necesario tener cada sonda presente múltiples veces (por ejemplo, 10, 20, 50, 100, 1000 veces etc.) en la serie. Además, puede ser necesario titular la cantidad de molde que se tiene que hibridar en la sonda.

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 provocarán la ausencia de interacciones de ADN con una secuencia 4C (señuelo) localizada cerca (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 provocar la ausencia completa de señales de hibridación para todas las sondas localizadas en la región reordenada si la deleción está presente en ambos alelos (homocigótica), o una reducción para sujetos enfermos frente a no enfermos de intensidades de señal si la deleción está presente en solamente un alelo (heterocigótica). La deleción pone secuencias más distales en proximidad más cercana en el molde físico de ADN a la secuencia 4C analizada (señuelo), que provocará señales más fuertes de hibridación para sondas localizadas directamente más allá de la región delecionada.

Detección de duplicación(es)

La detección de duplicación se basa típicamente en la identificación de aquellas interacciones ADN-ADN que son diferentes entre sujetos enfermos y no enfermos. Las sondas en la región duplicada mostrarán señales aumentadas de hibridación con una secuencia 4C (señuelo) localizada cerca (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 reordenada, en comparación con señales de un sujeto de control no enfermo. Las sondas más allá de la región duplicada están más separadas de la secuencia 4C y por consiguiente mostrarán señales disminuidas de hibridación en comparación con señales de un sujeto de control no enfermo.

Preferiblemente, un aumento o disminución de la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una duplicación o inserción.

Preferiblemente, 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 regiones más distantes es indicativo de una duplicación o inserción.

Diagnóstico prenatal

De forma ventajosa, la tecnología 4C también puede usarse en diagnóstico prenatal.

El ácido nucleico puede obtenerse de un feto usando diversos métodos que son conocidos en la técnica. A modo de ejemplo, puede usarse amniocentesis para obtener fluido amniótico del cual se extraen células fetales en suspensión y se cultivan durante varios días (Mercier y Bresson (1995) Ann. Gnt., 38, 151-157). El ácido nucleico de las células puede entonces extraerse. La colección de vellosidades coriónicas puede hacer posible prescindir de la etapa de cultivo y evita la recogida de fluido amniótico. Estas técnicas pueden aplicarse precozmente (hasta 7 semanas de gestación para la recogida de vellosidades coriónicas y 13-14 semanas para amniocentesis), pero con un riesgo ligeramente aumentada de aborto.

También puede usarse una recogida directa de sangre fetal a nivel del cordón umbilical para obtener ácido nucleico, pero típicamente requiere un equipo de médicos especializados en esta técnica (Donner et al. (1996) Fetal Diagn. Ther., 10, 192-199).

De forma ventajosa, las aberraciones genéticas (por ejemplo, aberraciones genómicas o cromosómicas) - tales como reordenamientos, translocaciones, inversiones, inserciones, deleciones y otras mutaciones en cromosomas y ácido nucleico - pueden detectarse en esta fase.
Preferiblemente, las 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 de o la totalidad de los cromosomas 21, 18, 13, X o Y pueden detectarse ya que éstos son cromosomas en que la mayoría de las aberraciones suceden en el feto.

Determinación de sitios de integración genómica

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

Determinación de la predisposición a adquirir ciertas translocaciones

De forma ventajosa, la tecnología 4C también puede aplicarse a sujetos no enfermos para medir el entorno genómico de loci frecuentemente implicados en aberraciones genéticas. De este modo, es posible determinar la predisposición del sujeto a adquirir ciertas aberraciones genéticas.

Por tanto, además de los usos médicos descritos en este documento, la presente invención puede usarse en diagnóstico.

Sujeto

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

Agente

El agente identificado por los métodos de la invención puede ser un compuesto orgánico u otro agente químico. El agente puede ser un compuesto, que se puede obtener de o producirse por cualquier fuente adecuada, 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 polinucleotídica - que puede ser una molécula con sentido o anti-sentido, o un anticuerpo, por ejemplo, un anticuerpo policlonal, un anticuerpo monoclonal o un anticuerpo humanizado monoclonal.

Se han desarrollado diversas estrategias para producir anticuerpos monoclonales con carácter humano, que eluden la necesidad de una línea celular humana productora de anticuerpos. Por ejemplo, anticuerpos monoclonales de ratón útiles se han "humanizado" ligando regiones variables de roedor y regiones constantes humanas (Winter, G. y Milstein, C. (1991) Nature 349, 293-299). Esto reduce la inmunogenicidad humana anti-ratón del anticuerpo pero se retiene inmunogenicidad residual en virtud de la región flanqueante V foránea. Además, la especificidad de unión a antígeno es esencialmente la del donante murino. El injerto de CDR y la manipulación de regiones flanqueantes (documento EP 0239400) ha mejorado y refinado la manipulación de anticuerpos hasta el punto en que es posible producir anticuerpos murinos humanizados que son aceptables para uso terapéutico en seres humanos. Los anticuerpos humanizados pueden obtenerse usando otros métodos bien conocidos en la técnica (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, sí como otros compuestos, tales como moléculas orgánicas pequeñas.

A modo de ejemplo, la entidad puede ser una sustancia natural, una macromolécula biológica, o un extracto preparado a partir de materiales biológicos tales como bacterias, hongos, o células o tejidos animales (particularmente de mamífero), una molécula orgánica o inorgánica, un agente sintético, un agente semi-sintético, un mimético estructural o funcional, un péptido, un peptidomimético, un péptido escindido de una proteína completa, o péptidos sintetizados de forma sintética (tal como, a modo de ejemplo, usando un sintetizador de péptidos o por técnicas recombinantes o combinaciones de las mismas), un agente recombinante, un anticuerpo, un agente natural o no natural, una proteína de fusión o equivalente de la misma y mutantes, derivados o combinaciones de los mismos.

Típicamente, la entidad será un compuesto orgánico. En algunos casos, los compuestos orgánicos comprenderán dos o más grupos hidrocarbilo. Aquí, la expresión "grupo hidrocarbilo" significa un grupo que comprende al menos C y H y puede comprender opcionalmente uno o más sustituyentes adecuados diferentes. Ejemplos de dichos sustituyentes pueden incluir 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 un C entonces esos carbonos no tienen que estar ligados necesariamente 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 evidentes para los expertos en la materia e incluyen, por ejemplo, azufre, nitrógeno y oxígeno. Para algunas aplicaciones, preferiblemente 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 ligado a otro grupo hidrocarbilo.

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

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

Profármaco - para referencia

Los expertos en la materia apreciarán que la entidad puede derivarse a partir de un profármaco. Ejemplos de profármacos incluyen ciertos grupos protegidos que pueden no poseer actividad farmacológica tal cual, pero pueden, en ciertos casos, administrarse (tal como por vía oral o parenteral) y después de ello metabolizarse en el organismo para formar una entidad que es farmacológicamente activa.

Los profármacos adecuados pueden incluir, aunque sin limitación, Doxorrubicina, Mitomicina, Mostaza de Fenol, Metotrexato, Antifolatos, Cloranfenicol, Camptotecina, 5-Fluorouracilo, Cianuro, Quinina, Dipiridamol y Paclitaxel.

Se apreciará adicionalmente que ciertos restos conocidos como "pro-restos", por ejemplo como se describe en "Design of Prodrugs" de H. Bundgaard, Elsevier, 1985, puede colocarse en funcionalidades apropiadas de los agentes.

El agente puede estar en forma de una sal farmacéuticamente aceptable - tal como una sal de adición de ácidos o una sal de base - o un solvato del mismo, incluyendo un hidrato del mismo. Para una revisión sobre sales adecuadas véase Berge et al, J. Pharm. Sci., 1977, 66, 1-19.

El agente puede ser capaz de presentar otras propiedades terapéuticas.

El agente puede usarse en combinación con uno o más agentes farmacéuticamente activos diferentes.

Si se administran combinaciones de agentes activos, entonces las combinaciones de agentes activos pueden administrarse simultáneamente, por separado o secuencialmente.

Estereoisómeros e isómeros geométricos - para referencia

La entidad puede existes en forma de estereoisómeros y/o isómeros geométricos - por ejemplo la entidad puede poseer uno o más centros asimétricos y/o geométricos y así pueden existir en dos o más formas estereoisoméricas y/o geométricas.

Sal farmacéutica

El agente identificado por métodos de la invención puede administrarse en forma de una sal farmacéuticamente aceptable.

Las sales farmacéuticamente aceptables son bien conocidas para los expertos en la materia, y por ejemplo, incluyen aquellas mencionadas por Berge et al., en J.Pharm.Sci., 66, 1-19 (1977). Las sales adecuadas de adición de ácidos se forman a partir de ácidos que forman sales no tóxicas e incluyen las sales clorhidrato, bromhidrato, yodhidrato, nitrato, sulfato, bisulfato, fosfato, hidrogenofosfato, acetato, trifluoroacetato, gluconato, lactato, salicilato, citrato, tartrato, ascorbato, succinato, maleato, fumarato, gluconato, formiato, benzoato, metanosulfonato, etanosulfonato, bencenosulfonato y p-toluenosulfonato.

Cuando están presentes uno o más restos ácidos, las sales adecuadas de adición de bases farmacéuticamente aceptable pueden formarse a partir de bases que forman sales no tóxicas e incluyen las sales de aluminio, calcio, litio, magnesio, potasio, sodio, zinc, y aminas farmacéuticamente activas tales como dietanolamina.

Una sal farmacéuticamente aceptable de un agente puede prepararse fácilmente mezclando juntas soluciones del agente y el ácido o base deseado, según lo apropiado. La sal puede precipitar en la solución y recogerse por filtración o puede recuperarse por evaporación del disolvente.

El agente puede existir en forma polimórfica.

El agente puede contener uno o más átomos de carbono asimétricos y por lo tanto existe en dos o más formas estereoisoméricas. Cuando un agente contiene un grupo alquenilo o alquenileno, también puede suceder isomería cis (E) y trans (Z). También se describen en este documento los estereoisómeros individuales del agente y, cuando sea apropiado, las formas tautoméricas individuales de los mismos, junto con mezclas de los mismos.

La separación de diaestereoisómeros o isómeros *cis* y *trans* puede conseguirse por técnicas convencionales, por ejemplo por cristalización fraccionada, cromatografía o HPLC de una mezcla estereoisomérica del agente o una sal o derivado adecuado del mismo. También puede prepararse un enantiómero individual del agente a partir de un correspondiente intermedio ópticamente puro o por resolución, tal como por HPLC del racemato correspondiente usando un soporte quiral adecuado o por cristalización fraccionada de las sales diastereoisoméricas formadas por reacción del correspondiente racemato con un ácido o base adecuado ópticamente activo, según lo apropiado.

El agente también puede incluir todas las variaciones isotópicas adecuadas del agente o una sal farmacéuticamente aceptable del mismo. Una variación isotópica de un agente o una sal farmacéuticamente aceptable del mismo se define como una en que al menos un átomo está remplazado por un átomo que tiene el mismo número atómico pero una masa atómica diferente de la masa atómica habitualmente encontrada en la naturaleza. Ejemplos de isótopos que pueden incorporarse en el agente y sales farmacéuticamente aceptables del mismo incluyen isótopos de hidrógeno, carbono, nitrógeno, oxígeno, fósforo, azufre, flúor y cloro tales como ^{2}H, ^{3}H, ^{13}C, ^{14}C ^{15}N, ^{17}O, ^{18}O ^{31}P, ^{32}P, ^{35}S, ^{18}F y ^{36}Cl, respectivamente. Ciertas variaciones isotópicas del agente y sales farmacéuticamente aceptables del mismo, por ejemplo, aquellas en que se incorpora un isótopo radiactivo tal como ^{3}H o ^{14}C, son útiles en estudios de distribución tisular de fármaco y/o sustrato. Se prefieren los isótopos tritiado, es decir, ^{3}H, y carbono-14, es decir, ^{14}C, particularmente por su facilidad de preparación y detectabilidad. Además, la sustitución con isótopos tales como deuterio, es decir, ^{2}H, puede producir ciertas ventajas terapéuticas resultantes de la mayor estabilidad metabólica, por ejemplo, semi-vida aumentada *in vivo* o necesidades reducidas de dosificación y por tanto pueden preferirse en algunas circunstancias. Pueden prepararse variaciones isotópicas del agente y sales farmacéuticamente aceptables del mismo como se describe en este documento generalmente por procedimientos convencionales usando variaciones isotópicas apropiadas de reactivos adecuados.

Sal farmacéuticamente activa

El agente puede administrarse como una sal farmacéuticamente aceptable. Típicamente, una sal farmacéuticamente aceptable puede prepararse fácilmente usando un ácido o base deseada, según lo apropiado. La sal puede precipitar en solución y recogerse por filtración o puede recuperarse por evaporación del disolvente.

Métodos de síntesis química - para referencia

El agente puede prepararse por técnicas de síntesis química.

Será evidente para los expertos en la materia que puede ser necesario proteger y desproteger grupos funcionales sensibles durante la síntesis de un compuesto descrito en este documento. Esto puede conseguirse por técnicas convencionales, por ejemplo, como se describe en "Protective Groups in Organic Synthesis" de T W Greene y P G M Wuts, John Wiley and Sons Inc. (1991), y por P.J.Kocienski, en "Protecting Groups", Georg Thieme Verlag (1994).

Es posible, durante algunas de las reacciones, que algún estereocentro presente, en ciertas condiciones, pudiera racemizarse, por ejemplo, si se usa una base en una reacción con un sustrato que tiene un centro óptico que comprende un grupo sensible a bases. Esto es posible durante, por ejemplo, una etapa de guanilación. Debe ser posible evitar problemas tales como éste mediante la elección de la secuencia de reacción, las condiciones, reactivos, regímenes de protección/desprotección, etc. como se sabe bien en la técnica.

Los compuestos y sales pueden separarse y purificarse por métodos convencionales.

La separación de diastereómeros puede conseguirse por técnicas convencionales, por ejemplo por cristalización fraccionada, cromatografía o HPLC de una mezcla estereoisomérica de un compuesto de fórmula (I) o una sal o derivado adecuado del mismo. Un enantiómero individual de un compuesto de fórmula (I) también puede prepararse a partir de un intermedio correspondiente ópticamente puro o por resolución, tal como por HPLC del correspondiente racemato usando un soporte quiral adecuado o por cristalización fraccionada de las sales diastereoméricas formadas por reacción del correspondiente racemato con un ácido o base adecuado ópticamente activo.

El agente puede producirse usando métodos químicos para sintetizar el agente por completo o en parte. Por ejemplo, si el agente comprende un péptido, entonces el péptido puede sintetizarse por técnicas en fase sólida,

escindirse de la resina, y purificarse por cromatografía líquida preparativa de alto rendimiento (por ejemplo, Creighton (1983) Proteins Structures And Molecular Principles, WH Freeman and Co, Nueva York NY). La composición de los péptidos sintéticos puede confirmarse por análisis de aminoácidos o secuenciación (por ejemplo, el procedimiento de degradación de Edman; Creighton, *supra*).

