

República Federativa do Brasil
Ministério do Desenvolvimento, Indústria
e do Comércio Exterior
Instituto Nacional da Propriedade Industrial.

(21) **PI0611702-3 A2**

(22) Data de Depósito: 03/07/2006
(43) Data da Publicação: 28/09/2010
(RPI 2073)



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

(54) Título: **CAPTURE E CARACTERIZAÇÃO DE CROMATINA CO-LOCALIZADA (4C)**

(30) Prioridade Unionista: 04/07/2005 GB 0513676.7,
17/03/2006 GB 0605449.8

(73) Titular(es): ERASMUS UNIVERSITY MEDICAL CENTER

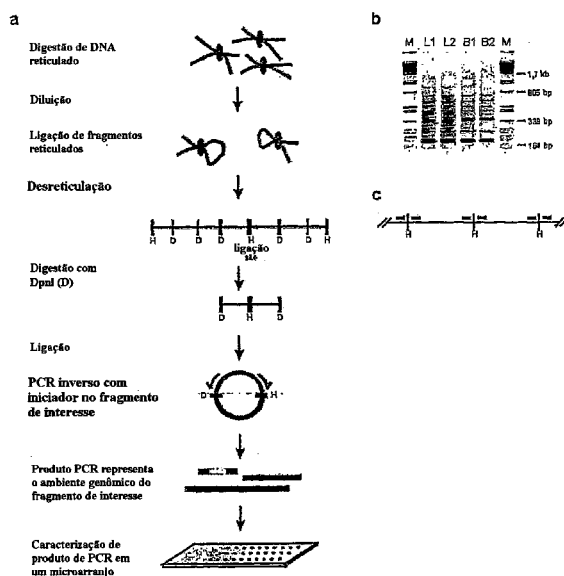
(72) Inventor(es): FRANK GROSVELD, WOUTER DE LAAT

(74) Procurador(es): RICARDO PINHO

(86) Pedido Internacional: PCT IB2006002268 de 03/07/2006

(87) Publicação Internacional: WO 2007/004057 de 11/01/2007

(57) Resumo: CAPTURE E CARACTERIZAÇÃO DE CROMATINA COLOCALIZADA (4C). A presente invenção diz respeito em um aspecto a um método para analisar a frequência de interação de uma sequência de nucleotídeos alvo com uma ou mais sequências de nucleotídeos de interesse (por exemplo, um ou mais loci genômicos) que compreende as etapas de: (a) fornecer uma amostra de DNA reticulado; (b) digerir o DNA reticulado com uma enzima de restrição primária; (c) ligar as sequências de nucleotídeos reticuladas; (d) reverter a reticulação; (e) digerir as sequências de nucleotídeos com uma enzima de restrição secundária; (f) ligar uma ou mais sequências de DNA de composição de nucleotídeo conhecida ao (s) sítio (s) de digestão de enzima de restrição secundária disponível que flanqueiam uma ou mais sequências de nucleotídeos de interesse; (g) amplificar uma ou mais sequências de nucleotídeos de interesse usando pelo menos dois iniciadores de oligonucleotídeo, em que cada iniciador hibridiza para as sequências de DNA que flanqueiam as sequências de nucleotídeos de interesse; (h) hibridizar as sequência (s) amplificada(s) em um arranjo; e (i) determinar a frequência de interação entre as sequências de DNA.



"CAPTURA E CARACTERIZAÇÃO DE CROMATINA CO-
LOCALIZADA (4C)"

Campo da invenção

A presente invenção diz respeito à análise da fre-
5 quência de interação de duas ou mais seqüências de nucleotí-
deos no espaço nuclear.

Antecedentes da invenção

Estudos na arquitetura nuclear de mamífero visam
entender como 2 metros de DNA são dobrados em um núcleo de
10 10 μ n, permitindo ao mesmo tempo expressão precisa dos genes
que especificam o tipo de célula, e como isto é confiavel-
mente propagado durante cada ciclo da célula. Progresso nes-
te campo vem amplamente dos estudos da microscopia, que re-
velaram que genomas não são arranjados randomicamente no es-
15 paço nuclear. Por exemplo, heterocromatina densamente empa-
cotada é separada de eucromatina mais aberta, e os cromosso-
mos ocupam territórios distintos no espaço nuclear. Existe
um relacionamento confuso entre posição nuclear e atividade
transcricional. Embora a transcrição ocorra por todo o inte-
20 rior nuclear, genes ativos que agrupam em cromossomos prefe-
rencialmente localizam-se na borda ou fora de seu território
de cromossomo. Genes individuais podem migrar mediante mu-
danças em seu estado de transcrição, medida contra marcos
nucleares relativamente maiores, tais como territórios de
25 cromossomo, centrômeros ou a periferia nuclear. Além do
mais, dezenas de genes transcritos ativamente e separados
dezenas de megabases no cromossomo podem se agrupar no nú-
cleo, demonstrado recentemente por hibridização *in situ* de

fluorescência (FISH) para o locus β -globina e alguns outros genes selecionados. Com base na transcrição, organização genômica é associada com a coordenação de replicação, recombinação e a probabilidade de loci translocar (que pode levar a malignidades) e o ajuste e reajuste de programas epigenéticos. Com base nestas observações, acredita-se que a organização arquitetônica de DNA no núcleo da célula seja um contribuinte chave para a função genômica.

Diferentes ensaios foram desenvolvidos para permitir uma percepção na organização espacial de loci *in vivo* genômico. Um ensaio, denominado RNA-TRAP, foi desenvolvido (Carter et al. (2002) *Nat. Genet.* 32, 623) que envolve alvejar peroxidase do rábano silvestre (HRP) transcritos de RNA embrionários, seguido por quantificação de deposição de biotina catalisada por HRP na cromatina por perto.

Um outro ensaio que foi desenvolvido é denominado tecnologia de captura de conformação de cromossomo (3C), que fornece uma ferramenta para estudar a organização estrutural de uma região genômica. Tecnologia 3C envolve análise PCR quantitativa de frequências reticuladas entre dois fragmentos de restrição de DNA determinados, que dão uma medida de sua proximidade no espaço nuclear (ver Figura 1). Originalmente desenvolvidos para analisar a conformação de cromossomos em levedura (Dekker et al., 2002), esta tecnologia foi adaptada para investigar o relacionamento entre expressão do gene e dobra da cromatina em agrupamentos confusos de genes de mamífero (ver, por exemplo, Tolhuis et al., 2002; Palstra et al., 2003; e Drissen et al., 2004). Resumidamente, tecno-

logia 3C envolve reticulação *in vivo* de formaldeído de células e digestão nuclear de cromatina com uma enzima de restrição, seguido por fragmentos de ligação de DNA que foram reticulados em um complexo. Os produtos de ligação são em seguida quantificados por PCR. A etapa de amplificação do PCR requer o conhecimento da informação de sequência para cada um dos fragmentos de DNA que devem ser amplificados. Assim, tecnologia 3C fornece uma medida de frequências da interação entre fragmentos de DNA selecionados.

Existe uma necessidade importante da tecnologia de alta produção que possa selecionar sistematicamente todo o genoma de uma maneira imparcial para loci de DNA que fazem contato uns com os outros no espaço nuclear.

A presente invenção procura fornecer melhorias na tecnologia 3C.

Sumário da invenção

Tecnologia 3C atualmente permite aplicar análise apenas a um número limitado de interações de DNA-DNA selecionadas devido às limitações da etapa de amplificação de PCR, que requer conhecimento de informação de sequência específica para cada fragmento a ser analisado. Além do mais, a seleção de fragmentos de restrição como candidatos para interações de DNA de longo alcance requer uma quantidade substancial de conhecimento prévio (por exemplo, a localização de sítios hipersensíveis) do locus de interesse, que não é freqüentemente disponível. Dada a relevância funcional de muitas interações de DNA-DNA de longo alcance descrito dessa maneira, a capacidade de selecionar randomicamente elementos

de DNA que laçam uma sequência de interesse, tais como um gene promotor, intensificador, isolante, silenciador, origem de replicação ou MAR/SAR - ou uma região genômica de interesse - tais como uma região densa em genes ou pobre em genes, ou elemento repetitivo - pode facilitar enormemente o mapeamento de sequências envolvidas em uma rede regulatória.

A presente invenção diz respeito à tecnologia 4C (isto é, captura e caracteriza cromatina co-localizada), que permite a análise de alta produção da frequência de interação de duas ou mais sequências de nucleotídeos no espaço nuclear.

Tecnologia 4C (captura e caracteriza cromatina co-localizada) é uma versão modificada de tecnologia 3C que permite uma pesquisa ampla de genoma imparcial de fragmentos de DNA que interagem com um locus de escolha. Resumidamente, análise 3C é realizada como usual, mas omitindo a etapa PCR. O padrão 3C contém um bait (por exemplo, um fragmento de restrição de escolha que engloba um gene de interesse) ligado a muitas sequências de nucleotídeos de interesse diferentes (representando o ambiente genômico deste gene). O padrão é clivado por uma outra enzima de restrição secundária, e ligado. Vantajosamente, uma ou mais sequências de nucleotídeos de interesse que são ligadas à sequência de nucleotídeos alvo são amplificadas usando pelo menos um (preferivelmente, pelo menos dois) iniciador de oligonucleotídeo, em que pelo menos um iniciador hibridiza para uma sequência de DNA que flanqueia as sequências de nucleotídeos de interesse. Tipicamente, isto produz um padrão de fragmentos de PCR

que é altamente produtivo entre reações de amplificação independentes e específicas para um determinado tecido. Em uma modalidade, *HindIII* e *DpnII* são usados como enzima de restrição primária e secundária. Depois, os fragmentos amplificados podem ser marcados e opcionalmente hibridizados para um arranjo, tipicamente contra uma amostra de controle contendo DNA genômico digerido com a mesma combinação de enzimas de restrição.

Em uma modalidade preferida da presente invenção, os fragmentos ligados que são clivados por uma enzima de restrição secundária são subsequentemente religados para formar pequenos círculos de DNA.

A tecnologia 3C foi portanto modificada de maneira tal que todas as seqüências de nucleotídeos de interesse que interagem com uma seqüência de nucleotídeos alvo fossem amplificadas. Praticamente, isto significa que, em vez de desenvolver uma reação de amplificação com iniciadores que são específicos para os fragmentos que se deseja analisar, é realizada uma amplificação usando oligonucleotídeo(s) iniciador(s) que hibridizam com uma seqüência de DNA que flanqueia as seqüências de nucleotídeos de interesse. Vantajosamente, 4C não é inclinado para o projeto de iniciadores PCR que são incluídos na etapa de amplificação do PCR e que podem portanto ser usados para pesquisar o genoma completo para elementos de DNA de interação.

Aspectos do sumário da presente invenção

Aspectos da presente invenção são apresentados nas reivindicações anexas.

Em um primeiro aspecto, é fornecido um método para analisar a frequência de interação de uma sequência de nucleotídeos alvo com uma ou mais sequências de nucleotídeos de interesse (por exemplo, um ou mais loci genômicos) que compreende as etapas de: (a) fornecer uma amostra de DNA reticulado; (b) digerir o DNA reticulado com uma enzima de restrição primária; (c) ligar as sequências de nucleotídeos reticuladas; (d) reverter a reticulação; (e) digerir as sequências de nucleotídeos com uma enzima de restrição secundária; (f) ligar uma ou mais sequências de DNA de composição de nucleotídeos conhecida ao(s) sítio(s) de digestão de enzima de restrição secundária disponíveis que flanqueiam uma ou mais sequências de nucleotídeos de interesse; (g) amplificar uma ou mais sequências de nucleotídeos de interesse usando pelo menos dois oligonucleotídeos iniciadores, em que cada iniciador hibridiza para as sequências de DNA que flanqueiam as sequências de nucleotídeos de interesse; (h) hibridizar as sequência(s) amplificada(s) em um arranjo; e (i) determinar a frequência de interação entre as sequências de DNA.

Em um segundo aspecto, é fornecido um método para analisar a frequência de interação de uma sequência de nucleotídeos alvo com uma ou mais sequências de nucleotídeos (por exemplo, um ou mais loci genômicos) que compreende as etapas de: (a) fornecer uma amostra de DNA reticulado; (b) digerir o DNA reticulado com uma enzima de restrição primária; (c) ligar as sequências de nucleotídeos reticuladas; (d) reverter a reticulação; (e) digerir as sequências de nu-

cleotídeos com uma enzima de restrição secundária; (f) circularizar as seqüências de nucleotídeos; (g) amplificar uma ou mais seqüências de nucleotídeos que são ligadas à seqüência de nucleotídeos alvo; (h) opcionalmente hibridizar as
5 seqüências amplificadas com um arranjo; e (i) determinar a frequência de interação entre as seqüências de DNA.

Em um terceiro aspecto é fornecida uma seqüência de nucleotídeos circularizada que compreende uma primeira e uma segunda seqüência de nucleotídeos, em que as extremida-
10 des da primeira e de uma segunda seqüência de nucleotídeos são separadas por diferentes sítios de reconhecimento de enzima de restrição, e em que a dita primeira seqüência de nucleotídeos é uma seqüência de nucleotídeos alvo e a dita segunda seqüência de nucleotídeos é obtenível por DNA genômico
15 reticulado.

Em um quarto aspecto, é fornecido um método para preparar uma seqüência de nucleotídeos circularizada que compreende as etapas de: (a) fornecer uma amostra de DNA reticulado, (b) digerir o DNA reticulado com uma enzima de
20 restrição primária; (c) ligar as seqüências de nucleotídeos reticuladas; (d) reverter a reticulação; (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária; e (f) circularizar as seqüências de nucleotídeos.

Em um quinto aspecto, é fornecido um método para
25 analisar a frequência de interação de uma seqüência de nucleotídeos alvo com uma ou mais seqüências de nucleotídeos (por exemplo, um ou mais loci genômicos) que compreende o uso da seqüência de nucleotídeos circularizada.

Em um sexto aspecto, é fornecido um arranjo de sondas imobilizado em um suporte que compreende uma ou mais sondas que hibridizam ou são capazes de hibridizar para a sequência de nucleotídeos circularizada.

5 Em um sétimo aspecto, é fornecido um conjunto de sondas complementares em sequência à sequência de ácido nucleico adjacente para cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária em DNA genômico.

10 Em um oitavo aspecto, é fornecido um processo para preparar um conjunto de sondas que compreende as etapas de: (a) identificar cada um dos sítios de reconhecimento da enzima de restrição primária para uma enzima de restrição primária no DNA genômico; (b) projetar sondas que são capazes
15 de hibridizar para a sequência adjacente a cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico; (c) sintetizar as sondas; e (d) combinar as sondas entre si para formar um conjunto de sondas, ou substancialmente um conjunto de sondas.

20 Em um nono aspecto, é fornecido um conjunto de sondas, ou substancialmente um conjunto de sondas, obtido ou obtenível pelo processo aqui descrito.

 Em um décimo aspecto, é fornecido um arranjo que compreende arranjo de sondas, ou substancialmente o conjunto
25 de sondas aqui descrito

 Em um décimo primeiro aspecto, é fornecido um arranjo que compreende o conjunto de sondas de acordo com o aqui descrito.

Em um décimo segundo aspecto, é fornecido um processo para preparar um arranjo que compreende a etapa de imobilizar em um suporte sólido substancialmente o arranjo de sondas ou substancialmente o conjunto de sondas aqui descrito.

Em um décimo terceiro aspecto, é fornecido um processo para preparar um arranjo que compreende a etapa de imobilizar em um suporte sólido o arranjo de sondas ou o conjunto de sondas aqui descrito.

Em um décimo quarto aspecto, é fornecido um arranjo obtido ou obtenível pelo método aqui descrito.

Em um décimo quinto aspecto, é fornecido um método para identificar uma ou mais interações de DNA-DNA que são indicativas de um estado da doença particular que compreende a etapa de realizar as etapas (a)-(i) do primeiro e segundo aspectos da presente invenção, em que, na etapa (a), uma amostra de DNA reticulado é fornecida de uma célula doente e uma não doente, e em que uma diferença entre a frequência de interação entre as seqüências de DNA das células doentes e não doentes indica que a interação de DNA-DNA é indicativa de um estado de doença particular.

Em um décimo sexto aspecto, é fornecido um método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA que compreende a etapa de realizar as etapas (a)-(i) do primeiro e segundo aspectos da presente invenção, em que a etapa (a) compreende fornecer uma amostra de DNA reticulado de um sujeito; e em que a etapa (i) compreende comparar a

freqüência de interação entre as seqüências de DNA com que de um controle não afetado; em que uma diferença entre o valor obtido do controle e o valor obtido do sujeito é indicativa que o sujeito sofre da doença ou síndrome ou é indicativa que o sujeito sofrerá da doença ou síndrome.

Em um décimo sétimo aspecto, é fornecido um método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA que compreende a etapa de: realizar as etapas (a)-(i) do primeiro e segundo aspectos da presente invenção, em que a etapa (a) compreende fornecer uma amostra de DNA reticulado de um sujeito; e em que o dito método compreende a etapa adicional de: (j) identificar um ou mais loci que foi submetido a um ré-arranjo genômico que é associado com uma doença.

Em um décimo oitavo aspecto, é fornecido um método de ensaio para identificar um ou mais agentes que modulam uma interação de DNA-DNA que compreende as etapas de: (a) colocar uma amostra em contato com um ou mais agentes; e (b) realizar as etapas (a) a (i) do primeiro e segundo aspectos da presente invenção, em que a etapa (a) compreende fornecer DNA reticulado da amostra;

em que uma diferença entre (i) a freqüência de interação entre as seqüências de DNA na presença do agente e (ii) a freqüência de interação entre as seqüências de DNA na ausência do agente é indicativa de um agente que modula a interação de DNA-DNA.