La síntesis de agentes inhibidores peptídicos (o variantes, homólogos, derivados, fragmentos o miméticos de los mismos) puede realizarse usando diversas técnicas en fase sólida (Roberge JY et al (1995) Science 269: 202-204) y puede conseguirse síntesis automatizada, por ejemplo, usando el sintetizador de péptidos ABI 43 1 A (Perkin Elmer) de acuerdo con las instrucciones proporcionadas por el fabricante. Además, las secuencias de aminoácidos que comprenden el agente, pueden alterarse durante síntesis directa y/o combinarse usando métodos químicos con una secuencia de otras subunidades, o cualquier parte de las mismas, para producir un agente variante.

Derivado químico - para referencia

El término "derivado" o "derivatizado" como se usa en este documento incluye modificación química de un agente. Formas ilustrativas de dichas modificaciones químicas serían el remplazo de hidrógeno por un grupo halo, un grupo alquilo, un grupo acilo o un grupo amino.

Modificación química - para referencia

El agente puede ser un agente modificado - tal como, aunque sin limitación, una gente químicamente modificado.

La modificación química de un agente puede potenciar o reducir la interacción de enlaces de hidrógeno, interacción de cargas, interacción hidrófoba, interacción de Van Der Waals o interacción de dipolos.

En un aspecto, el agente puede actuar como modelo (por ejemplo, un molde) para el desarrollo de otros compuestos.

Composiciones farmacéuticas - para referencia

En un aspecto adicional, también se describe en este documento una composición farmacéutica que comprende un agente identificado por el método de ensayo descrito en este documento mezclado con un vehículo, diluyente, excipiente o adyuvante farmacéuticamente aceptable y/o combinaciones de los mismos.

En un aspecto adicional, también se describe en este documento una composición de vacuna que comprende un agente.

En un aspecto adicional, también se describe en este documento un proceso para preparar una composición farmacéutica que comprende mezclar un agente identificado por el ensayo con un diluyente, vehículo, excipiente o adyuvante farmacéuticamente aceptable y/o combinaciones de los mismos.

En un aspecto adicional, se proporciona un método para prevenir y/o tratar una enfermedad que comprende administrar un agente o una composición farmacéutica o una vacuna a un sujeto.

Las composiciones farmacéuticas pueden ser para uso en seres humanos o animales en medicina humana y veterinaria y típicamente comprenderán uno cualquiera o más de un diluyente, vehículo, o excipiente farmacéuticamente aceptable. Los vehículos o diluyentes aceptables para uso terapéutico son bien conocidos en la técnica farmacéutica, y se describen, por ejemplo, en Remington's Pharmaceutical Sciences, Mack Publishing Co. (A. R. Gennaro edit. 1985). La elección del vehículo, excipiente o diluyente farmacéutico puede seleccionarse con respecto a la vía pretendida de administración y la práctica farmacéutica convencional. Las composiciones farmacéuticas pueden comprender como - o además de - el vehículo, excipiente o diluyente cualquier aglutinante(s), lubricante(s), agente(s) de suspensión, agente(s) de recubrimiento, agente(s) solubilizante(s) adecuado(s).

Pueden proporcionarse conservantes, estabilizantes, colorante e incluso agentes aromatizantes en la composición farmacéutica. Ejemplos de conservantes incluyen benzoato sódico, ácido sórbico y ésteres de ácido p-hidroxibenzoico. También pueden usarse antioxidantes y agentes de suspensión.

Puede haber diferentes requisitos de composición/formulación dependiendo de los diferentes sistemas de suministro. A modo de ejemplo, la composición farmacéutica como se describe en este documento puede formularse para administrarse usando una mini-bomba o mediante una vía mucosa, por ejemplo, como pulverización nasal o aerosol para inhalación o solución ingerible, o por vía parenteral en que la composición se formula mediante una forma inyectable, para suministro por, por ejemplo, una vía intravenosa, intramuscular o subcutánea. Como alternativa, la formulación puede diseñarse para administrarse por varias vías.

Si el agente tiene que administrarse a la mucosa a través de la mucosa gastrointestinal, debe ser capaz de permanecer estable durante el tránsito a través del tracto gastrointestinal; por ejemplo, debe ser resistente a la degradación proteolítica, estable a pH ácido y resistente a los efectos detergentes de la bilis.

Cuando sea apropiado, las composiciones farmacéuticas pueden administrarse por inhalación, en forma de un supositorio o pesario, por vía tópica en forma de una loción, solución, crema, pomada o polvo salpicado, mediante el uso de un parche cutáneo, por vía oral en forma de comprimidos que contienen excipientes tales como almidón o lactosa, o en cápsulas u óvulos solos o en mezcla con excipientes, o en forma de elixires, soluciones o suspensiones que contienen agentes aromatizantes o colorantes, o las composiciones farmacéuticas pueden inyectarse por vía parenteral, por ejemplo, por vía intravenosa, intramuscular o subcutánea. Para administración parenteral, las composiciones pueden usarse mejor en forma de una solución acuosa estéril que puede contener otras sustancias, por ejemplo, sales o monosacáridos suficientes para hacer la solución isotónica con la sangre. Para administración bucal o sublingual las composiciones pueden administrarse en forma de comprimidos o grageas que puede formularse de manera convencional.

Los agentes pueden usarse en combinación con una ciclodextrina. Las ciclodextrinas son conocidas por formar complejos de inclusión y no de inclusión con moléculas de fármaco. La formación de un complejo fármaco-ciclodextrina puede modificar la solubilidad, tasa de disolución, biodisponibilidad y/o propiedad de estabilidad de una molécula de fármaco. Los complejos fármaco-ciclodextrina generalmente son útiles para la mayoría de las formas de dosificación y vías de administración. Como alternativa a la formación directa de complejos con el fármaco, la ciclodextrina puede usarse como aditivo auxiliar, por ejemplo como un vehículo, diluyente o solubilizante. Las alfa-, beta- y gamma-ciclodextrinas son las más habitualmente usadas y se describen ejemplos adecuados en los documentos WO-A-91/11172, WO-A-94/02518 y WO-A-98/55148.

Si el agente es una proteína, entonces dicha proteína puede prepararse *in situ* en el sujeto que se está tratando. A este respecto, las secuencias nucleotídicas que codifican dicha proteína pueden suministrarse mediante el uso de técnicas no virales (por ejemplo, mediante el uso de liposomas) y/o técnicas virales (por ejemplo, mediante el uso de vectores retrovirales) de modo que dicha proteína se exprese a partir de dicha secuencia nucleotídica.

Las composiciones farmacéuticas descritas en este documento pueden usarse también en combinación con tratamientos convencionales.

Administración - para referencia

El término "administrado" incluye suministro por técnicas virales o no virales. Los mecanismos de suministro viral incluyen, aunque sin limitación, vectores adenovirales, vectores virales adeno-asociados (AAV), vectores virales de herpes, vectores retrovirales, vectores lentivirales, y vectores baculovirales. Los mecanismos de suministro no viral incluyen transfección mediada por lípidos, liposomas, inmunoliposomas, lipofectina, anfífilos faciales catiónicos (CFA) y combinaciones de los mismos.

Los componentes pueden administrarse solos pero generalmente se administrarán como una composición farmacéutica - por ejemplo cuando los componentes están en mezcla con un excipiente, diluyente o vehículo farmacéutico adecuado seleccionado con respecto a la vía pretendida de administración y la práctica farmacéutica convencional.

Por ejemplo, los componentes pueden administrarse en forma de comprimidos, cápsulas, óvulos, elixires, soluciones o suspensiones, que pueden contener agentes aromatizantes o colorantes, para aplicaciones de liberación inmediata, retardada, modificada, sostenida, pulsada o controlada.

Si la forma farmacéutica es un comprimido, entonces el comprimido puede contener excipientes tales como celulosa microcristalina, lactosa, citrato sódico, carbonato de calcio, fosfato cálcico dibásico y glicina, disgregantes tales como almidón (preferiblemente almidón de maíz, patata o tapioca), almidón glicolato sódico, croscarmelosa sódica y ciertos silicatos complejos, y aglutinantes de granulación tales como polivinilpirrolidona, hidroxipropilmetilcelulosa (HPMC), hidroxipropilcelulosa (HPC), sacarosa, gelatina y goma arábiga. Además, pueden incluirse agentes lubricantes tales como estearato de magnesio, ácido esteárico, behenato de glicerilo y talco.

También pueden emplearse composiciones sólidas de un tipo similar como rellenos en cápsulas de gelatina. Los excipientes preferidos a este respecto incluyen lactosa, almidón, una celulosa, azúcar de la leche o polietilenglicoles de alto peso molecular. Para suspensiones acuosas y/o elixires, el agente puede combinarse con diversos agentes edulcorantes o aromatizantes, materia colorante o colorantes, con agentes emulsionantes y/o de suspensión y con diluyentes tales como agua, etanol, propilenglicol y glicerina, y combinaciones de los mismos.

Las vías para administración (suministro) puede incluir, aunque sin limitación, una o más de oral (por ejemplo, en forma de un comprimido, cápsula, o en forma de una solución ingerible), tópica, a la mucosa (por ejemplo, en forma de una pulverización nasal o aerosol para inhalación), nasal, parenteral (por ejemplo, mediante una forma inyectable), gastrointestinal, intraespinal, intraperitoneal, intramuscular, intravenosa, intrauterina, intraocular,

intradérmica, intracraneal, intratraqueal, intravaginal, intracerebroventricular, intracerebral, subcutánea, oftálmica (incluyendo intravítrea o intracameral), transdérmica, rectal, bucal, vaginal, epidural, sublingual.

Niveles de dosificación - para referencia

Típicamente, un médico determinará la dosificación real que será la más adecuada para un sujeto individual. El nivel específico de dosis y frecuencia de dosificación para cualquier paciente particular pueden variarse y dependerán de una diversidad de factores incluyendo la actividad del compuesto específico empleado, la estabilidad metabólica y la duración de acción de ese compuesto, la edad, peso corporal, salud general, sexo, dieta, modo y tiempo de administración, tasa de excreción, combinación de fármacos, la gravedad de la afección particular, y la terapia que está experimentando el individuo.

Formulación - para referencia

El componente o componentes pueden formularse en una composición farmacéutica, tal como por mezcla con uno o más de un vehículo, diluyente o excipiente adecuado, usando técnicas que son conocidas en la técnica.

Enfermedad

Pueden usarse aspectos de la presente descripción para el tratamiento y/o prevención y/o diagnóstico y/o pronóstico de una enfermedad - tal como las enumeradas en el documento WO-A-98/09985.

Para facilidad de referencias, parte de esa lista se proporciona ahora: actividad inhibidora de macrófagos y/o de células T y por tanto, actividad anti-inflamatoria; actividad anti-inmune, es decir efectos inhibidores contra una respuesta inmune celular y/o humoral, incluyendo una respuesta no asociada con inflamación; enfermedades asociadas con virus y/u otros patógenos intracelulares; inhibición de la capacidad de los macrófagos y células T de adherirse a componentes de la matriz extracelular y fibronectina, así como expresión regulada positivamente del receptor fas en células T; inhibición de reacción inmune indeseada e inflamación incluyendo artritis, incluyendo artritis reumatoide, inflamación asociada con hipersensibilidad, reacciones alérgicas, asma, lupus sistémico eritematoso, enfermedades de colágeno y otras enfermedades autoinmunes, inflamación asociada con aterosclerosis, arteriosclerosis, enfermedad cardiaca aterosclerótica, lesión por reperfusión, paro cardiaco, infarto de miocardio, trastornos inflamatorios vasculares, síndrome de distrés respiratorio u otras enfermedades cardiopulmonares, inflamación asociada con úlcera péptica, colitis ulcerosa y otras enfermedades del tracto 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 otorrino-laringológicas, dermatitis u otras enfermedades dérmicas, enfermedades periodontales u otras enfermedades dentales, orquitis o epidídimo-orquitis, infertilidad, traumatismo orquidal u otras enfermedades testiculares relacionadas con el sistema inmune, disfunción placentaria, insuficiencia placentaria, aborto habitual, eclampsia, pre-eclampsia y otras enfermedades ginecológicas inmunes y/o relacionadas con inflamación, uveítis posterior, uveítis intermedia, uveítis anterior, conjuntivitis, corio-retinitis, uveo-retinitis, neuritis óptica, inflamación intraocular, por ejemplo retinitis o edema macular cistoide, oftalmia simpática, escleritis, retinitis pigmentosa, componentes inmune e inflamatorios de enfermedad degenerativa del fondo, componentes inflamatorios de traumatismo ocular, inflamación ocular causada por infección, vitreo-retinopatías proliferativas, neuropatía óptica por isquemia aguda, cicatrización excesiva, por ejemplo después de operación por filtración de glaucoma, reacción inmune y/o inflamación contra implantes oculares y otras enfermedades oftálmicas inmunes y relacionadas con inflamación, inflamación asociada con enfermedades autoinmunes o afecciones o trastornos donde, tanto en el sistema nervioso central (SNC) como en cualquier otro órgano, sería beneficiosa la supresión inmune y/o de la inflamación, enfermedad de Parkinson, complicación y/o efectos secundarios del tratamiento de la enfermedad de Parkinson, complejo de demencia relacionada con SIDA por encefalopatía relacionada con VIH, enfermedad de Devic, corea de Sydenham, enfermedad de Alzheimer y otras enfermedades degenerativas, afecciones o trastornos del SNC, componentes inflamatorios de apoplejías, síndrome post-polio, componentes inmunes e inflamatorios de trastornos psiquiátricos, mielitis, encefalitis, pan-encefalitis esclerosante subaguda, encefalomielitis, neuropatía aguda, neuropatía subaguda, neuropatía crónica, síndrome de Guillaim-Barre, corea Sydenham, miastenia grave, pseudo-tumor cerebral, síndrome de Down, enfermedad de Huntington, esclerosis lateral amiotrófica, componentes inflamatorios de la compresión del SNC o traumatismo del SNC o infecciones del SNC, componentes inflamatorios de atrofias y distrofias musculares, y enfermedades relacionadas con el sistema inmune e inflamación, afecciones o trastornos del sistema nervioso central y periférico, inflamación post-traumática, choque séptico, enfermedades infecciosas, complicaciones inflamatorias o efectos secundarios de cirugía, trasplante de médula ósea u otras complicaciones de trasplante y/o efectos secundarios, complicaciones inflamatorias y/o inmunes y efectos secundarios de terapia génica, por ejemplo debido a infección con un portador viral, o inflamación asociada con SIDA, para suprimir o inhibir una respuesta inmune 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 tratamiento de rechazo de injertos en casos de trasplante de células, tejido y órganos naturales o artificiales tales como córnea, médula ósea, órganos, cristalino, marcapasos, tejido cutáneo natural o artificial. Los trastornos específicos relacionados con cáncer incluyen aunque sin limitación: tumores sólidos; tumores producidos en la sangre tales como leucemias; metástasis tumoral; tumores benignos, por ejemplo hemangiomas, neuromas

acústicos, neurofibromas, tracomas, y granulomas piogénicos; artritis reumatoide; psoriasis; enfermedad angiogénica ocular, por ejemplo, retinopatía diabética, retinopatía del prematuro, degeneración macular, rechazo de injerto de córnea, glaucoma neovascular, fibroplasia retrolental, rubeosis; síndrome de Osler-Webber; angiogénesis de miocardio; neovascularización de la placa; telangiectasia; articulaciones hemofílicas; angiofibroma; granulación de heridas; circulación colateral coronaria; circulación colateral cerebral; malformaciones arteriovenosas; angiogénesis de extremidad isquémica; glaucoma neovascular; fibroplasia retrolental; neovascularización diabética; enfermedades relacionadas con heliobacter, fracturas, vasculogénesis, hematopoyesis, ovulación, menstruación y placentación.