Em um décimo nono aspecto, é fornecido um método

para detectar a localização de um ponto de ruptura equilibrado e/ou desequilibrado (por exemplo, uma translocação) que compreende a etapa de: (a) realizar as etapas (a) a (i) do primeiro e segundo aspectos da presente invenção; e (b) comparar a frequência de interação entre as seqüências de DNA com a de um controle; em que uma transição de baixa a alta frequência de interação de DNA-DNA na amostra comparado com o controle é indicativa da localização de um ponto de ruptura.

10 Em um vigésimo aspecto, é fornecido um método para detectar a localização de uma inversão equilibrada e/ou desequilibrada que compreende as etapas de: (a) realizar as etapas (a) a (i) do primeiro e segundo aspectos da presente invenção; e (b) comparar a frequência de interação entre as seqüências de DNA com a de um controle; em que um padrão inverso de frequências de interação de DNA-DNA para a amostra comparadas com o controle é indicativo de uma inversão.

 Em um vigésimo primeiro aspecto, é fornecido um método para detectar a localização de uma deleção que compreende as etapas de: (a) realizar as etapas (a) a (i) do primeiro e segundo aspectos da presente invenção; (b) comparar a frequência de interação entre as seqüências de DNA com a de um controle; em que uma redução na interação de frequência de DNA-DNA para a amostra comparada com o controle é
25 indicativa de deleção.

 Em um vigésimo segundo aspecto, é fornecido um método para detectar a localização de uma duplicação que compreende as etapas de: (a) realizar as etapas (a) a (i) do

primeiro e segundo aspectos da presente invenção; e (b) comparar a frequência de interação entre as seqüências de DNA com a de um controle; em que um aumento ou uma diminuição na interação de frequência de DNA-DNA para a amostra do sujeito comparado com o controle é indicativo de uma duplicação ou inserção.

Em um vigésimo terceiro aspecto, é fornecido um agente obtido ou obtenível pelo método de arranjo aqui descrito.

10 Em um vigésimo quarto aspecto, é possibilitado o uso da seqüência de nucleotídeos circularizada para identificar uma ou mais interações de DNA-DNA em uma amostra.

Em um vigésimo quinto aspecto, é possibilitado o uso da seqüência de nucleotídeos circularizada para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

15 Em um vigésimo sexto aspecto, é possibilitado o uso do arranjo de sondas ou o conjunto de sondas aqui descrito para identificar uma ou mais interações de DNA-DNA em uma amostra.

20 Em um vigésimo sétimo aspecto, é possibilitado o uso do arranjo de sondas ou o conjunto de sondas aqui descrito para o diagnóstico ou prognóstico da doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

25 Em um vigésimo oitavo aspecto, é possibilitado o uso do arranjo aqui descrito para identificar uma ou mais interações de DNA-DNA em uma amostra.

Em um vigésimo nono aspecto, é possibilitado o uso do arranjo aqui descrito para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

5 Em um trigésimo aspecto, é fornecido um método, um arranjo de sondas, um conjunto de sondas, um processo, um arranjo, um método de ensaio, um agente, ou um uso substancialmente da forma aqui descrita e com referência a qualquer dos Exemplos ou Figuras.

10 Modalidades preferidas

Preferivelmente, a reação da ligação na etapa (f) resulta na formação de círculos de DNA.

Preferivelmente, a sequência de nucleotídeos alvo é selecionada do grupo que consiste em um ré-arranjo genômico, promotor, um intensificador, um silenciador, um isolante, uma região de anexação de matriz, uma região de controle de locus, uma unidade de transcrição, uma origem de replicação, um ponto de acesso de recombinação, um ponto de ruptura de translocação, um centrômero, um telômero, uma região densa em genes, uma região pobre em genes, um elemento repetitivo e um sítio de integração (viral).

Preferivelmente, a sequência de nucleotídeos alvo é uma sequência de nucleotídeos que é associada com uma doença, ou que causa a mesma, ou é localizada até 15 Mb ou mais em um padrão de DNA linear de um locus que é associado com uma doença ou que causa a mesma.

Preferivelmente, a sequência de nucleotídeos alvos é selecionada do grupo que consiste em: *AML1*, *MLL*, *MYC*, *BCL*,

BCR, ABLI, IGH, LYL1, TAL1, TAL2, LMO2, TCR α/δ , TCR β e HOX
ou outros loci associados com doença como descrito em "Catalogue of unbalanced Chromosome Aberrations in Man" 2nd edition. Albert Schinzel. Berlin: Walter de Gruyter, 2001.
5 ISBN. 3-11-011607-3.

Preferivelmente, a enzima de restrição primária é uma enzima de restrição que reconhece um sítio de reconhecimento de 6-8 bp.

Preferivelmente, a enzima de restrição primária é
10 selecionada do grupo que consiste em *BglIII, HindIII, EcoRI, BamHI, SphI, PstI e NdeI*.

Preferivelmente, a enzima de restrição secundária é uma enzima de restrição que reconhece um sítio de reconhecimento de sequência de nucleotídeos de 4 ou 5 bp.

15 Preferivelmente, o sítio de reconhecimento da enzima de restrição secundária é localizado em mais que cerca de 350 bp do sítio de restrição primário na sequência de nucleotídeos alvo.

Preferivelmente, a sequência de nucleotídeos é
20 marcada.

Preferivelmente, as sondas são complementares em sequência à sequência de ácido nucléico adjacente a cada lado de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA
25 genômico.

Preferivelmente, as sondas são complementares em sequência à sequência de ácido nucléico que está menos que 300 pares de base de cada um dos sítios de reconhecimento da

enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

Preferivelmente, as sondas são complementares à sequência que está menos que 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

Preferivelmente, as sondas são complementares à sequência que está entre 200 e 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

Preferivelmente, as sondas são complementares à sequência que está entre 100 e 200 bp ou 0 a 100 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

Preferivelmente, duas ou mais as sondas são capazes de hibridizar para a sequência adjacente a cada sítio de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

Preferivelmente, as sondas sobrepõem-se, ou sobrepõem-se parcialmente.

Preferivelmente, a sobreposição é menos que 10 nucleotídeos.

Preferivelmente, a sequência de sonda corresponde a toda ou parte da sequência entre cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária e cada um dos primeiros sítios de reconhecimento de enzima de restrição secundária vizinhos de uma enzima de restrição secundária.

Preferivelmente, cada sonda é pelo menos 25 mer.

Preferivelmente, cada uma das sondas é 25-60 mer.

Preferivelmente, as sondas são produtos de amplificação PCR.

5 Preferivelmente, o arranjo compreende cerca de 300.000-400.000 sondas.

Preferivelmente, o arranjo compreende cerca de 385.000 ou mais sondas, preferivelmente, cerca de 750.000 sondas, mais preferivelmente, 6 x 750.000 sondas.

10 Preferivelmente, o arranjo compreende ou consiste em uma representação do genoma completo de uma dada espécie em baixa resolução.

Preferivelmente, uma a cada 2, 3, 4, 5, 6, 7, 8, 9 ou 10 sondas ordenadas em um padrão de cromossomo linear é
15 contida no arranjo.

Preferivelmente, uma transição de frequências da interação baixa para alta é indicativa da localização de um ponto de ruptura equilibrado e/ou desequilibrado.

Preferivelmente, um padrão inverso de frequências
20 da interação de DNA-DNA para a amostra do sujeito comparada com o controle é indicativo de uma inversão equilibrada e/ou desequilibrada.

Preferivelmente, uma redução na interação da frequência de DNA-DNA para a amostra do sujeito comparada com o
25 controle, em combinação com um aumento na interação de frequência de DNA-DNA para regiões mais distantes, é indicativa de uma deleção equilibrada e/ou desequilibrada.

Preferivelmente, um aumento ou uma diminuição na

freqüência de interação de DNA-DNA para a amostra do sujeito comparado com o controle é indicativo de uma duplicação ou inserção equilibrada e/ou desequilibrada.

Preferivelmente, cariotipagem espectral e/ou FISH
5 é usado antes de realizar o dito método.

Preferivelmente, a doença é uma doença genética.

Preferivelmente, a doença é câncer.

Preferivelmente, as duas ou mais seqüências ampli-
ficadas são diferencialmente marcadas.

10 Preferivelmente, as duas ou mais seqüências ampli-
ficadas são identicamente marcadas quando as seqüências re-
sidem em diferentes cromossomos.

Preferivelmente, as duas ou mais seqüências ampli-
ficadas são identicamente marcadas quando as seqüências re-
15 sidem no mesmo cromossomo a uma distância que é grande o
bastante para mínima sobreposição entre interação de sinais
de DNA-DNA.

Preferivelmente, em que o diagnóstico ou prognós-
tico é diagnóstico ou prognóstico pré-natal.

20 Vantagens

A presente invenção tem inúmeras vantagens. Estas
vantagens ficam aparentes na descrição seguinte.

A título de exemplo, a presente invenção é vanta-
josa uma vez que ela fornece *inter alia* seqüências de nucle-
25 otídeos, processos, sondas e arranjos comercialmente úteis.

A título de exemplo adicional, a presente invenção
é vantajosa uma vez que ela fornece a análise de alta produ-
ção da freqüência de interação de duas ou mais seqüências de

nucleotídeos no espaço nuclear.

A título de exemplo adicional, a presente invenção é vantajosa uma vez que, usando tecnologia 3C convencional, cada interação simples de DNA-DNA deve ser analisada por uma
5 única reação de PCR contendo um único par de iniciadores. Análise de alta produção é portanto possível somente se o PCR for automatizado, mas o custo de tantos iniciadores será muito alto. Desta maneira, alta produção (genoma amplo) análise de interações de DNA-DNA não é viável com tecnologia 3C
10 convencional. Ao contrário, a presente invenção permite agora a seleção simultânea de milhares de interações de DNA-DNA. Análise de alta produção de interações de DNA-DNA de acordo com a presente invenção aumentará bastante a escala e resolução de análise.

15 A título de exemplo adicional, a presente invenção é vantajosa uma vez que, usando tecnologia 3C convencional, a seleção é voltada para as seqüências de DNA cujos oligonucleotídeos iniciadores foram designados, solicitados e incluídos na análise. A escolha de tais oligonucleotídeos ini-
20 ciadores é tipicamente com base no conhecimento concernente à posição, por exemplo, (distante) de sítios melhoradores e/ou outros sítios elementos/ hipersensível reguladores que acredita-se serão reticulados com a seqüência de nucleotí-
25 deos que está sendo investigada. Assim, 3C convencional é voltado para o projeto de iniciadores PCR que são incluídos na etapa de amplificação de PCR, ao passo que 4C não é voltado e pode ser usado para pesquisar o genoma completo para elementos de DNA de interação. Isto se dá em virtude de a

amplificação de seqüências reticuladas no 4C não ser baseada no conhecimento previsto de seqüências que reticulam com a seqüência de nucleotídeos sendo investigada. Particularmente, em uma modalidade de 4C, seqüências que reticulam com a primeira seqüência de nucleotídeos (visada) podem ser amplificadas usando iniciadores PCR que hibridizam para a seqüência de nucleotídeos. Assim, a presente invenção permite uma seleção ampla de genoma não voltado para interações de DNA-DNA.

10 A título de exemplo adicional, a presente invenção é vantajosa porque usando tecnologia 3C convencional somente permite a amplificação seletiva de uma interação simples de DNA-DNA. Isto não é informativo quando hibridizado em um arranjo. A tecnologia foi melhorada de maneira que todos os
15 fragmentos que interagem com uma primeira seqüência de nucleotídeos (visada) sejam agora amplificados, por exemplo, amplificados seletivamente.

 A título de exemplo adicional, a presente invenção é vantajosa porque a tecnologia 4C pode ser usada para de-
20 tectar aberrações genéticas balanceadas ou não balanceadas - tais como todos os tipos de translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos - em ácido nucléico, por exemplo, cromossomos. Tecnologia 4C (que mede a proximidade de fragmentos de DNA) pode ainda determinar a
25 predisposição de um sujeito para adquirir certas translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos (por exemplo, translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos balanceados ou

não balanceados). Uma vantagem em relação às estratégias atuais é que não é necessário conhecer a posição exata da mudança em virtude de a resolução de tecnologia 4C ser de maneira tal que ela pode ser usada para detectar ré-arranjos mesmo quando o '4C-bait' (definido pelos sítios de reconhecimento de enzima de restrição primária e secundária que são analisados) é localizada fora (por exemplo, até uma megabase ou ainda mais) da mudança. Uma outra vantagem é que tecnologia 4C permite o mapeamento preciso de mudanças, uma vez que ela pode ser usada para definir os dois sítios de restrição (primária) entre os quais as mudanças ocorreram. Uma outra vantagem é que células não precisam ser cultivadas antes da fixação. Assim, tumores sólidos, por exemplo, podem também ser analisados por ré-arranjos genômicos.

A título de exemplo adicional, a presente invenção é vantajosa em virtude de a tecnologia 4C poder também detectar mudanças (por exemplo, ré-arranjos) em um estado pré-maligno, isto é, antes de todas as células conterem estas mudanças. Assim, a tecnologia pode ser usada não somente no diagnóstico de doença, mas também no prognóstico de doença.

A título de exemplo adicional, o desenho do arranjo de acordo com a presente invenção é particularmente vantajoso comparado com os arranjos de ladrilhamento genômico existentes - tais como arranjos de ladrilhamento genômico Nimblegen - uma vez que o desenho permite representação de uma parte muito grande do genoma por arranjo simples. A título de exemplo, para uma enzima de restrição que reconhece uma sequência de hexa-nucleotídeos, cerca de 3 arranjos com

cerca de 385.000 sondas cada serão suficientes cobrir, por exemplo, o genoma humano ou de camundongo completo. Para uma enzima de restrição que reconhece mais que 6 bp, um arranjo simples de cerca de 385.000 sondas pode ser usado para co-
 5 brir, por exemplo, o genoma humano ou de camundongo completo. As vantagens do desenho do arranjo são que: (1) cada sonda é informativa uma vez que cada qual analisa um evento de ligação independente, facilitando muito a interpretação do resultado; e (2) uma grande representação do genoma pode
 10 ser localizada em um arranjo simples que é barato.

Tecnologia 4C pode ser vantajosamente usada para o mapeamento fino de ré-arranjos pouco caracterizados originalmente detectados por abordagens citogenéticas (microscopia luminosa, FISH, SKY, etc.).

15 Tecnologia 4C pode ser vantajosamente usada para a seleção simultânea em um arranjo simples para combinações de ré-arranjos que ocorreram próximo aos múltiplos loci.

Descrição resumida das figuras

Figura 1

20 O princípio da tecnologia 3C

Figura 2

(a) O princípio de uma modalidade de tecnologia 4C. Análise 3C é realizada da maneira usual, como, por exemplo, *HindIII* (H) como enzima de restrição. Após reversão de
 25 reticulações, a mistura de DNA conterá uma primeira sequência de nucleotídeos (visada) ligada a muitos fragmentos diferentes. Estes fragmentos serão amplificados e marcados usando métodos de amplificação - tal como PCR inverso - em

por exemplo, círculos *DpnII*, usando primeiramente iniciadores específicos de sequência de nucleotídeo (visados). Produtos de amplificação marcados podem ser hibridizados com os arranjos aqui descritos. *HindIII* e *DpnII* são dados como exemplos, mas outras combinações de enzimas de restrição - tais como 6 ou 8- e 4 ou 5-cortadores - podem também ser usadas. (b) Resultado de PCR separado por eletroforese de gel de duas amostras de fígado (L1, L2) e cérebro (B1, B2) fetais independentes. (c) Representação esquemática da localização das sondas de microarranjo. Sondas foram projetadas em 100 bp de sítios *HindII*. Assim, cada sonda analisa um possível parceiro de ligação.

Figura 3

Tecnologia 4C detecta o ambiente genômico de β -globina (cromossomo 7). Razões não processadas são apresentadas (sinais 4C para β -globina em HS2 divididos por sinais obtidos para amostra de controle) para sondas localizadas em regiões genômicas ~35 Mb em cromossomo de camundongo 10, 11, 12, 14, 15, 7 e 8 (do topo até a base; regiões apresentadas são em distância idêntica de cada centrômero correspondente). Note o grande agrupamento de fortes sinais em torno do bait (globina) no cromossomo 7 (fileira 6), que demonstra que tecnologia 4C detecta fragmentos genômicos próximo do padrão de cromossomo linear (de acordo com o fato de que frequências da interação são inversamente proporcionais à separação do sítio genômico). Note que a região ligada em cis em torno do bait que apresenta altas intensidades de sinal é grande (>5Mb), implicando, por exemplo, que transloca-

ções podem ser detectadas mesmo com baits afastados mais que 1 MB do ponto de ruptura.