Preferiblemente, 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 huesos, tumor cerebral, cáncer de mama, cáncer del sistema genital femenino, cáncer del sistema genital masculino, linfoma del sistema nervioso central, cáncer cervical, rabdomiosarcoma infantil, sarcoma infantil, leucemia linfocítica crónica (CLL), leucemia mieloide crónica (CML), cáncer de colon y rectal, cáncer de colon, cáncer de endometrio, sarcoma de endometrio, cáncer esofágico, cáncer ocular, cáncer de vesícula biliar, cáncer gástrico, cáncer de tracto gastrointestinal, leucemia de células capilares, cáncer de cabeza y cuello, cáncer hepatocelular, enfermedad Hodgkin, cáncer hipofaríngeo, sarcoma de Kaposi, cáncer renal, cáncer laríngeo, leucemia, cáncer de hígado, cáncer pulmonar, histiocitoma fibroso maligno, timoma maligno, melanoma, mesotelioma, mieloma múltiple, mieloma, cáncer de cavidad nasal y seno paranasal, cáncer nasofaríngeo, cáncer del sistema nervioso, neuroblastoma, linfoma no Hodgkin, cáncer de la cavidad oral, cáncer orofaríngeo, osteosarcoma, cáncer de ovario, cáncer pancreático, cáncer de paratiroides, cáncer de pene, cáncer faríngeo, tumor de la pituitaria, neoplasia de células plasmáticas, linfoma primario del SNC, cáncer de próstata, cáncer rectal, sistema respiratorio, retinoblastoma, cáncer de glándula salivar, cáncer de piel, cáncer del intestino delgado, sarcoma de tejido blando, cáncer de estómago, cáncer de estómago, cáncer testicular, cáncer de tiroides, cáncer del sistema urinario, sarcoma uterino, cáncer vaginal, sistema vascular, macroglobulinemia de Waldenstrom y tumor de Wilms.

Kits

Los materiales para su uso en los métodos de la presente invención son idealmente adecuados para la preparación de kits.

Dicho kit puede comprender recipientes, cada uno con uno o más de los diversos reactivos (típicamente en forma concentrada) utilizados en los métodos descritos en este 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 revertir el entrecruzamiento (por ejemplo proteinasa K).

También pueden proporcionarse oligonucleótidos en recipientes que pueden estar en cualquier forma, por ejemplo, liofilizados, o en solución (por ejemplo, una solución en agua destilada o tamponada), etc.

En un aspecto preferido de la presente descripción se proporciona un kit que comprende un conjunto de sondas como se describe en este documento, una serie y opcionalmente uno o más marcadores.

Típicamente también se incluirá un conjunto de instrucciones.

Usos

De forma ventajosa, la presente descripción puede usarse para obtener información acerca de la organización espacial de secuencias nucleotídicas - tales como loci genómicos *in vitro* o *in vivo.*

A modo de ejemplo, la tecnología 4C puede usarse para estudiar la organización tridimensional de uno o más loci génicos. En particular, esta tecnología puede usarse para estudiar el papel de uno o más factores de transcripción en la organización tridimensional de uno o más loci génicos.

A modo de ejemplo adicional, la tecnología 4C puede usarse para estudiar el papel de factores de acción en *trans* y elementos de ADN reguladores en *cis*.

A modo de ejemplo adicional, la tecnología 4C puede usarse para estudiar la regulación génica de rango largo *in vitro* o *in vivo.*

A modo de ejemplo adicional, la tecnología 4C puede usarse para estudiar la proximidad e interacción intra-cromosómica.

A modo de ejemplo adicional, la tecnología 4C puede usarse para estudiar la proximidad e interacción inter-cromosómica.

A modo de ejemplo adicional, la tecnología 4C puede usarse para identificar secuencias nucleotídicas que funcionen con un promotor, potenciador, silenciador, aislante, región de control de locus, origen de replicación, MAR, SAR, centrómero, telómero o cualquier otra secuencia de interés en una red reguladora.

A modo de ejemplo adicional, la tecnología 4C puede usarse para identificar genes responsables de un fenotipo (enfermedad) en casos donde sucede mutación y/o deleción que afecta al elemento regulador distante y su mapeado por lo tanto no logra proporcionar dicha información.

A modo de ejemplo adicional, la tecnología 4C puede usarse para reconstruir finalmente la conformación espacial de loci génicos, regiones genómicas grandes o incluso cromosomas completos.

A modo de ejemplo adicional, la tecnología 4C puede usarse para definir secuencias de anclaje potenciales que mantienen ciertos cromosomas juntos en el espacio nuclear.

A modo de ejemplo adicional, la tecnología 4C puede usarse para reconstruir finalmente a alta resolución el posicionamiento de cromosomas unos con respecto a otros.

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

Técnicas de metodología de ADN recombinante general

La presente invención emplea, salvo que se especifique de otro modo, técnicas convencionales de química, biología molecular, microbiología, ADN recombinante e inmunología, que están dentro de las capacidades de los expertos en la materia. 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, cap. 9, 13, y 16, John Wiley & Sons, Nueva York, N.Y.); B. Roe, J. Crabtree, y A. Kahn, 1996, DNA Isolation and Sequencing: Essential Technique, John Wiley & 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.

La invención se describirá ahora adicionalmente mediante Ejemplos, que pretenden servir para ayudar a los expertos en la materia a realizar la invención y no pretenden de ningún modo limitar el alcance de la invención.

Ejemplo 1

Materiales y métodos

Tecnología 4C

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

Las reacciones de PCR se realizaron usando el sistema de PCR de molde largo Expand (Roche), usando condiciones optimizadas cuidadosamente para asegurar la amplificación lineal de fragmentos con tamaños de hasta 1,2 kb (80% de los fragmentos 4C-PCR son más pequeños de 600 pb). Las condiciones de PCR fueron las siguientes: 94ºC durante 2 minutos, 30 ciclos de 94ºC durante 15 segundos, 55ºC durante 1 minuto y 68ºC durante 3 minutos, seguido de una etapa final de 68ºC durante 7 minutos. Se determinó la cantidad máxima de molde que aún muestra intervalo lineal de amplificación. Para esto, se añadieron diluciones en serie de molde a reacciones de PCR,

el material de ADN amplificado se procesó en un gel de agarosa y los productos de PCR se cuantificaron usando el software ImageQuant. Típicamente, 100-200 ng de molde por 50 µl de reacción de PCR dieron productos en el intervalo lineal de amplificación. Se combinaron de 16 a 32 reacciones de PCR y se purificó este molde 4C usando el sistema de retirada de nucleótidos QIAquick (250) (Qiagen). El molde 4C purificado se marcó e hibridó en series de acuerdo con protocolos convencionales ChIP-chip (Nimblegen Systems of Iceland, LLC). ADN genómico marcado de forma diferencial, que se había digerido con la enzima primaria y secundaria usadas en el procedimiento 4C, sirvió como molde de control para corregir diferencias en las eficacias de hibridación. Para cada experimento se marcador dos muestras procesadas independientemente con orientaciones alternativas de colorante.

Secuencias de cebador 4C usadas:

HS2: 5' -ACTTCCTACACATTAACGAGCC-3',
5'- GCTGTTATCCCTTTCTCTTCTAC-3'

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

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

Series 4C

Las series y el análisis se basaron en m34 construido en NCBI. Las sondas (60 monómeros) se seleccionaron entre las secuencias 100 pb cadena arriba y abajo de sitios HindIII. El contenido en CG se optimizó hacia el 50%, para señales uniformes de hibridación. Para evitar la hibridación cruzada, se retiraron las sondas que tenían cualquier similitud con repeticiones altamente abundantes (RepBase 10,09)[3] del conjunto de sondas. Además, también se retiraron las sondas que daban más de dos aciertos BLAST en el genoma del conjunto de sondas. Se realizaron alineaciones de secuencia usando MegaBLAST (Zhang et al. (2000) J Comput Biol 7, 203-14) usando las configuraciones convencionales. Un acierto se definió como una alineación de 30 nt o más larga.

Análisis de datos 4C

La relación de señal de muestra 4C/ADN genómico se calculó para sonda y los datos se visualizaron con el software SignalMap proporcionado por Nimblegen Systems. Los datos se analizaron usando el paquete R (http://www.r-project.org), Spotfire y Excel. Las relaciones de hibridación no procesadas mostraron grupos de 20-50 señales 4C positivas a lo largo del molde de cromosoma. Para definir estos grupos, se aplicó una media de ejecución. Se usaron diversos tamaños de ventana, que variaban de 9-39 sondas, que todas identificaban los mismos grupos. Los resultados mostrados se basaron en un tamaño de ventana de 29 sondas (en promedio 60 kb) y se compararon con la media de ejecución realizada en todos los datos aleatorizados.

Esto se hizo para cada serie por separado. Por consiguiente, todas las mediciones se apreciaron respecto a la amplitud y ruido de esa serie específica. La tasa de descubrimiento de falsos (FDR), definida como (nº falsos positivos) / (nº de falsos positivos + nº de verdaderos positivos) se determinó del siguiente modo: (número de positivos en el conjunto aleatorizado) / (número de positivos en los datos). El nivel umbral se determinó usando un enfoque vertical para establecer el valor mínimo para el cual: FDR<0,05.

A continuación, se compararon experimentos biológicos duplicados. Las ventanas que cumplían el umbral en ambos duplicados se consideraron positivas. Cuando se comparan datos aleatorizados, ninguna ventana estaba por encima del umbral en ambos duplicados. Las ventanas positivas directamente adyacentes en el molde de cromosoma se unieron (sin huecos permitidos), creando áreas positivas.

Análisis de expresión

Para cada tejido, se realizaron tres microseries independientes de acuerdo con el protocolo Affymetrix (series 430_2 de ratón). Los datos se normalizaron usando RMA ca-tools; www.bioconductor.org) y para cada conjunto de sondas se promediaron las mediciones de las tres microseries. Además, cuando múltiples conjuntos de sondas representaban el mismo gen, también se promediaron. Se usó Mas5calls (biblioteca Affy: www.bioconductor.org) para establecer las calificaciones "presente", "ausente" y "marginal". Los genes con una calificación "presente" en las tres series y un valor de expresión mayor de 50 se calificaron expresados. Los 'genes específicos de hígado fetal' se clasificaron como genes que cumplían nuestros criterios de expresarse en hígado fetal y obtuvieron valores de expresión más de cinco veces mayores en comparación con cerebro fetal. Para proporcionar una medida de la actividad global de transcripción alrededor de cada gen, se aplicó una suma de ejecución. Para esto, se usaron

valores de expresión transformados en log. 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 áreas 4C positivas (n = 153, 301 y 186, respectivamente, donde n=153 corresponde a la cantidad de genes activos, no de interacción, presentes entre las regiones de interacción más centroméricas y el telómero del cromosoma 7); los dos grupos se compararon usando un ensayo de suma de rango de Wilcoxon unilateral.

Sondas FISH

Se usaron los siguientes clones BAC (BACPAC Resources Centre); RP23-370E12 para Hbb-1, RP23-317H16 para chr.7 en 80,1 Mb (grupo génico OR), RP23-334E9 para Uros, RP23-32C19 para chr.7 a 118,3 Mb, RP23-143F10 para chr.7 a 130,1 Mb, RP23-470N5 para chr.7 a 73,1 Mb, RP23-247L11 para chr.7 a 135,0 Mb (grupo génico OR), RP23-136A15 para Rad23A, RP23-307P24 para chr.8 a 21,8 Mb y RP23-460F21 para chr.8 a 122,4 Mb. Para una sonda específica de centrómero de cromosoma 7 se usó el clon P1 5279 (Genome Systems Inc.) que hibrida con el segmento de ADN D7Mit21. Se prepararon sondas marcadas por cebado aleatorio usando el sistema de marcaje genómico BioPrime Array CGH (Invitrogen). Antes del marcaje, se digirió el ADN con DpnII y se purificó con un kit de limpieza y concentración de ADN-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 kit de purificación de ADN de PCR y bandas de gel GFX (Amersham Biosciences) para retirar los nucleótidos no incorporados. La especificidad de las sondas marcadas se ensayó en extensiones de metafase preparadas a partir de células ES murinas.

Crio-FISH

Se realizó crio-FISH como se ha descrito anteriormente[5]. En resumen, se fijaron hígado y cerebro E14.5 durante 20 min en paraformaldehído al 4%/HEPES 250 mM, pH 7,5 y se cortaron en bloques tisulares pequeños, seguido de otra etapa de fijación de 2 h en paraformaldehído al 8% a 4ºC. Los bloques tisulares fijados se sumergieron en sacarosa 2,3 M durante 20 min a temperatura ambiente, se montaron en un portamuestras y se congelaron instantáneamente en nitrógeno líquido. Los bloques tisulares se almacenaron en nitrógeno líquido hasta el seccionamiento. Se cortaron criosecciones ultrafinas de aproximadamente 200 nm usando un Ultramicrótomo Reichert E equipado con crio-adhesión (Leica). Usando una carga completa llena de sacarosa, se transfirieron secciones a cubreobjetos y se almacenaron a -20ºC. Para hibridación, las secciones se lavaron con PBS para retirar la sacarosa, se trataron con 250 ng/ml de RNasa en SSC 2x durante 1 h a 37ºC, se incubaron durante 10 min en HCl 0,1 M, se deshidrataron en una serie de etanol y se desnaturalizaron durante 8 min a 80ºC en formamida al 70%/SSC 2x, pH 7,5. Las secciones de deshidrataron de nuevo directamente antes de la hibridación de las sondas. Se co-precipitaron 500 ng de sonda marcada con 5 μg de ADN Cot1 de ratón (Invitrogen) y se disolvieron en mezcla de hibridación (formamida al 50%, sulfato de dextrano al 10%, SSC 2x, tampón fosfato 50 mM, pH 7,5). Las sondas de desnaturalizaron durante 5 min a 95ºC, se volvieron a unir durante 30 min a 37ºC y se hibridaron durante al menos 40 h a 37ºC. Después de los lavados post-hibridación, los núcleos se tiñeron con contraste con 20 ng/ml de DAPI (Sigma) en PBS/Tween-20 al 0,05% y se montaron en reactivo anti-extinción Prolong Gold (Molecular Probes).