Figura 4

Tecnologia 4C detecta o ambiente genômico de Rad23A (cromossomo 8). Razões não processadas são apresentadas (sinais 4C para Rad23A divididos por sinal obtido para amostra de controle) para sondas localizadas em regiões genômicas de ~15 Mb ou mais em cromossomo de camundongo 10, 11, 12, 14, 15, 7 e 8 (do topo até a base; regiões apresentadas são em distância idêntica de cada centrômero correspondente). Note o grande agrupamento de fortes sinais em torno do bait (Rad23A) no cromossomo 8 (fileira 7), que demonstra que tecnologia 4C detecta fragmentos genômicos próximo do padrão de cromossomo linear (de acordo com o fato que frequências da interação são inversamente proporcionais à separação do sítio genômico). Note que a região ligada em cis em torno do bait que apresenta altas intensidades de sinal é grande (>5Mb), implicando por exemplo que translocações podem ser detectadas mesmo com baits afastados mais que 20 1 MB do ponto de ruptura.

Figura 5

Interações 4C de β -globina no cromossomo 7 (~135Mb) para um tecido de transcrição (fígado fetal) e um tecido de não transcrição (cérebro fetal) (analisado por uma 25 abordagem de meio de corrida). Note que interações de longo alcance com β -globina diferem entre tecidos (provavelmente dependente do estado de transcrição do gene). Independente do tecido, fortes sinais 4C limitam uma grande região (>5

Mb) em torno do bait.

Figura 6

Uros e *Eraf* interagem com β -globina em células de fígado fetal. A abordagem 4C revela que dois genes, *Eraf* e *Uros*, interagem >30 Mb com o locus β -globina localizado a-
 5 afastado ~30 Mb. Estas duas interações foram previamente encontradas por uma tecnologia diferente (Fluorescence *In situ* Hybridisation) descrito em Osborne *et al.*, *Nature Genetics* 36, 1065 (2004). Este exemplo mostra que interações de longo
 10 alcance detectadas pela tecnologia 4C podem ser verificadas por FISH e refletem verdadeiramente a proximidade nuclear.

Figura 7

Tecnologia 4C identifica precisamente transições entre regiões genômicas não relacionadas que são ligadas em
 15 *cis*. Para estes experimentos, foram usados camundongos transgênicos que contêm um cassete da região de controle de locus de β -globina humano (LCR) (~20 kb) inserido (por meio de recombinação homóloga) no locus Rad23A em cromossomo de camundongo 8. Tecnologia 4C foi realizada em fígados fetais
 20 E14.5 de camundongos transgênicos que foram homozigotos para esta inserção. Um fragmento *HindIII* no cassete da integração (HS2) foi usado como '4C-bait'. Os dados mostram que a tecnologia 4C define precisamente tanto extremidades do cassete transgênico (fileira de base: somente as sondas no LCR huma-
 25 no (~20kb) dá sinais 4C e não sondas no restante de ~380 kb da sequência de β -globina humano), e revela claramente a posição de integração em cromossomo de camundongo 8 (painel superior: comparar os sinais no cromossomo 8 (para posição

de integração, ver seta) com sinais em 6 outros cromossomo de camundongos) (cromossomos completos são representados). Este exemplo mostra que a tecnologia 4C pode ser usada para detectar a posição genômico de fragmentos de DNA etopicamente integrados (vírus, transgene, etc.). Ele mostra que transições entre regiões genômicas não relacionadas que são ligadas em *cis* podem ser identificadas precisamente, que podem ser usadas para identificar parceiros de pontos de ruptura genômico e localização.

Figura 8

Tecnologia 4C produz dados reprodutíveis, uma vez que o perfil para HS2 e β -globina são muito similares. Quatro experimentos 4C biologicamente independentes foram realizados em fígados fetais E14.5, usando tanto o gene β -globina β -principal (2 fileiras superiores) quanto β -globina HS2 (base de duas fileiras) como o bait. Estes baits são ~40 kb separados no padrão de cromossomo linear mas mostraram previamente estar perto do espaço nuclear (Tolhuis *et al.*, Molecular Cell 10, 1453 (2002)) Uma região ~5 Mb é descrita em cromossomo de camundongo 7 que é 20-20 Mb fora do locus β -globina. Os dados mostram alta reprodutibilidade entre experimentos independentes e demonstram que dois fragmentos perto do espaço nuclear compartilham parceiros de interação localizados em outro lugar no genoma.

Figura 9

A tecnologia 4C é aplicada para medir frequências da interação de DNA-DNA com sequência X (em cromossomo A) em células de uma pessoa saudável (topo) e um paciente com

translocação (A;B) (base). Intensidades de sinal representando frequências da interação de DNA-DNA (eixo-Y) são colocadas em gráfico para sondas solicitadas em padrões de cromossomo lineares (eixo-X). Em células normais, interações de DNA-DNA frequentes são detectadas em cromossomo A em torno da sequência X. Em células de paciente, uma redução de 50 % nas frequências da interação é observada para sondas em cromossomo A localizadas no outro lado do ponto de ruptura (BP) (comparar curva cinza (paciente) com linha preta (pessoa saudável)). Além do mais, a translocação traz parte do cromossomo B na proximidade física imediata para a sequência X, e interações de DNA-DNA frequentes são agora observadas para esta região no cromossomo B. A transição abrupta das frequências da interação baixa para alta neste cromossomo marca a localização de seu ponto de ruptura.

Figura 10

Inversão(s) (balanceada(s)) podem ser detectadas pela tecnologia 4C. Parceiros inversos de frequências da interação de DNA-DNA (medidos pela tecnologia 4C como intensidades de sinal de hibridização) são observados em sujeito doente (curva cheia) comparado com não doente (curva pontilhada), que revela a presença e tamanho da inversão.

Figura 11

Deteção de deleção(s) heterozigoto pela tecnologia 4C. Sondas com frequências da interação de DNA-DNA reduzidas (medidas pela tecnologia 4C como intensidades de sinal de hibridização) em sujeitos doentes (curva cinza) comparado com não doentes (curva preta), revelam a posição e tamanho

da região deletada. Sinais de hibridização residuais na região deletada do sujeito doente proveniente de alelo intacto (deleção heterozigoto). Deleção é tipicamente acompanhada por um aumento nas intensidades de sinal para sondas localizadas diretamente além da região deletada (note que a curva cinza está acima da curva preta no lado direito da deleção), uma vez que estas regiões ficam em proximidade física imediata com a sequência 4C (bait).

Figura 12

10 Duplicação detectada pela tecnologia 4C. Sondagens com maiores sinais de hibridização em um paciente (curva cheia) comparadas com um sujeito normal (curva pontilhada) indicam a posição e tamanho de duplicação. Duplicação detectada pela tecnologia 4C é tipicamente acompanhada por menores sinais de hibridização em doente em função de sujeito não doente para sondas além da região duplicada (duplicação aumenta sua separação do sítio genômico da sequência 4C).

Figura 13

20 Interações de longo alcance com β -globina reveladas pela tecnologia 4C. a, Razões não processadas de 4C sobre sinais de hibridização de controle, revelando interações de β -globina HS2 com cromossomo 7 e dois cromossomos não relacionados (8 e 14). b-c, Dados não processados para duas amostras de fígado fetal (topo, em vermelho) e cérebro fetal (base, em azul) independentes colocados em gráfico junto com duas diferentes regiões 1-2 Mb em cromossomo 7. Agrupamentos altamente produtivos de interações são observados tanto nas duas amostras de fígado fetal (b) quanto nas duas amostras

de cérebro (c). d-e, Dados de meio de corrida para as mesmas regiões. Taxa de falsa descoberta foi ajustada em 5 % (linha pontilhada). f, Representação esquemática de regiões de interação com β -globina ativa (fígado fetal, topo) e inativa (cérebro fetal, base) no cromossomo 7.

Figura 14

β -globina ativa e inativa interage com regiões cromossomais ativas e inativas, respectivamente. a, Comparação entre interações β -globina de longo alcance em fígado fetal (meio de corrida 4C, topo), análise de expressão de microarranjo em fígado fetal (escala log, meio) e a localização de genes (base) colocados em gráfico junto com uma região 4 Mb que contém o gene Uros (30 Mb fora de β -globina), mostrando que β -globina ativa interage preferencialmente com outros genes ativamente transcritos. b, A mesma comparação em cérebro fetal em torno um agrupamento de gene OR localizado - 38 distante de globina, mostrando que β -globina inativa interage preferencialmente com regiões inativas. c, Caracterização de regiões interagindo com β -globina em fígado fetal (esquerdo) e cérebro. (direito) em termos de teor do gene e atividade.

Figura 15

Rad23A unipresentemente expresso interage com regiões ativas muito similares, em fígado e cérebro fetal. a, Representação esquemática de regiões em cromossomo 8 interagindo com Rad23A ativo em fígado fetal (topo, vermelho) e cérebro (base, azul). b, Comparação entre interações Rad23A de longo alcance (meio de corrida 4C) e análise de expressão

de microarranjo (escala log) em fígado fetal (topo de dois painéis), interações Rad23A de longo alcance (meio de corrida 4C) e análise de expressão de microarranjo (escala log) em cérebro fetal (painel 3 e 4) e a localização de genes (painel inferior) colocados em gráfico junto com uma região 3 Mb de cromossomo 8. c, Caracterização de regiões interagindo com Rad23A em fígado (esquerdo) e cérebro (direito) fetal em termos de teor de gene e atividade.

Figura 16

10 Cryo-FISH confirma que tecnologia 4C identifica verdadeiramente regiões de interação. a, exemplo de parte de uma seção cryo (200 nm) mostrando mais que 10 núcleos, alguns dos quais contendo o locus β -globina (verde) e/ou Uros (vermelho). Por causa do seccionamento, muitos núcleos não
15 contêm sinais para esses dois loci. b-d, exemplos de sinais sobrepostos completamente (b) e parcialmente (c) e sinais de contato (d), que foram todos pontuados como positivo para interação. e-g, exemplos de núcleos contendo em alelos de não contato (e-f) e um núcleo contendo somente β -globina
20 (g), que foram todos pontuados como negativos para interação. Representação esquemática de resultado cryo-FISH. Porcentagens de interação com β -globina (h) e Rad23A (i) são indicadas acima dos cromossomos para regiões positivamente identificadas (ponta de flecha vermelha) e negativamente i-
25 dentificadas (ponta de flecha azul) pela tecnologia 4C. Os mesmos BACs foram usados para os dois tecidos. Frequências da interação medidas por cryo-FISH entre dois agrupamentos de gene OR distantes em fígado fetal e cérebro são indicados

abaixo dos cromossomos.

Figura 17

Análise 4C de HS2 e β -principal deu resultado altamente similar. (a) Dados 4C não processadas de quatro amostras de fígado E14.5 independentes apresentaram um padrão muito similar de interação com HS2 (topo) e β -principal (base). (b) Um Existe uma grande sobreposição entre sondas pontuadas positivo para interação no experimento HS-2 e sondas que pontuam positivo para interação no experimento β -principal.

Figura 18

Uma comparação entre interações em cis e em trans. (a) Dados 4C não processados de dois experimentos independentes mostrando interações β -globina com uma região positivamente identificada em cis (cromossomo 7, topo) e uma região em trans contendo o locus α -globina (chr.11, base). (b) Dados 4C não processadas de dois experimentos independentes mostrando interações Rad23A com uma região positivamente identificada em cis (cromossomo 8, topo) e uma região em trans que aparece em topo quando classificada de acordo com o maior valor do meio de corrida. Nenhuma das regiões em trans atendeu as condições rigorosas que permitiram a identificação de regiões de longa interação em cis.

Figura 19

Regiões que interagem com β -globina também fazem contato freqüentemente umas com as outras. Duas regiões (separadas uma da outra quase 60 Mb), contendo genes ativamente transcritos e identificadas pela tecnologia 4C para intera-

gir com β -globina em fígado fetal, mostraram frequências de co-localização por cryo-FISH de 5,5 %, que foi significativamente mais que as frequências de co-localização de fundo.

Descrição detalhada da invenção

5 Tecnologia 3c

O método 3C foi descrito com detalhe em Dekker *et al.* (2002), Tolhuis *et al.* (2002), Palstra *et al.* (2003), Splinter *et al.* (2004) and Drissen *et al.* (2004). Resumidamente, 3C é realizado digerindo DNA reticulado com uma enzi-
10 ma de restrição primária seguido por ligação em concentrações de DNA muito baixas. Nestas condições, ligação de fragmentos reticulados, que é intramolecular, é fortemente favorecida em relação a ligação de fragmentos aleatórios, que é intermolecular. A reticulação é em seguida revertida, e produtos de ligação individuais são detectados e quantificados
15 pela reação da cadeia polimerase (PCR) usando iniciadores específicos de locus. A frequência da reticulação (X) de dois loci específicos é determinada por reações PCR quantitativas usando controle e padrões reticulados, e X é expresso
20 so como a razão da quantidade do produto obtido com o padrão reticulado e com o padrão de controle.

De acordo com a presente invenção, um padrão 3C é preparado usando os métodos descritos por Splinter *et al.*, (2004) *Methods Enzymol.* 375, 493-507. (isto é, fixação de
25 formaldeído, digestão de enzima de restrição (primária), re-ligação de fragmentos de DNA reticulados e purificação de DNA). Resumidamente, uma amostra - tal como células, tecidos ou R núcleos - é fixada usando um agente de reticulação -

tal como formaldeído. A digestão de enzima de restrição primária é em seguida realizada de maneira que o DNA é digerido no contexto dos núcleos reticulados. Ligação intramolecular é em seguida realizada em baixas concentrações de DNA (por exemplo, cerca de 3,7 ng/ μ L), que favorece ligação entre fragmentos de DNA reticulados (isto é, ligação intramolecular) relacionados a ligação entre fragmentos de DNA não reticulados (isto é, ligação intermolecular ou aleatória). Em seguida, as reticulações são revertidas e o DNA pode ser purificado. O padrão 3C que é produzido contém fragmentos de restrição que são ligados em virtude de eles estarem originalmente perto do espaço nuclear.

Uma vez que uma enzima de restrição primária é usada para digerir o DNA antes da etapa de ligação intramolecular, um sítio de reconhecimento de enzima para a enzima de primária restrição separará a primeira sequência de nucleotídeos (visada) e a sequência de nucleotídeos que foi ligada. Desta maneira, o sítio de reconhecimento primário é localizado entre a primeira sequência de nucleotídeos (visada) e a sequência de nucleotídeos ligada (isto é, a segunda sequência ligada).

Sequência de nucleotídeos

A presente invenção envolve o uso de sequências de nucleotídeos (por exemplo, padrão 3Cs, padrão 4Cs, padrões de DNA, padrões de amplificação, fragmentos de DNA e DNA genômico), que podem ser disponíveis em bases de dados.

A sequência de nucleotídeos pode ser DNA ou RNA de origem genômica, sintética ou recombinante, por exemplo, cD-

NA. Por exemplo, seqüências de nucleotídeos recombinantes podem ser preparadas usando técnicas de clonagem PCR. Isto envolverá fazer um par de iniciadores flanquear uma região da seqüência a qual se deseja clonar, colocar os iniciadores em contato com mRNA ou cDNA obtido, por exemplo, de um mamífero (por exemplo, célula animal ou humana) ou célula não mamífero, desenvolver uma reação da cadeia polimerase (PCR) em condições que realizam amplificação da região desejada, isolar o fragmento amplificado (por exemplo, purificando a mistura da reação em um gel de agarose) e recuperar o DNA amplificado. Os iniciadores podem ser projetados para conter sítios de reconhecimento de enzima de restrição adequados para que o DNA amplificado possa ser clonado em um vetor de clonagem adequado.

15 A seqüência de nucleotídeos pode ser fita dupla ou fita única, quer representando a fita de sentido como anti-sentido, ou suas combinações.

 Para alguns aspectos, prefere-se que a seqüência de nucleotídeos seja DNA de fita única - tais como iniciadores de fita única e sondas.

20 Para alguns aspectos, prefere-se que a seqüência de nucleotídeos seja DNA de fita dupla - tais como padrões 3C e 4C de fita dupla.

 Para alguns aspectos, prefere-se que a seqüência de nucleotídeos seja DNA genômico - tais como um ou mais loci genômicos.

25 Para alguns aspectos, prefere-se que a seqüência de nucleotídeos seja DNA cromossômico.

A seqüência de nucleotídeos pode compreender uma primeira seqüência de nucleotídeos (visada) e/ou uma segunda seqüência de nucleotídeos.

Os sítios de reconhecimento de enzima de restrição primária e secundária serão diferentes uns dos outros e ocorrerão tipicamente somente uma vez na seqüência de nucleotídeos.

Em um aspecto, é fornecida uma seqüência de nucleotídeos circularizada que compreende uma primeira seqüência de nucleotídeos e (por exemplo, ligada a) uma segunda seqüência de nucleotídeos separadas (por exemplo, divididas, ou partidas) por um sítio de reconhecimento de enzima de restrição primária e uma secundária, em que a dita primeira seqüência de nucleotídeos é uma seqüência de nucleotídeos alvo e a dita segunda seqüência de nucleotídeos é obtenível reticulando DNA genômico (por exemplo, *in vivo* ou *in vitro*). Os sítios de reconhecimento de enzima de restrição primária e secundária serão diferentes uns dos outros e ocorrerão tipicamente somente uma vez na seqüência de nucleotídeos.

Em um aspecto adicional, é fornecida uma seqüência de nucleotídeos circularizada que compreende uma primeira seqüência de nucleotídeos e (por exemplo, ligada) uma segunda seqüência de nucleotídeos separada (por exemplo, dividida, ou partida) por um sítio de reconhecimento de enzima de restrição primária e uma secundária, em que a dita primeira seqüência de nucleotídeos é uma seqüência de nucleotídeos alvo e em que a dita primeira e segunda seqüências de nucleotídeos são obteníveis por um processo que compreende as e-

tapas de (a) reticular DNA genômico de (por exemplo, *in vivo* ou *in vitro*); (b) digerir o DNA reticulado com uma enzima de restrição primária; (c) ligar as seqüências de nucleotídeos reticuladas; (d) reverter a reticulação; e (e) digerir as
5 seqüências de nucleotídeos com uma enzima de restrição secundária para circularizar as seqüências de nucleotídeos.