Se recogieron imágenes con un microscopio de epifluorescencia Zeiss Axio Imager Z1 (apocromático plano x100, objetivo en aceite 1,4), equipado con una cámara CCD y el software Isis FISH Imaging System (Metasystems). Se analizó un mínimo de 250 alelos de β-globina o Rad23A y se valoraron como solapantes o no solapantes con BAC localizados en otras partes en el genoma, por una persona que no conocía la combinación de sondas aplicada a las secciones. Se realizaron ensayos de bondad de ajuste replicados (estadística G)[6] para evaluar la significancia de diferencias entre valores medidos para regiones 4C positivas frente a 4C negativas. Se proporciona una visión global de los resultados en la Tabla 2.

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 inferiores que las medidas por otros usando diferentes protocolos FISH. El seccionamiento puede separar algunos loci de interacción y las mediciones por crio-FISH por lo tanto serán frecuencias de interacción verdaderas ligeramente subestimadas. Por otro lado, los actuales procedimientos FISH 2D y 3D sobre-estimarán estos porcentajes debido a la resolución limitada en la dirección z. En el futuro, técnicas de microscopía mejoradas en combinación con sondas FISH más específicas revelarán mejor las frecuencias de interacción verdaderas.

Ejemplo 2

El procedimiento 3C (es decir fijación con formaldehído, digestión con enzima de restricción (primaria), re-ligamiento de fragmento 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 cerca en el espacio nuclear.

Se realiza PCR inversa para amplificar todos los fragmentos ligados a un fragmento de restricción dado ('señuelo'; elegido porque contiene un promotor, potenciador, aislante, región de adhesión a matriz, origen de replicación o cualquier otra primera secuencia nucleotídica (diana)).

Para esto, se crean círculos de ADN digiriendo el molde 3C con una enzima de restricción secundaria (preferiblemente una cortadora frecuente que reconoce secuencias tetra- o penta-nucleotídicas), seguido de ligamiento en condiciones diluidas de modo que se favorezcan las interacciones intra-moleculares. Para minimizar la predisposición en la formación de círculos debido a restricciones topológicas (Rippe et al, (2001) Trends in Biochem. Sciences 26, 733-40), debe elegirse una enzima de restricción secundaria que corte preferiblemente el señuelo en >350-400 pb desde el sitio de restricción primario. Para aumentar la eficacia y reproducibilidad de la amplificación por PCR inversa, los círculos se linealizan mejor antes de la amplificación por PCR mediante una enzima de restricción (por ejemplo, una cortadora de 6 o más pb) que corta el señuelo 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 de ligamiento en condiciones diluidas y la linealización de círculos que contienen señuelo se realizan en condiciones convencionales para dichas manipulaciones de ADN para producir un molde de ADN para amplificación por PCR inversa ('molde 4C').

Por consiguiente, se digieren 10 μg de molde 3C en 100 μl con 20 U de la enzima de restricción secundaria (durante una noche), seguido de inactivación por calor de la enzima y purificación de ADN. El ligamiento se realiza en 10 ml (1 ng/μl de ADN) con 50 U de ligasa T4 (4 h a 16ºC, 30 min a TA), seguido de purificación de ADN. Finalmente, se hace la linealización de los círculos de interés en 100 μl con 20 U de enzima de restricción (durante una noche), seguido de nuevo por purificación de ADN.

Para PCR inversa, se diseñan dos cebadores específicos de señuelo, cada uno lo más cerca posible al sitio de reconocimiento por enzima de restricción primaria y secundaria directamente adyacente, respectivamente, y cada uno con su extremo 3' enfocado hacia afuera de modo que la extensión proceda inmediatamente a través de los sitios de restricción en un fragmento ligado al señuelo. La PCR inversa con estos cebadores se realiza preferiblemente sobre 100-400 ng de ADN de molde 4C (por 50 μl de mezcla de reacción de PCR), para incluir una cantidad máxima de eventos de ligamiento por reacción de PCR. Se realza PCR inversa aplicando el sistema de PCR de molde largo Expand (Roche), usando tampón 1 de acuerdo con los procedimientos del fabricante.

Se realizan los siguientes ciclos de PCR:

1. 2 min 94ºC
2. 15 segundos 94ºC
3. 1 min 55ºC
4. 3 min 68ºC
5. repetir etapa 2-4 29x (o algo entre 25-40x)
6. 7 min 68ºC
7. fin

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

Para obtener suficiente material para marcaje por cebado aleatorio e hibridación en serie, pueden combinarse múltiples reacciones de PCR (cada una obtenida después de 30 ciclos de PCR), (en lugar de aumentar la cantidad de ciclos de PCR por reacción). Como alternativa para marcaje cebado aleatorio, pueden incorporarse nucleótidos marcados en los últimos ciclos de PCR (por ejemplo, 30 ciclos (sin marcador) + 10 ciclos (marcador)).

Ejemplo 3

Detección de translocaciones usando tecnología 4C

La tecnología 4C se usa para medir las frecuencias de interacción para una secuencia X dada presente en un cromosoma dado A en células de un sujeto sano y en células de un paciente que porta una única translocación recíproca entre el cromosoma A y B estando el punto de interrupción cerca de la secuencia X (como se muestra en la Figura 9).

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

En células de paciente sin embargo, las señales de hibridación con todas las sondas del cromosoma A localizadas en el otro lado del punto de interrupción se reducen en ~50% (una copia de cromosoma A está aún intacta y producirá señales normales), mientras que se observa una concentración única (es decir no presente en células normales) de señales elevadas de hibridación para sondas que limitan el punto de interrupción en el cromosoma B. De hecho, la abrupta transición entre sondas que no muestras señales frente a señales fuertes de hibridación en el cromosoma B revela la localización del punto de interrupción en el cromosoma B.

Ejemplo 4

Análisis de tecnología 4C, resultados

La tecnología 4C se usó para caracterizar el entorno genómico de la región de control del locus (LCR) de β-globina de ratón, que se centra en un fragmento de restricción que contiene su sitio hipersensible 2 (HS2). La LCR es un fuerte elemento regulador de la transcripción específico eritroide necesario para altos niveles de expresión génica de β-globina. El locus de β-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 solamente en neuronas olfativas. Se analizaron las interacciones en dos tejidos: hígado fetal E14.5, donde la LCR es activa y los genes de β-globina se transcriben elevadamente, y cerebro fetal E14.5, donde la LCR es inactiva y los genes de globina son silenciosos. En ambos tejidos, la gran mayoría de las interacciones se encontraron con secuencias en el cromosoma 7 y se detectaron muy pocas interacciones de LCR con seis cromosomas no relacionados (8, 10, 11, 12, 13, 14) (Figura 13a). Las señales más fuertes en el cromosoma 7 se encontraron en una región de 5-10 Mb centrada alrededor de la posición cromosómica de β-globina, en acuerdo con la idea de que las frecuencias de interacción son inversamente proporcionales a la distancia (en pares de bases) entre secuencias de ADN físicamente ligadas. No fue posible interpretar las interacciones en esta región de forma cuantitativa. Se razonó que estas secuencias cercanas estaban junto con la β-globina tan frecuentemente que su gran sobre-expresión en las muestras de hibridación saturaba 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 señal estaba reducida en sondas fuera y en el borde, pero no dentro de esta región (datos no mostrados).

El procedimiento 4C produjo datos altamente reproducibles. La Figura 2b-c muestra relaciones no procesadas de señales 4C sobre señales de hibridación de control para dos regiones de 1,5 Mb en el cromosoma 7, aproximadamente 25 Mb y 80 Mb desde el gen de β-globina. A este nivel de resolución los resultados de muestras procesadas independientemente fueron casi idénticos. Tanto en hígado fetal como en cerebro, se identificaron grupos de señales positivas en el cromosoma 7, a menudo en localizaciones cromosómicas decenas de megabases desde la β-globina. Estos grupos típicamente consistían de mínimamente 20-50 sondas con relaciones de señal aumentadas yuxtapuestas en el molde de cromosoma (Figura 13b-c). Cada sonda en la serie analiza un evento independiente de ligamiento. Además, están presentes solamente dos copias del fragmento de restricción HS2 por célula, cada uno de los cuales puede ligar solamente con un fragmento de restricción diferente. Por lo tanto, la detección de eventos independientes de ligamiento con 20 o más fragmentos de restricción adyacentes indica fuertemente que el locus correspondiente contacta con la LCR de β-globina en múltiples células.

Para determinar la significancia estadística de estos grupos, los datos de experimentos individuales se ordenaron en mapas cromosómicos y se analizaron usando un algoritmo de media de ejecución con un tamaño de ventana de aproximadamente 60 kb. La distribución de la media de ejecución de datos aleatoriamente barajados se usó para establecer un valor umbral, que permite una tasa de descubrimiento de falsos del 5%. Este análisis identificó 66 grupos en hígado fetal y 45 en cerebro que se encontraron de forma reproducible en experimentos duplicados (Figura 13d-f). De hecho, FISH de alta resolución confirmó que dichos grupos realmente representan loci que interaccionan de forma frecuente (véase a continuación).

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

Se realizó una serie completamente independiente de experimentos 4C con un conjunto diferente de cebadores de PCR inversa que investigó el entorno genómico del gen β mayor, localizado ~50 kb cadena abajo de HS2. En hígado fetal, el gen β mayor se transcribe elevadamente y contacta frecuentemente por la LCR. Se encontraron grupos casi idénticos de interacciones de rango largo con β mayor como con HS2, ambos en hígado fetal y en cerebro, lo que sostiene adicionalmente que estos loci contactan frecuentemente con el locus de β-globina (Figura 17).

Ejemplo 5

El locus de β-globina activo e inactivo ocupa distinto entorno genómico

Una comparación entre los dos tejidos reveló que el locus de β-globina transcrito de forma activa en hígado fetal interacciona con un conjunto completamente diferente de loci que su equivalente transcripcionalmente silencioso en

cerebro (τ=-0,03; correlación de rango de Spearman) (Figura 13f). Esto excluyó que los resultados estuvieran influenciados por la composición de secuencia de las sondas. En hígado fetal, los segmentos de ADN de interacción estaban localizados en una región de 70 Mb centrada alrededor del locus de β-globina, con la mayoría (40/66) localizados hacia el telómero del cromosoma 7. En cerebro fetal, se encontraron loci de interacción a distancias similares o incluso mayores desde la β-globina en comparación con hígado fetal y con la gran mayoría de interacciones (43/45) localizadas hacia el centrómero del cromosoma 7. Estos datos demostraron que el locus de β-globina activo e inactivo contacta con diferentes partes del cromosoma 7.

Otros seis cromosomas (8, 10, 11, 12, 13 y 14) estaban representados en la microseries. Las señales fuertes de hibridación en estos cromosomas fueron raras, típicamente aparecían aisladas en el molde de ADN lineal y a menudo estaban ausentes de experimentos duplicados. Además, los niveles de media de ejecución en todos estos cromosomas nunca llegaron de forma reproducible a los niveles valorados para el cromosoma 7 (Figura 19). Por tanto, nuestros datos mostraron que el locus de β-globina contactaba principalmente con loci en otra parte en el mismo cromosoma, en acuerdo con la localización preferida de este locus dentro de su propio territorio de cromosoma. Observamos que el locus de α-globina también estaba presente en la serie (cromosoma 11) y no valoraba positivo para la interacción con β-globina, en acuerdo con la reciente demostración por FISH de que la α- y β-globina de ratón no coinciden frecuentemente en el espacio nuclear (Brown, J. M. et al. (2006) J Cell Biol 172, 177-87).

Para comprender mejor la relevancia de las interacciones observadas de rango largo en el cromosoma 7, se compararon los loci de interacción con las posiciones cromosómicas de los genes. Además, se realizó análisis en serie 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 y actividad génica. En hígado fetal, el 80% de los loci de interacción con β-globina contenían uno o más genes transcritos de forma activa, mientras que en cerebro fetal la gran mayoría (87%) no mostraba actividad génica detectable (Figura 15). Por tanto, el locus de β-globina está integrado en un entorno genómico muy diferente en los dos tejidos. En cerebro, donde el locus no es activo, principalmente contacta con loci transcripcionalmente silenciosos hacia el centrómero del cromosoma 7. En hígado fetal, donde el locus es altamente activo, interacciona preferentemente con regiones transcritas activamente localizadas más prominentemente hacia el lado telomérico del cromosoma 7. De forma importante, la tecnología 4C identificó tanto Uros como Eraf (~30 Mb desde la β-globina) como genes que interaccionan con el locus de β-globina activa en hígado fetal, en acuerdo con las observaciones previas hechas por FISH (Osborne, C. S. et al. (2004) Nat Genet 36, 1065-71 (2004)). De forma interesante, en cerebro se observaron contactos con los otros dos grupos de genes de receptores olfativos presentes en el cromosoma 7 que estaban localizados en cada lado de, y 17 y 37 Mb desde, la β-globina.

No todas las regiones transcritas en el cromosoma 7 interaccionan con el locus de β-globina activa en hígado fetal. Por lo tanto, se buscó un denominador compartido exclusivamente por los loci de interacción pero no por otras regiones activas en hígado fetal. Los genes de β-globina, Uros y Eraf son todos genes específicos 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 serie de expresión Affymetrix de hígado fetal E14.5 con los de cerebro fetal para identificar genes expresados de forma preferente (>5 veces más) en hígado fetal. Y así, el 28% de los genes activos en el cromosoma 7 se clasificaron como "específicos de hígado fetal", de los cuales el 25% se encontraron en un área de co-localización. Por tanto, se encontró ausencia de enriquecimiento de genes "específicos de hígado fetal" en las áreas de co-localización. De forma más importante, 49 de 66 (74%) regiones de interacción no contenían "específico de hígado fetal" y por lo tanto se concluye que nuestros datos no mostraban evidencias para coordinar la expresión de genes específicos de tejido en el espacio nuclear. Los genes de β-globina se transcriben a tasas altas excepcionales y la siguiente pregunta fue si el locus interaccionaba preferentemente con otras regiones de alta actividad transcripcional, que son genes o áreas altamente expresadas con una alta densidad de genes activos. Usando los recuentos Affymetrix como medida para la actividad génica, se realizó un algoritmo de suma de ejecución para medir la actividad transcripcional global en regiones de 200 kb alrededor de genes transcritos de forma activa. Este análisis reveló que la actividad transcripcional alrededor de genes de interacción no era mayor que alrededor de genes activos no de interacción en el cromosoma 7 ($p = 0,9867$; suma de rango de Wilcoxon).