Preferivelmente, a segunda seqüência de nucleotídeos cruza (por exemplo, corta) a primeira seqüência de nucleotídeos (visada). Desta maneira, a seqüência de nucleotídeos compreende a segunda seqüência de nucleotídeos, que se-
10 para a primeira seqüência de nucleotídeos (visada) em duas porções ou fragmentos - tal como aproximadamente duas porções igualmente dimensionadas ou fragmentos. Tipicamente, as porções ou fragmentos terão pelo menos cerca de 16 nucleotí-
15 deos de comprimento.

Primeira seqüência de nucleotídeos

A primeira seqüência de nucleotídeos é uma seqüência de nucleotídeos alvo.

Da maneira aqui usada, o termo "seqüência de nucleotídeos alvo" refere-se a seqüência que é usada como uma
20 seqüência bait a fim de identificar uma ou mais seqüências ao qual ela reticula (por exemplo, uma ou mais seqüências de nucleotídeos de interesse ou uma ou mais seqüências de composição de seqüência de nucleotídeos desconhecida).

25 A seqüência de nucleotídeos alvo é de seqüência conhecida.

Reticulação é indicativo de que a seqüência de nucleotídeos alvo e a seqüência reticulada a ela estavam ori-

ginalmente próximas no espaço nuclear. Determinando-se a frequência para que seqüências fiquem próximas umas das outras, é possível entender, por exemplo, a conformação de cromossomos e regiões cromossomais no contexto espacial do núcleo (por exemplo, *in vivo* ou *in vitro*). Além do mais, é possível entender as organizações estruturais confusas no genoma, por exemplo, quando melhoradores ou outros elementos reguladores transcricionais comunicam com promotores distantes localizados em *cis* ou mesmo em *trans*. Além do mais, é ainda possível entender o posicionamento de uma dada região genômica em relação às seqüências de nucleotídeos presentes no mesmo cromossomo (em *cis*) bem como a seqüências de nucleotídeos em outros cromossomos (em *trans*). Assim, é possível mapear seqüências de nucleotídeos em diferentes cromossomos que freqüentemente compartilham sítios no espaço nuclear. Além do mais, é ainda possível detectar aberrações genéticas equilibradas e/ou desequilibradas - tais como translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos equilibrados e/ou desequilibrados (por exemplo, deleções ou translocações em um ou mais cromossomos). A este respeito, aberrações genéticas resultam em mudanças nas interações de DNA-DNA na posição que a mudança ocorreu, que pode ser detectada.

A primeira seqüência de nucleotídeos (visada) de acordo com a presente invenção pode ser qualquer seqüência em que se deseja determinar a frequência de interação no espaço nuclear com uma ou mais outras seqüências.

Em uma modalidade, a primeira seqüência de nucleo-

tídeos (visada) terá mais que cerca de 350 bp de comprimento, uma vez que uma enzima de restrição secundária é escolhida que corta a primeira sequência de nucleotídeos (visada) a cerca de 350 bp ou mais do sítio de restrição primário.

Isto pode minimizar a propensão a formação de círculo, por causa das restrições topológicas (Rippe *et al.* (2001) *Trends em Biochem. Sciences* 26, 733-40). Adequadamente, a primeira sequência de nucleotídeos (visada) após a amplificação compreende pelo menos cerca de 32 bp pelo fato de que o comprimento mínimo de pelo menos dois iniciadores de amplificação usados para amplificar a segunda sequência de nucleotídeos ser cerca de 16 bases cada.

Em uma modalidade preferida, a primeira sequência de nucleotídeos (visada) pode compreender completamente ou parcialmente (por exemplo, um fragmento), ou estar perto de (por exemplo, nas proximidades de) um promotor, um intensificador, um silenciador, um isolante, uma região de anexação de matriz, uma região de controle de locus, uma unidade de transcrição, uma origem de replicação, uma recombinação de ponto de acesso, um ponto de ruptura de translocação, um centrômero, um telômero, uma região densa em genes, uma região fraca de gene, um elemento repetitivo, um sítio de integração (viral), uma sequência de nucleotídeos em que deleções e/ou mutações são relacionados com um efeito, por exemplo, doença, efeito fisiológico, funcional ou estrutural - tal como um SNP (polimorfismo nucleotídeos simples), ou sequência(s) de nucleotídeos contendo tais deleções e/ou muta-

ções, ou qualquer seqüência em que se deseja determinar a frequência de interação no espaço nuclear com outras seqüências.

Da maneira supramencionada, a primeira seqüência
 5 de nucleotídeos (visada) pode compreender completamente ou parcialmente (por exemplo, um fragmento), ou estar perto de (por exemplo, nas proximidades) uma seqüência de nucleotídeos em que aberrações genéticas - tais como deleções e/ou mutações - são relacionadas a um efeito (por exemplo, uma
 10 doença). De acordo com esta modalidade da invenção a primeira (seqüência de nucleotídeos alvo) pode portanto ser uma seqüência de nucleotídeos (por exemplo, um gene ou um locus), adjacente (no padrão de DNA físico), ou na região genômica em que mudanças foram associadas ou correlacionados
 15 com uma doença - tal como uma doença genética ou congênita. Em outras palavras, a primeira seqüência de nucleotídeos (visada) pode ser, ou pode ser escolhida com base na sua associação com um fenótipo clínico. Em uma modalidade preferida, as mudanças são mudanças em um ou mais cromossomos, e a
 20 doença pode ser como uma consequência, por exemplo, de uma ou mais deleções, uma ou mais translocações, uma ou mais duplicações e/ou uma ou mais inversões, etc.

Exemplos não limitantes de tais genes/loci são A-
ML1, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, imunoglobulina loci, *LYL1*,
 25 *TAL1*, *TAL2*, *LMO2*, *TCR α/δ* , *TCR β* , *HOX* e outros loci em várias leucemias linfoblásticas.

Outros exemplos são descritos em bases de dados eletrônicos - tais como:

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=cancerchromosomes>

<http://cgap.nci.nih.gov/Chromosomes/Mitelma>

<http://www.progenetix.net/progenetix/P14603437/ideogram.html>

<http://www.changbioscience.com/cytogenetics/cytol.pl?query=47,xy>

<http://www.possum.net.au/>

<http://www.Imdatabases.com/>

<http://www.wiley.com/legacy/products/subject/life/borgaonkar/index.html>

<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>

<http://www.sanger.ac.uk/PostGenomics/decipher/>

<http://agserver01.azn.nl:8080/ecaruca/ecaruca.Jsp>

Outros exemplos são descritos em "Catalogue of unbalanced Chromosome Aberrations in Man" 2nd edition. Albert Schiuzel. Berlin: Walter of Gruyter, 2001. ISBN 3-11-011607-3.

Em uma modalidade, o termo "adjacente" significa "diretamente adjacente", de maneira que não haja nucleotídeos interferentes entre duas seqüências adjacentes.

Em uma outra modalidade, o termo "adjacente", no contexto da seqüência de ácido nucléico e do sítio de reconhecimento da enzima de restrição primária, significa "diretamente adjacente", de maneira que não haja nucleotídeos interferentes entre a seqüência de ácido nucléico e a sítio de reconhecimento de enzima de restrição primário.

Segunda sequência de nucleotídeos.

A segunda sequência de nucleotídeos é obtenível, obtida, identificada, ou identificável para reticular DNA genômico de (por exemplo, *in vivo* ou *in vitro*).

5 A segunda sequência de nucleotídeos (por exemplo, sequência de nucleotídeos de interesse) torna-se ligada à primeira sequência de nucleotídeos (visada) após tratar uma amostra com um agente de reticulação e digerir/ligar os fragmentos de DNA reticulados. Tais sequências são reticula-
10 das na primeira sequência de nucleotídeos (visada) em virtude de eles estarem originalmente perto no espaço nuclear e ligados à primeira sequência de nucleotídeos (visada) em virtude das condições de ligação favorecerem ligação entre fragmentos de DNA reticulados (intramolecular) nos eventos
15 de ligação aleatória.

Doenças com base nas alterações - tais como translocações, deleções, inversões, duplicações e outros rearranjos genômicos - são de maneira geral causadas por interações de DNA-DNA aberrantes. Tecnologia 4C mede frequências
20 de interação de DNA-DNA, que primeiramente são função da separação do sítio genômico, isto é, frequências de interação de DNA-DNA são inversamente proporcionais à distância linear (em quilobases) entre dois loci de DNA presentes no mesmo padrão de DNA físico (Dekker et al., 2002). Assim, altera-
25 ção(s) que cria(m) fisicamente novos e/ou diferentes padrões de DNA, é(são) acompanhada(s) por interações de DNA-DNA alteradas, e isto pode ser medido pela tecnologia 4C.