Ejemplo 6

El entorno genómico de un gen constitutivo está conservado en gran medida entre los tejidos

A continuación se investigó si un gen que se expresa de forma similar en ambos tejidos también cambiar su entorno genómico. Rad23A es un gen expresado de forma ubicua que reside en un grupo denso en genes de genes principalmente constitutivos en el cromosoma 8. Tanto en hígado fetal como en cerebro E14.5, este gen y muchos de sus vecinos directos son activos. Se realizó análisis 4C e identificó muchas interacciones de rango largo con loci hasta 70 Mb desde Rad23A. De forma importante, las interacciones con Rad23A estuvieron altamente correlacionadas entre hígado fetal y cerebro (τ=0,73; correlación de rango de Spearman) (Figura 15a). Un distintivo

compartido de estos loci fue de nuevo que contenían genes activamente transcritos. Por tanto, en ambos tejidos aproximadamente el 70% contenía al menos un gen activo (Figura 15b-c). Las regiones alrededor de los genes de interacción presentaban niveles mayores estadísticamente significativos de actividad génica en comparación con genes activos en otra parte en el cromosoma, determinado por un algoritmo de suma de ejecución ($p < 0,001$ para ambos tejidos). Por tanto, a diferencia del locus de β-globina, el gen Rad23A que está localizado en una región rica en genes interacciona preferentemente en 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 intra-cromosómicas y ningún área en el cromosoma 7, 10, 11, 12, 13 o 14 cumplía de forma reproducible nuestros criterios rigurosos para interacción. Por tanto, Rad23A está principalmente implicado en interacciones intra-cromosómicas que son similares en dos tejidos muy diferentes. Si Rad23A tiene loci adyacentes preferidos en estos cromosomas no relacionados, no interaccionan de forma suficientemente frecuente para detectarse en las condiciones usadas aquí para tecnología 4C.

Ejemplo 7

Validación de tecnología 4C, por microscopía de alta resolución

Para validar los resultados obtenidos por tecnología 4C, se realizaron experimentos de crio-FISH. Crio-FISH es una técnica de microscopía recientemente desarrollada, que tiene la ventaja sobre los actuales protocolos 3D-FISH de que conserva mejor la ultra-estructura nuclear ofreciendo al mismo tiempo resolución mejorada en el eje z mediante la preparación de crio-secciones ultra-finas (Branco, M. R. y Pombo, A (2006). PLoS Biol 4, e138). Los datos 4C se verificaron midiendo cómo de frecuente los alelos de β-globina o Rad23A (siempre $n>250$) se co-localizan con más de 15 regiones cromosómicas seleccionadas en secciones ultra-finas de 200 nm preparadas a partir de hígado y cerebro E14.5. De forma importante, todas las frecuencias de interacción medidas por crio-FISH estaban en perfecto acuerdo con los resultados 4C (Figura 17). Por ejemplo, regiones distantes que se identificó que interaccionaban con β-globina por tecnología 4C se co-localizaron más frecuentemente que áreas intermedias no detectadas por 4C (7,4% y 9,7%, frente a 3,6% y 3,5%, respectivamente). Además, los dos grupos génicos de receptores olfativos distantes identificados por tecnología 4C que interaccionaban con β-globina en cerebro fetal pero no en hígado dieron valores de frecuencias de co-localización respectivamente del 12,9% y el 7% en cerebro, frente al 3,6% y el 1,9% en secciones hepáticas. En resumen, las frecuencias de co-localización medidas para loci identificados positivamente por tecnología 4C fueron todas significativamente mayores que las frecuencias medidas para loci de fondo ($p<0,05$; ensayo G). Se concluyó que la tecnología 4C identificó fielmente loci de ADN de interacción. Finalmente, se usó crio-FISH para demostrar que loci identificados que interaccionaban con β-globina también contactaban frecuentemente entre sí. Esto fue cierto para dos regiones activas separadas sobre una gran distancia cromosómica en hígado fetal (Figura 19) así como para dos grupos génicos OR inactivos muy alejados en el cromosoma en cerebro (Figura 17). De forma interesante, también se encontraron contactos frecuentes entre estos dos grupos génicos OR distantes en hígado fetal, donde no interaccionaban con el grupo génico OR que contenía el locus de β-globina activamente transcrito. Estos datos indicaron que interacciones nucleares entre distintos grupos génicos OR no eran una peculiaridad del tejido de cerebro fetal analizado. Resulta tentador especular que dichos contactos espaciales facilitan la comunicación entre los muchos genes OR necesarios para asegurar que solamente se transcriba un único alelo por neurona olfativa (Shykind, B. (2005) Hum Mol Genet 14 Spec No 1, R33-9.

Ejemplo 8

Organización nuclear de dominios de cromatina activa e inactiva

Las observaciones descritas en este documento demuestran que no solamente regiones genómicas activas, sino también inactivas forman distintas regiones en el espacio nuclear que implican muchos contactos de rango largo, lo que sugiere fuertemente que cada segmento de ADN tiene su propio conjunto preferido de interacciones. Nuestros datos sugieren que cuando el locus de β-globina se activa, abandona un entorno genómico de silencio transcripcional y entre en un área nuclear donde están favorecidas las interacciones con dominios activos. Se anticipa que dicho drástico reposicionamiento tras la activación transcripcional bien puede ser un distintivo solamente de genes específicos de tejido que alcanzan un cierto nivel de expresión y, de forma más importante, están aislados de otros genes activos en el molde de cromosoma lineal, como es el caso para β-globina. Se propone que la red extensiva de interacciones de rango largo que se identifican tanto entre loci genómicos inactivos como entre loci genómicos activos, refleja las diferencias entre células en las conformaciones cromosómicas más que ser una consecuencia de los movimientos dinámicos durante la interfase (Chakalova et al. (2005) Nat Rev Genet 6, 669-77 (2005). Presumiblemente, diferentes grados de des-condensación después de la división celular dirigen las regiones genómicas activas desde la cromatina inactiva (Gilbert, N. et al. (2004) Cell 118, 555-66 (2004)) y los contactos entre loci distantes de composición similar de cromatina se establecer principalmente a través de afinidades entre proteínas unidas a la cromatina. La yuxtaposición espacial entre loci distantes puede ser funcional, pero también puede ser simplemente la consecuencia de los patrones de desplegamiento de un cromosoma. Aunque los loci individuales puede moverse dentro de un volumen nuclear restringido, la conformación general de un cromosoma se mantendría en gran medida en todo el ciclo celular y requiriendo división celular para el

restablecimiento. Esta idea está en acuerdo con los estudios de imágenes de células vivas que muestran el movimiento restringido de loci de ADN marcados en el interior nuclear (Chubb et al. (2002) Curr Biol 12, 439-45 (2002)) y se ajusta bien con los estudios que muestran que la información posicional de cromatina nuclear se propaga frecuentemente 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)).

TABLA 2

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

Referencias

Blanton J, Gaszner M, Schedl P. 2003. Protein:protein interactions and the pairing of boundary elements in vivo. Genes Dev 17:664-75.

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.

Murrell A, Heeson S, Reik W. 2004. Interaction between differentially metilated regions partitions the imprinted genes Igf2 and H19 in 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.

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.

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 hipersensitive sites in the active beta-globin locus. Molecular Cell 10: 1453-65.

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.

Aunque la invención se ha descrito en conexión con realizaciones preferidas específicas, debe entenderse que la invención reivindicada no debe limitarse indebidamente a dichas realizaciones específicas. De hecho, se pretende

que diversas modificaciones de los modos descritos para realizar la invención que son obvios para los expertos en biología molecular o campos relacionados estén dentro del alcance de las siguientes reivindicaciones.

REIVINDICACIONES

1. Un método para analizar la frecuencia de interacción de una secuencia nucleotídica diana con una o más secuencias nucleotídicas de interés (por ejemplo, uno o más loci 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 nucleotídicas entrecruzadas;
(d) revertir el entrecruzamiento;
(e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria;
(f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios disponibles de digestión por enzima de restricción secundaria que flanquean la una o más secuencias nucleotídicas de interés;
(g) amplificar la una o más secuencias nucleotídicas de interés usando al menos dos cebadores oligonucleotídicos, en el que cada cebador hibrida con las secuencias de ADN que flanquean las secuencias nucleotídicas de interés;
(h) hibridar la secuencias o secuencias amplificadas en una serie; y
(i) determinar la frecuencia de interacción entre las secuencias de ADN.

2. El método de acuerdo con la reivindicación 1, en el que la reacción de ligamiento en la etapa (f) provoca la formación de círculos de ADN.

3. El método de acuerdo con la reivindicación 1 o reivindicación 2, en el que la secuencia nucleotídica diana se selecciona entre el grupo que consiste en un reordenamiento genómico, promotor, un potenciador, un silenciador, un aislante, una región de adhesión a matriz, una región de control de locus, una unidad de transcripción, un origen de replicación, un punto caliente de recombinación, un punto de interrupción de translocación, un centrómero, un telómero, una región densa en genes, una región pobre en genes, un elemento repetitivo y un sitio de integración (viral).

4. El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la secuencia nucleotídica diana es una secuencia nucleotídica que está asociada con o causa una enfermedad, o está localizada hasta o más de 15 Mb en un molde de ADN lineal desde un locus que está asociado con o causa una enfermedad.

5. El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la secuencia nucleotídica diana se selecciona entre el grupo que consiste en: *AML1, MLL, MYC, BCL, BCR, ABL1*, *IGH, LYL1*, *TAL1, TAL2, LMO2, TCR*α/δ*, TCR*β y *HOX* u otros loci asociados con enfermedad.

6. El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la enzima de restricción primaria es una enzima de restricción que reconoce un sitio de reconocimiento de 6-8 pb.

7. El método de acuerdo con la reivindicación 6, en el que la enzima de restricción primaria se selecciona entre el grupo que consiste en *Bgl*II, *Hind*III, *Eco*RI, *Bam*HI, *Spe*I, *Pst*I y *Nde*I.

8. El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la enzima de restricción secundaria es una enzima de restricción que reconoce un sitio de reconocimiento de secuencia nucleotídica de 4 o 5 pb.

9. El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que el sitio de reconocimiento por enzima de restricción secundaria está localizado a más de aproximadamente 350 pb desde el sitio de restricción primario en la secuencia nucleotídica diana.

10. El método de acuerdo con cualquiera de las reivindicaciones precedentes, en el que la secuencia nucleotídica está marcada.

11. Un método para analizar la frecuencia de interacción de una secuencia nucleotídica diana con una o más secuencias nucleotídicas (por ejemplo, uno o más loci 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 nucleotídicas entrecruzadas;
(d) revertir el entrecruzamiento;
(e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria;
(f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios disponibles de digestión por enzima de restricción secundaria que flanquean la una o más secuencias nucleotídicas de interés que circulariza las secuencias nucleotídicas;

(g) amplificar la una o más secuencias nucleotídicas que están ligadas a la secuencia nucleotídica diana usando al menos dos cebadores oligonucleotídicos, en el que cada cebador hibrida con las secuencias de ADN que flanquean las secuencias nucleotídicas de interés;
(h) opcionalmente hibridar las secuencias amplificadas en una serie; y
(i) determinar la frecuencia de interacción entre las secuencias de ADN.

12. Un método para preparar una secuencia nucleotídica circularizada 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 nucleotídicas entrecruzadas;
(d) revertir el entrecruzamiento;
(e) digerir las secuencias nucleotídicas con una enzima de restricción secundaria; y
(f) ligar una o más secuencias de ADN de composición nucleotídica conocida con el sitio o sitios disponibles de digestión por enzima de restricción secundaria que flanquean la una o más secuencias nucleotídicas de interés que circulariza las secuencias nucleotídicas.

13. 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 1-11, en el que en la etapa (a) se proporciona una muestra de ADN entrecruzado de células enfermas y no enfermas, y en el que una diferencia entre la frecuencia de interacción entre las secuencias de ADN de las células enfermas y no enfermas indica que la interacción ADN-ADN es indicativa de una patología particular.

14. Un método de diagnóstico de una enfermedad o síndrome causado por o asociado con un cambio en las interacciones ADN-ADN que comprende la etapa de realizar las etapas (a)-(i) de cualquiera de las reivindicaciones 1-11, 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 está padeciendo la enfermedad o síndrome o es indicativa de que el sujeto padecerá la enfermedad o síndrome.

15. El método de acuerdo con la reivindicación 14, en el que una transición de frecuencias de interacción bajas a altas para una región cromosómica es indicativa de la localización de un punto de interrupción.

16. El método de acuerdo con 14, en el que un patrón invertido de frecuencias de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una inversión.

17. El método de acuerdo con la reivindicación 14, en el que una reducción en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control, en combinación con un aumento en la frecuencia de interacción ADN-ADN para regiones cromosómicas más distantes, es indicativa de una deleción.

18. El método de acuerdo con la reivindicación 14, en el que un aumento en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una duplicación o inserción.

19. El método de la reivindicación 18, en el que existe una disminución en las frecuencias de interacción ADN-ADN para una región cromosómica entre regiones más distantes de ADN en la muestra del sujeto en comparación con el control.

20. Un método de acuerdo con cualquiera de las reivindicaciones 14-19, en el que se usa cariotipado espectral y/o FISH antes de realizar dicho método.

21. El método de acuerdo con cualquiera de las reivindicaciones 14-20, en el que la enfermedad es una enfermedad genética.

22. El método de acuerdo con cualquiera de las reivindicaciones 14-21, en el que la enfermedad es cáncer.

23. Un método de diagnóstico de una enfermedad o síndrome causado por o asociado con un cambio en las interacciones ADN-ADN que comprende las etapas de: realizar las etapas (a)-(i) de cualquiera de las reivindicaciones 1-11, 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 loci que han experimentado un reordenamiento genómico que está asociado con una enfermedad.

24. 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) de cualquiera de las reivindicaciones 1-11, 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.

25. Un método para detectar la localización de un punto de interrupción equilibrado y/o desequilibrado (por ejemplo, una translocación) que comprende la etapa de:

(a) realizar las etapas (a) a (i) de cualquiera de las reivindicaciones 1-11; y
(b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control;

en el que una transición de frecuencia de interacción ADN-ADN baja a alta para una región cromosómica en la muestra en comparación con el control es indicativa de la localización de un punto de interrupción.

26. Un método para detectar la localización de una inversión equilibrada y/o desequilibrada que comprende las etapas de:

(a) realizar las etapas (a) a (i) de cualquiera de las reivindicaciones 1-11; 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 una región cromosómica para la muestra en comparación con el control es indicativo de una inversión.

27. Un método para detectar la localización de una deleción que comprende las etapas de:

(a) realizar las etapas (a) a (i) de cualquiera de las reivindicaciones 1-11; 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 una región cromosómica para la muestra en comparación con el control es indicativa de deleción.

28. Un método para detectar la localización de una duplicación o inserción que comprende las etapas de:

(a) realizar las etapas (a) a (i) de cualquiera de las reivindicaciones 1-11; y
(b) comparar la frecuencia de interacción entre las secuencias de ADN con la de un control;

en el que un aumento en la frecuencia de interacción ADN-ADN para una región cromosómica para la muestra del sujeto en comparación con el control es indicativo de una duplicación o inserción.

Figura 1

1.
Fijación con formaldehído de células
Los sitios genómicos cerca en el espacio nuclear se entrecruzan
2.
Digestión con enzima de restricción
digestión de cromatina entrecruzada en el núcleo
3.
Ligamiento intramolecular
Ligamiento de fragmentos de ADN entrecruzado
4.
Reversión de entrecruzamientos/ purificación del ADN
5.
Cuantificación de los productos de ligamiento por PCR

Figura 2

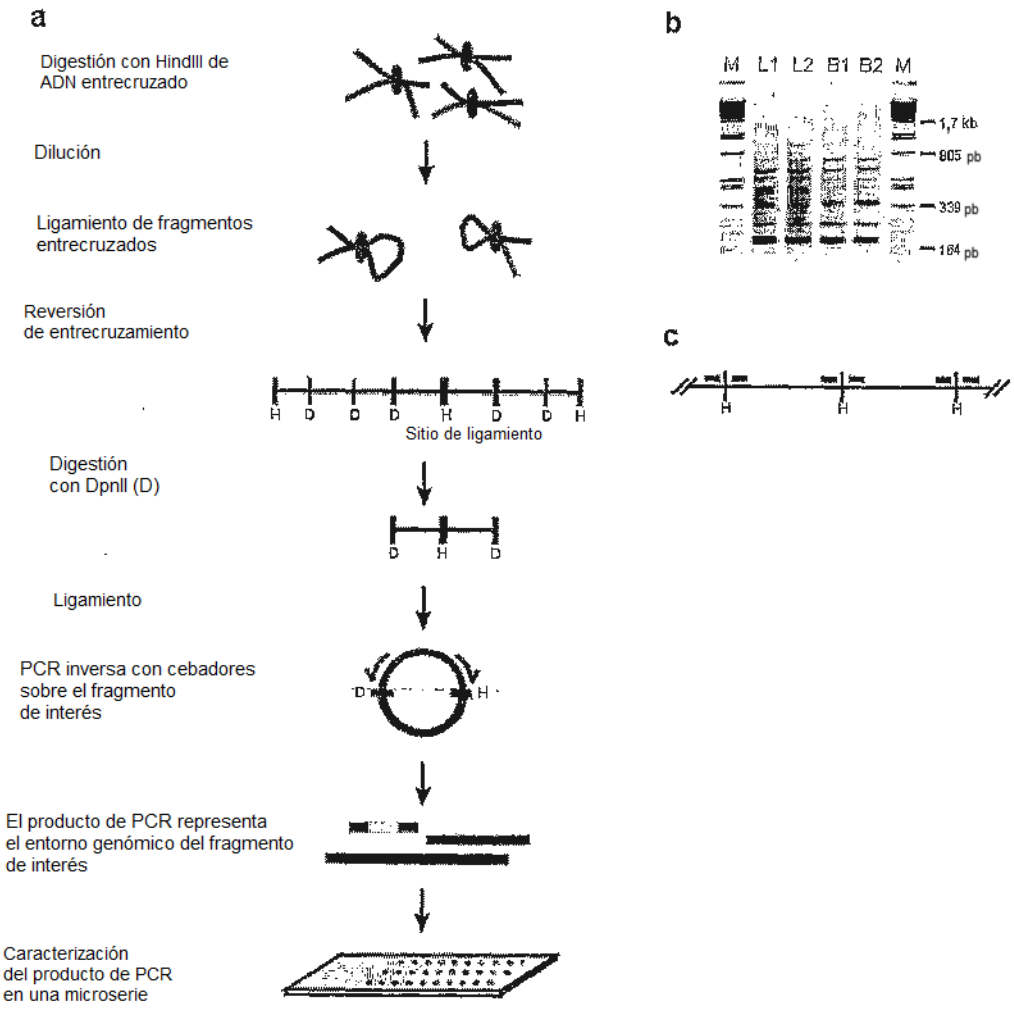

a
Digestión con HindIII de ADN entrecruzado
Dilución
Ligamiento de fragmentos entrecruzados
Reversión de entrecruzamiento
H D D D H D D H
Sitio de ligamiento
Digestión con DpnII (D)
D H D
Ligamiento
PCR inversa con cebadores sobre el fragmento de interés
El producto de PCR representa el entorno genómico del fragmento de interés
Caracterización del producto de PCR en una microserie
b
M L1 L2 B1 B2 M
1,7 kb
805 pb
339 pb
164 pb
c
H H H

Figura 3

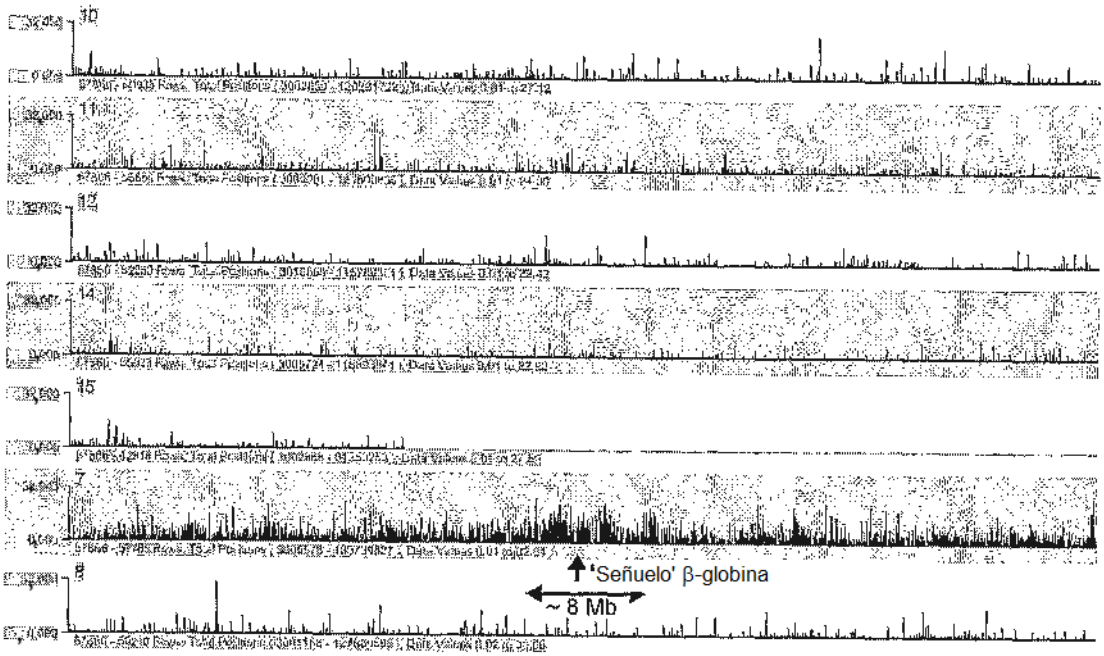


Figura 4

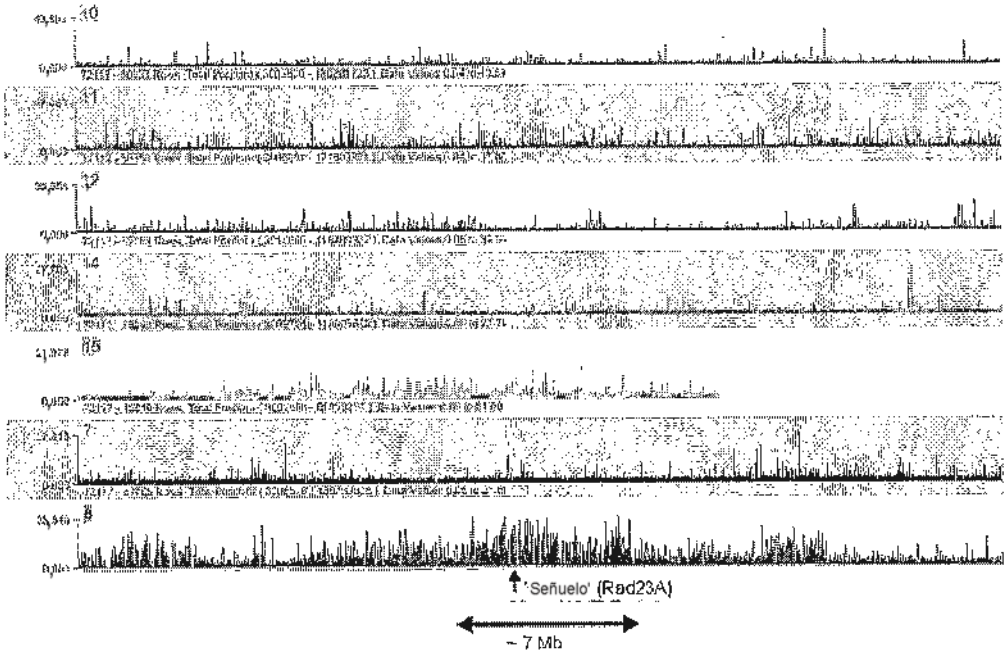


Figura 5

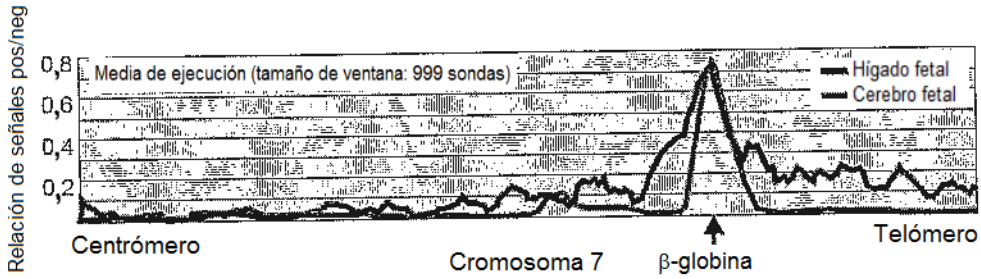

Relación de señales pos/neg
0,8
0,6
0,4
0,2
Media de ejecución (tamaño de ventana: 999 sondas)
Hígado fetal
Cerebro fetal
Centrómero
Cromosoma 7
β-globina
Telómero

Figura 6

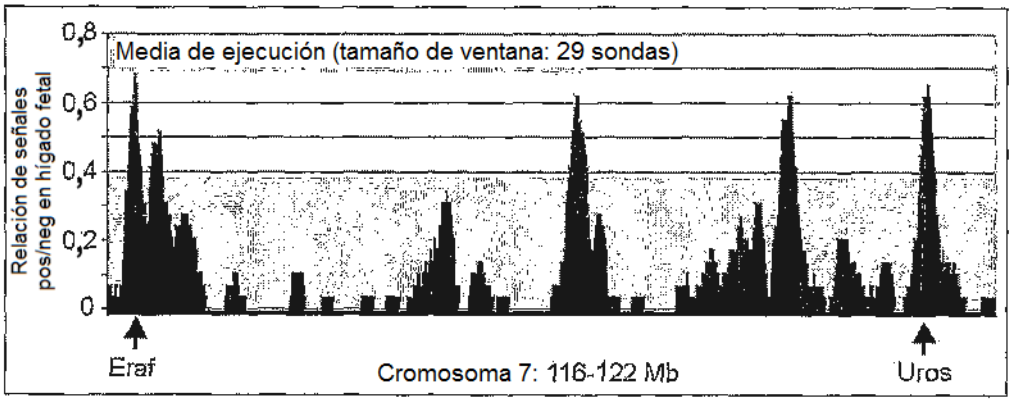

Media de ejecución (tamaño de ventana: 29 sondas)
Relación de señales pos/neg en hígado fetal
0,8
0,6
0,4
0,2
0
Eraf
Cromosoma 7: 116-122 Mb
Uros

Figura 7

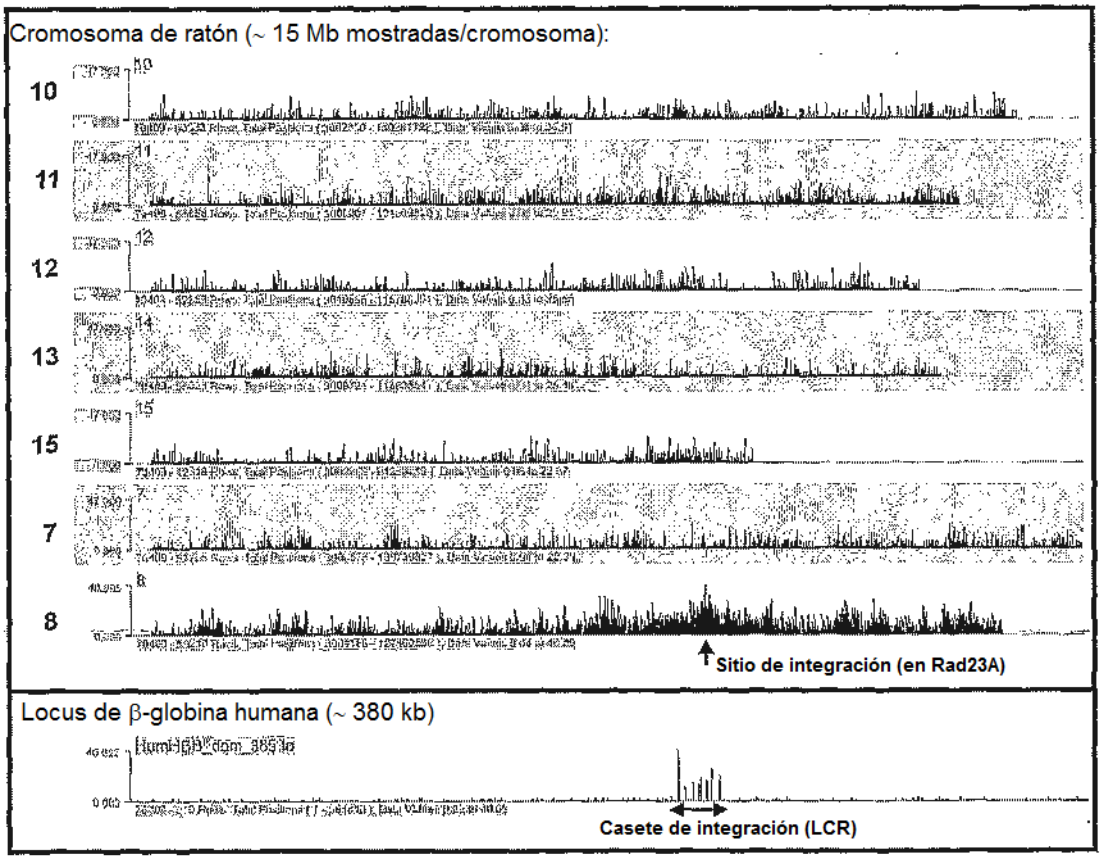

Cromosoma de ratón (~ 15 Mb mostradas/cromosoma):
10
11
12
13
15
7
8
Sitio de integración (en Rad23A)
Locus de β-globina humana (~ 380 kb)
Casete de integración (LCR)

Figura 8

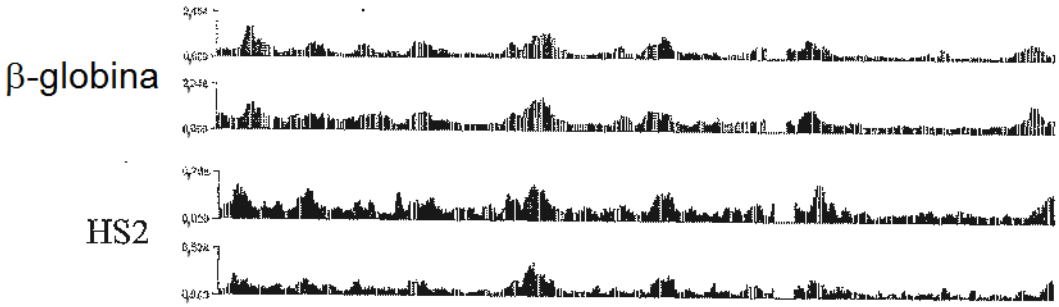


Figura 9

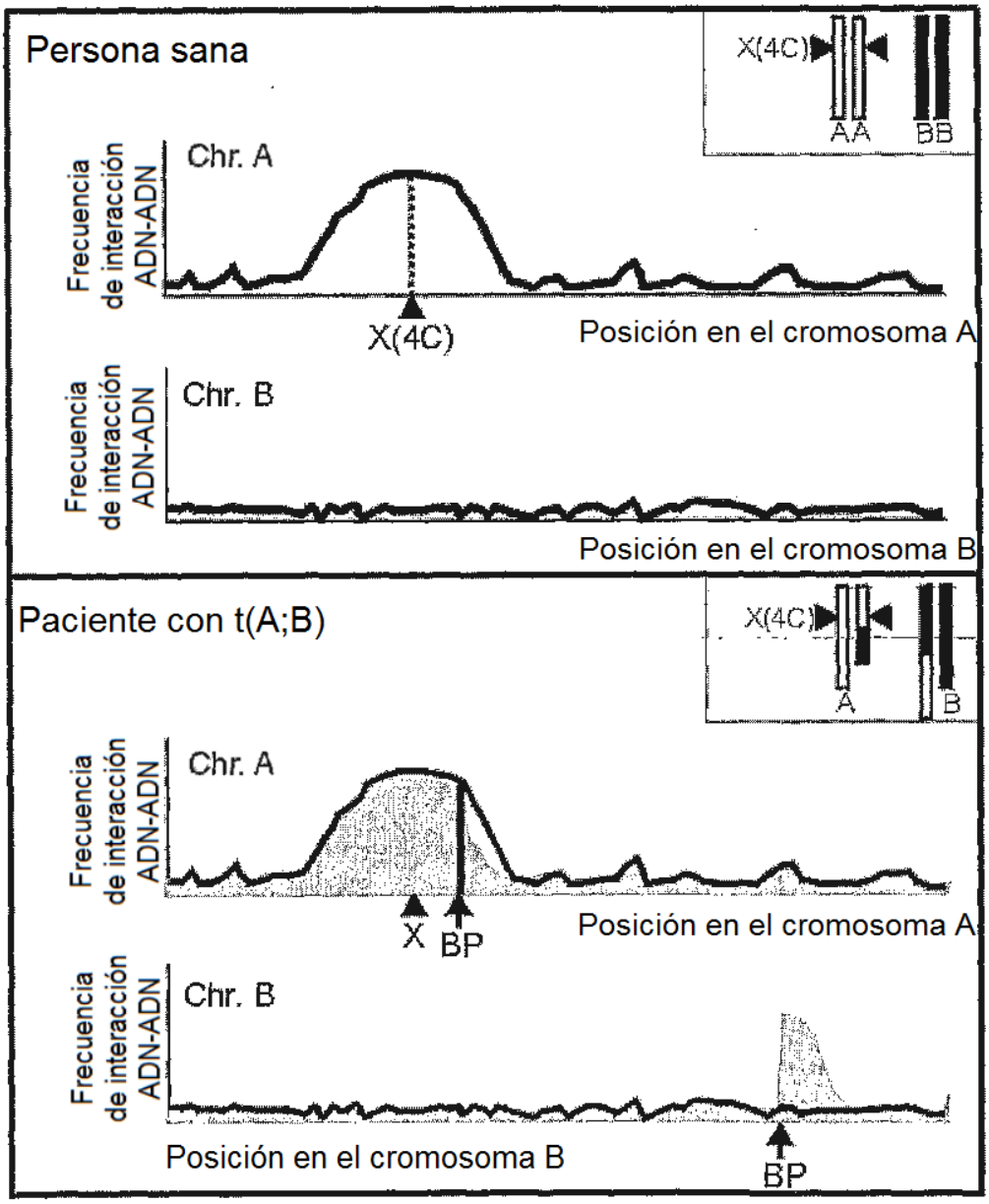

Persona sana
X(4C)
AA
BB
Chr. A
Frecuencia de interacción ADN-ADN
X(4C)
Posición en el cromosoma A
Chr. B
Frecuencia de interacción ADN-ADN
Posición en el cromosoma B
Paciente con t(A;B)
X(4C)
A
B
Chr. A
Frecuencia de interacción ADN-ADN
X
BP
Posición en el cromosoma A
Chr. B
Frecuencia de interacción ADN-ADN
Posición en el cromosoma B
BP

FIGURA 10

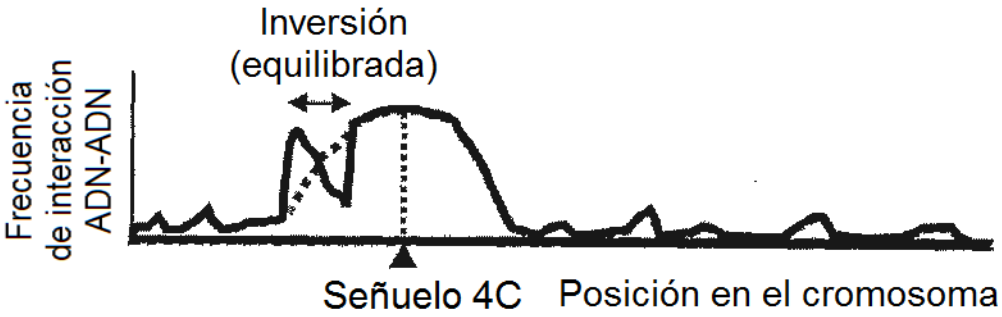

Inversión
(equilibrada)
Frecuencia
de interacción
ADN-ADN
Señuelo 4C
Posición en el cromosoma

FIGURA 11

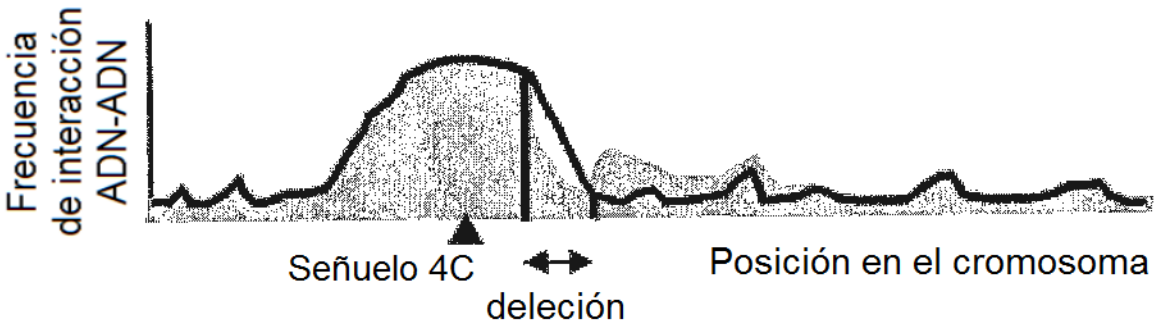

Frecuencia de interacción ADN-ADN
Señuelo 4C
deleción
Posición en el cromosoma

FIGURA 12

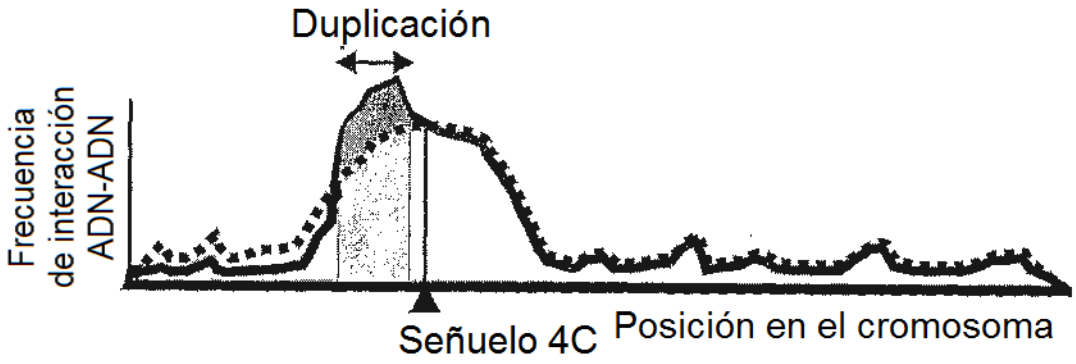

Duplicación
Frecuencia de interacción ADN-ADN
Señuelo 4C
Posición en el cromosoma

FIGURA 13

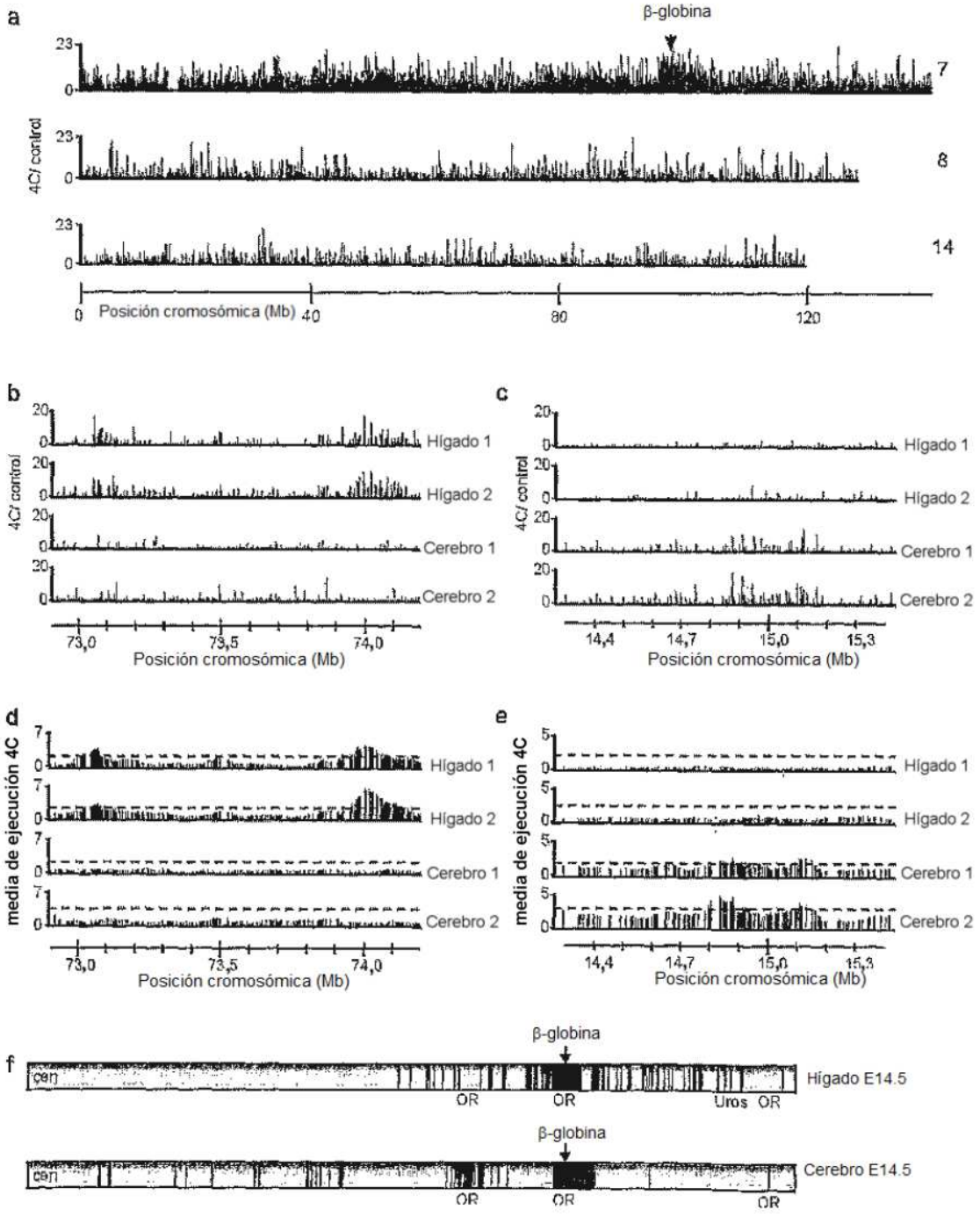

a
β-globina
4C/ control
7
8
14
Posición cromosómica (Mb)
b
c
Hígado 1
Hígado 2
Cerebro 1
Cerebro 2
d
e
media de ejecución 4C
f
cen
OR
Uros OR
Hígado E14.5
Cerebro E14.5

FIGURA 14

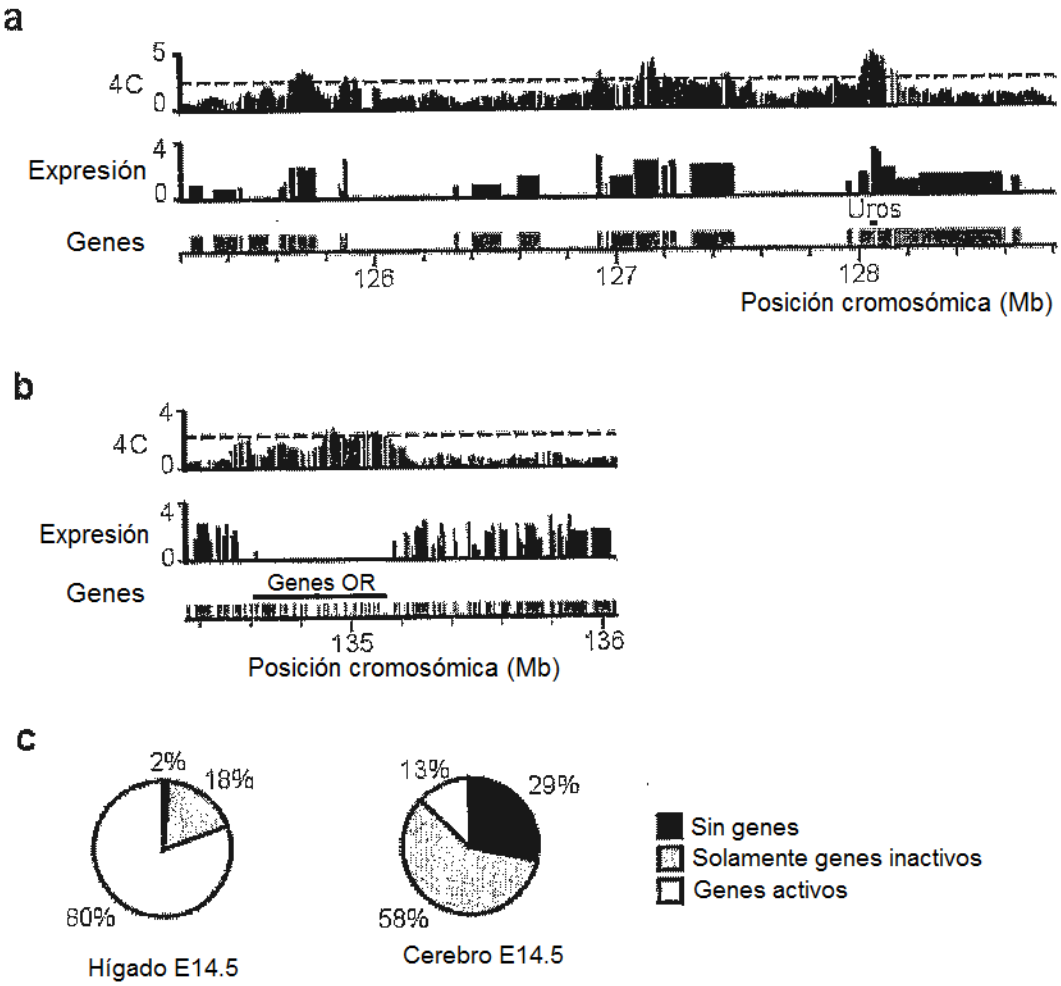

a
4C
5
0
Expresión
4
0
Genes
Urod
126
127
128
Posición cromosómica (Mb)
b
4C
4
0
Expresión
4
0
Genes
Genes OR
135
136
Posición cromosómica (Mb)
c
2%
18%
80%
Hígado E14.5
13%
29%
58%
Cerebro E14.5
Sin genes
Solamente genes inactivos
Genes activos

FIGURA 15

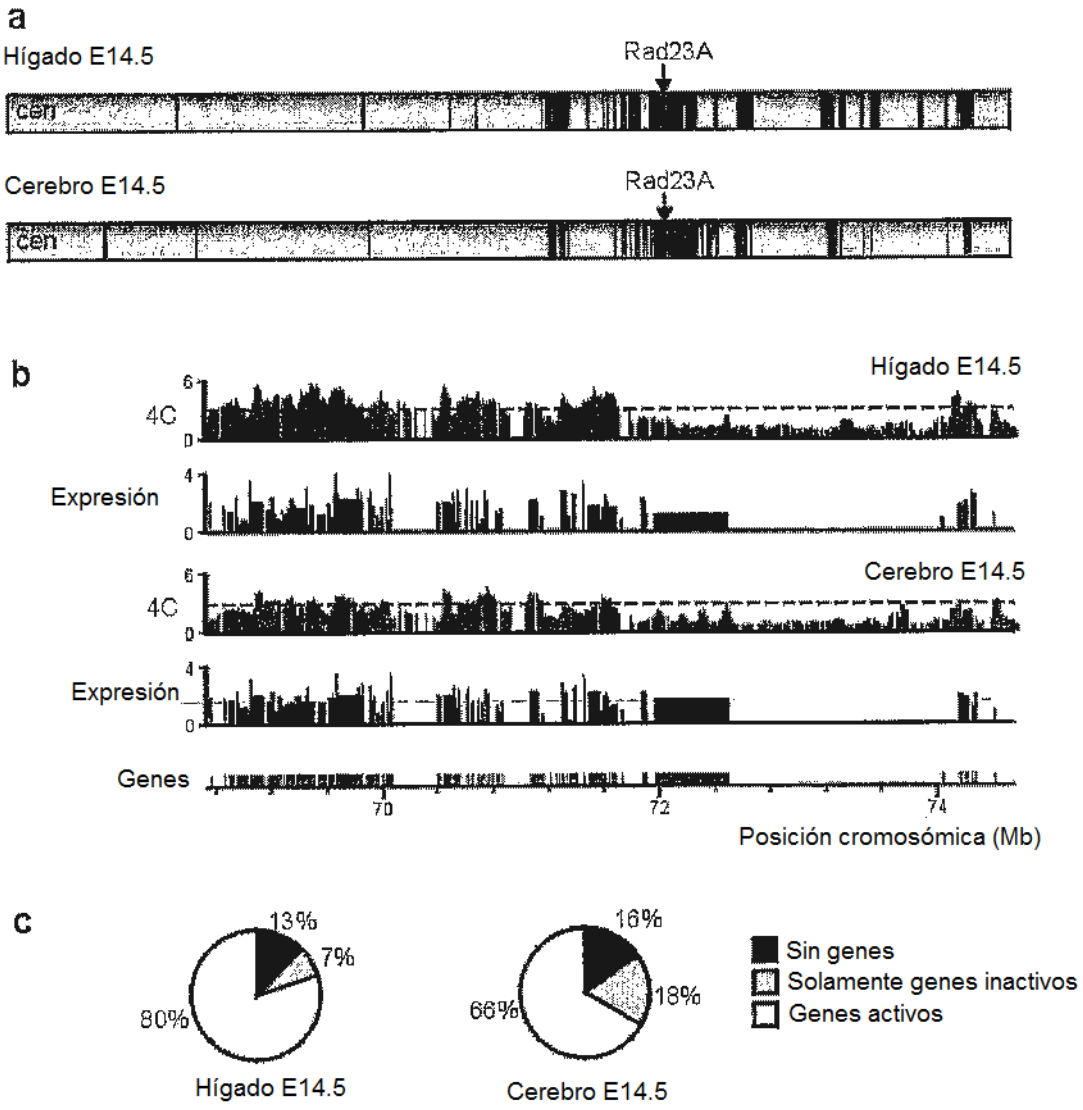

a
Hígado E14.5
Rad23A
cen
Cerebro E14.5
Rad23A
cen
b
Hígado E14.5
4C
Expresión
Cerebro E14.5
4C
Expresión
Genes
70
72
74
Posición cromosómica (Mb)
c
13%
7%
80%
Hígado E14.5
16%
18%
66%
Cerebro E14.5
Sin genes
Solamente genes inactivos
Genes activos

FIGURA 16

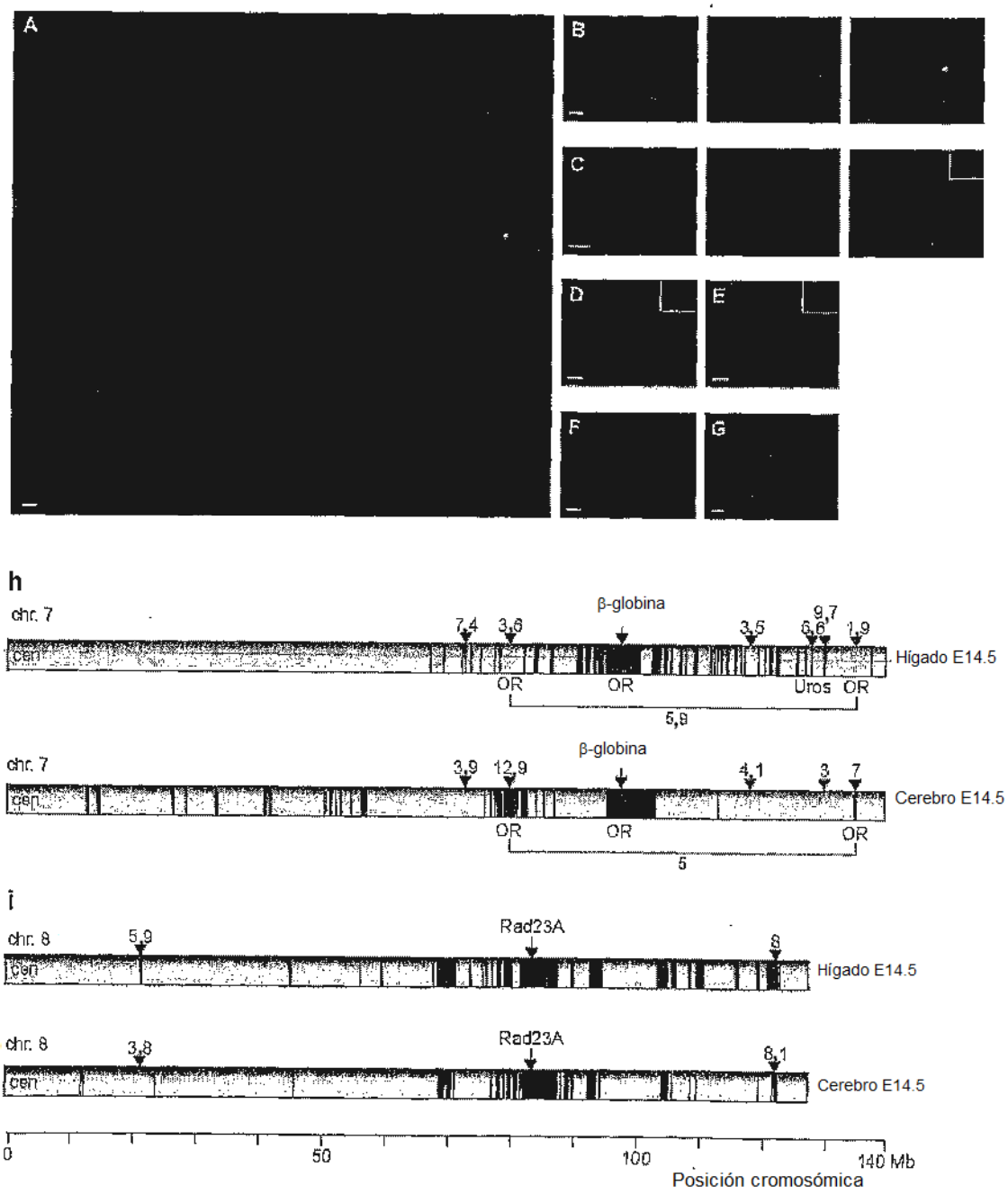
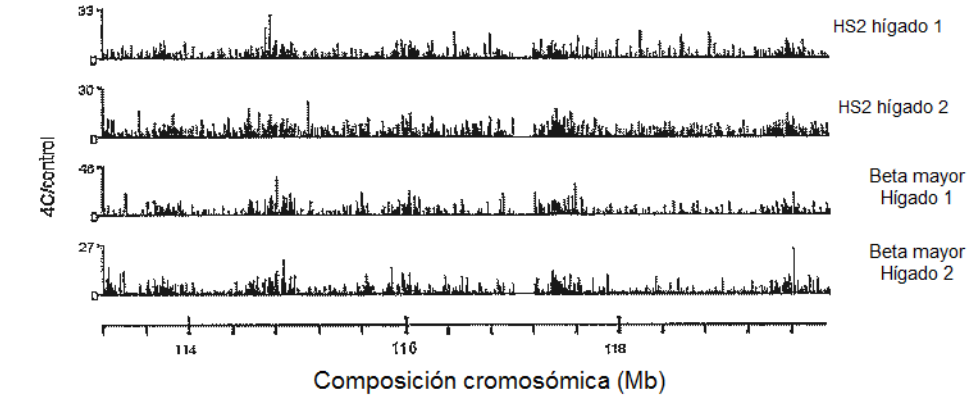

A
B
C
D
E
F
G
h
chr. 7
cen
7,4
3,6
β-globina
3,5
6,6
9,7
1,9
Hígado E14.5
OR
OR
Uros
OR
5,9
chr. 7
cen
3,9
12,9
β-globina
4,1
3
7
Cerebro E14.5
OR
OR
OR
5
i
chr. 8
cen
5,9
Rad23A
8
Hígado E14.5
chr. 8
cen
3,8
Rad23A
8,1
Cerebro E14.5
0
50
100
140 Mb
Posición cromosómica

FIGURA 17

a


33
30
46
27
4C/control
HS2 hígado 1
HS2 hígado 2
Beta mayor Hígado 1
Beta mayor Hígado 2
114
116
118
Composición cromosómica (Mb)

b

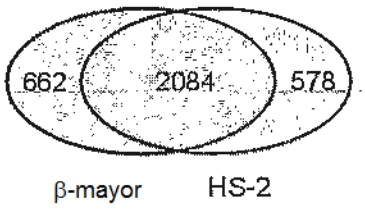
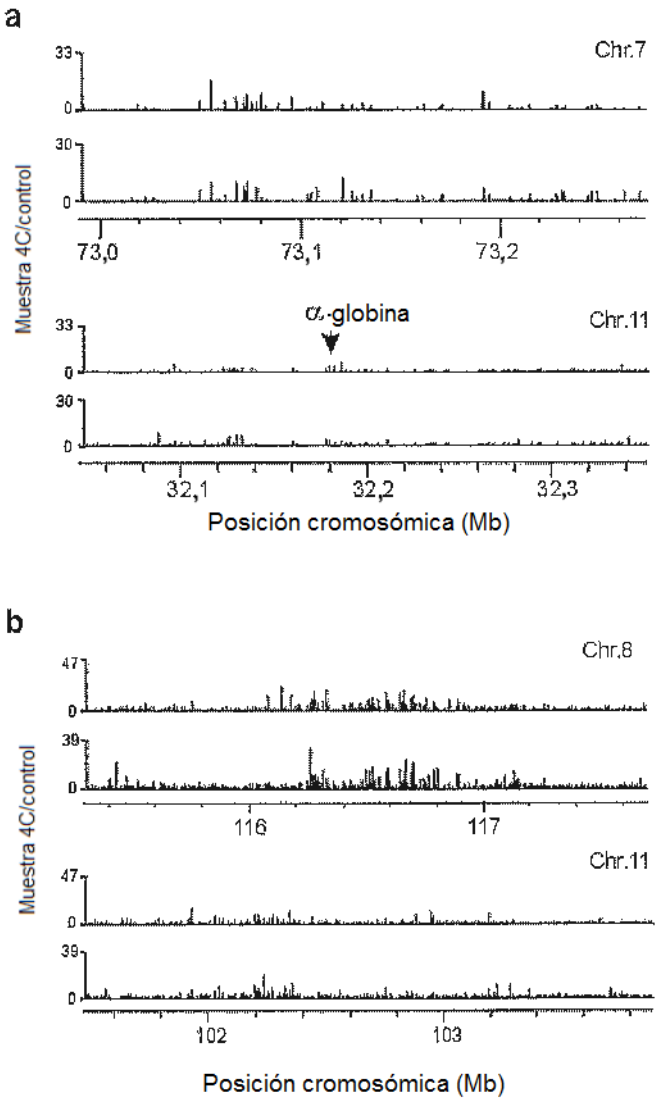

662
2084
578
β-mayor
HS-2

FIGURA 18


a
Chr.7
33
0
30
0
Muestra 4C/control
73,0
73,1
73,2
α-globina
Chr.11
33
0
30
0
32,1
32,2
32,3
Posición cromosómica (Mb)
b
Chr.8
47
0
39
0
Muestra 4C/control
116
117
Chr.11
47
0
39
0
102
103
Posición cromosómica (Mb)

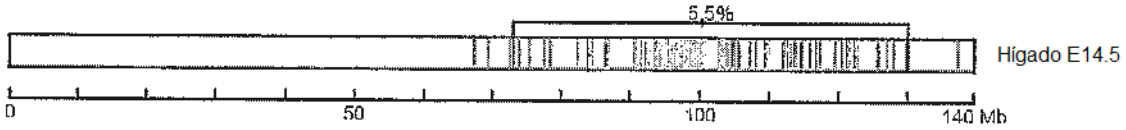



FIGURA 19