Adequadamente, a segunda sequência de nucleotídeos

tem pelo menos 40 pares de base.

Agentes de reticulação - tal como formaldeído - podem ser usados para reticular proteínas em outras proteínas e ácido nucléico vizinhos. Assim, duas ou mais seqüências de nucleotídeos podem ser reticuladas somente por meio de proteínas ligadas a (uma de) estas seqüências de nucleotídeos. Agentes de reticulação sem ser formaldeído podem também ser usados de acordo com a presente invenção, incluindo os agentes de reticulação que reticulam diretamente seqüências de nucleotídeos. Exemplos de agentes que reticulam DNA incluem, mas sem limitações, luz UV, mitomicina C, nitrogênio mustarda, melfalano, diepóxido de 1,3 butadieno, *cis* diaminadicloroplatina(II) e ciclofosfamida.

Adequadamente, o agente de reticulação formará reticulados que ligarão distâncias relativamente pequenas - tal como cerca de 2 Å - selecionando assim interações íntimas que podem ser revertidas.

Reticulações podem ser realizadas, por exemplo, para incubar as células em formaldeído 2 % a temperatura ambiente - tal como para incubar células 1×10^7 em 10 mL de DMEM-10 % FCS suplementado com formaldeído 2 % por 10 minutos à temperatura ambiente.

Enzima de restrição primária

Da maneira aqui usada, o termo "enzima de restrição primária" refere-se a uma primeira enzima de restrição que é usada para digerir o DNA reticulado.

A enzima de restrição primária será escolhida dependendo do tipo de seqüência visada (por exemplo, locus) a

ser analisada. Deseja-se que experimentos preliminares sejam realizados para otimizar as condições de digestão.

A enzima de restrição primária pode ser selecionada de enzimas de restrição que reconhecem pelo menos seqüências 6 bp ou mais de DNA.

Enzimas de restrição que reconhecem seqüências 6 bp de DNA incluem, mas sem limitações, AcII, HindII, SspI, BspLUIII, AgeI, MluI, SpeI, BgIII, Eco47III, StuI, ScaI, ClaI, AvaIII, VspI, MfeI, PmaCI, PvuII, NdeI, NcoI, SmaI, SacII, AvrII, PvuI, XmaII, SpII, XhoI, PstI, AfiIII, EcoRI, AatII, SacI, EcoRV, SphI, NaeI, BsePI, NheI, BamHI, NarI, ApaI, KpnI, SnaI, SalI, ApaLI, HpaI, SnaBI, BspHI, BspMII, NruI, XbaI, BclI, MstI, BclI, Bsp1407I, PstI, AsuII e AhaII-I.

Enzimas de restrição que reconhecem mais que uma seqüência 6 bp de DNA incluem, mas sem limitações, 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, Sml I, Srf I e Sse8387 I.

Para alguns aspectos da presente invenção, no caso de enzimas de restrição reconhecerem seqüências 6 bp, BgIII, HindIII ou EcoRI são preferidos.

O termo "sítio de reconhecimento de enzima de restrição primária" refere-se ao sítio em uma seqüência de nucleotídeos que é reconhecida e clivada pela enzima de restrição primária.

Enzima de restrição secundária

Da maneira aqui usada, o termo "enzima de restri-

ção secundária" refere-se a uma segunda enzima de restrição que é usada após a digestão de enzima de restrição primária, ligação de DNA reticulado, des-reticulação e purificação de DNA (opcional). Em uma modalidade, a enzima de restrição secundária é usada para fornecer extremidades de DNA definidas para as seqüências de nucleotídeos de interesse, que permite a ligação de seqüências de composição de nucleotídeos conhecida nos sítios de reconhecimento de enzima de restrição secundária que flanqueiam as seqüências de nucleotídeos de interesse.

Em uma modalidade, ligação de seqüências de composição de nucleotídeos conhecida com os sítios de reconhecimento de enzima de restrição secundária que flanqueiam (por exemplo, estão de cada lado ou extremidade) as seqüências de nucleotídeos de interesse envolvem ligação em condições diluídas para favorecer a ligação intramolecular entre os sítios de reconhecimento de enzima de restrição secundária que flanqueiam seqüências de nucleotídeos visadas e as seqüências de nucleotídeos de interesse ligadas. Isto resulta efetivamente na formação de círculos de DNA em que seqüências de nucleotídeos visadas conhecidas flanqueiam seqüências de interesse desconhecidas.

Em uma outra modalidade, ligação de seqüências de composição de nucleotídeos conhecida com os sítios de reconhecimento de enzima de restrição secundária que flanqueiam (por exemplo, estão de cada lado ou extremidade) as seqüências de nucleotídeos de interesse envolvem a adição de seqüências de DNA únicas de composição de nucleotídeos conhe-

cidas, seguido por ligação em condições que favorecem ligação intermolecular entre os sítios de reconhecimento de enzima de restrição secundária que flanqueiam as seqüências de nucleotídeos de interesse e introduzem seqüências de DNA únicas de composição de nucleotídeos conhecida.

Em uma modalidade, a enzima de restrição secundária é escolhida de maneira que nenhum sítio de enzima de restrição secundária fique em cerca de 350 bp (por exemplo, 350-400 bp) do sítio de restrição primário.

Em uma outra modalidade, a enzima de restrição secundária é escolhida de maneira que o mesmo sítio de enzima de restrição secundária é provavelmente para ser localizada na seqüência de nucleotídeos ligada (isto é, a seqüência reticulada ligada). Uma vez que as extremidades da primeira seqüência de nucleotídeos (visada) e da seqüência de nucleotídeos ligada podem ser extremidades compatíveis coesivas (ou rombas), as seqüências podem ainda ser ligadas a fim de circularizar o DNA. Desta maneira, a etapa de digestão é seguida por ligação em condições diluídas que favorecem interações intramoleculares e circularização opcional do DNA por meio das extremidades compatíveis.

Preferivelmente, o sítio de reconhecimento da enzima de restrição secundária é um sítio de reconhecimento de seqüência de nucleotídeos de 4 ou 5 bp. Enzimas que reconhecem seqüências de DNA de 4 ou 5 bp incluem, mas sem limitações, TspEI, MaeII, A1uI, N1aIII, HpaII, FnuDII, MaeI, DpnI, MboI, HhaI, HaeIII, RsaI, TaqI, CviRI, MseI, Sth132I, AciI, DpnII, Sau3AI e MnII.

Em uma modalidade preferida, a enzima de restrição secundária é NlaIII e/ou DpnII. O termo "sítio de reconhecimento de enzima de restrição secundária" refere-se ao sítio na sequência de nucleotídeos que é reconhecida e clivada pela enzima de restrição secundária.

Após a digestão com a enzima de restrição secundária, uma reação adicional da ligação é realizada. Em uma modalidade, esta reação da ligação liga as sequências de DNA de composição de sequência de nucleotídeos conhecidas com o sítio de digestão da enzima de restrição secundária de uma ou mais sequências que são ligadas à sequência de nucleotídeos alvo.

Enzima de restrição terciária

Da maneira aqui usada, o termo "enzima de restrição terciária" refere-se a uma terceira enzima de restrição que pode ser opcionalmente usada após a etapa de enzima de restrição secundária a fim de linearizar DNA circularizado antes da amplificação.

Preferivelmente, a enzima de restrição terciária é como enzima que reconhece um sítio de reconhecimento de 6 bp ou mais nucleotídeos. Preferivelmente, a enzima de restrição terciária digere a primeira sequência de nucleotídeos (visada) entre sítio de enzima de restrição primária e secundária de reconhecimento. Como entendido pelos versados na técnica, é desejável que a enzima de restrição terciária não digira a primeira sequência de nucleotídeos (visada) muito perto dos sítios de reconhecimento de enzima de restrição primária e secundária de maneira que os iniciadores de amplificação não

possam mais hibridizar. Desta maneira, prefere-se que o sítio de restrição de reconhecimento de enzima terciária seja localizado pelo menos a mesma distância dos sítios de reconhecimento de enzima de restrição primária e secundária que o comprimento do iniciador a ser usado, de maneira que os iniciadores de amplificação possam ainda hibridizar.

Em uma modalidade preferida, a enzima de restrição terciária é uma que reconhece uma sequência de 6-bp de DNA. O termo "sítio de reconhecimento de enzima de restrição terciária" refere-se ao sítio na sequência de nucleotídeos que é reconhecida e clivada pela enzima de restrição terciária.

Sítio de reconhecimento

Endonucleases de restrição são enzimas que clivam a espinha dorsal açúcar-fosfato de DNA. Em cenários mais práticos, uma dada enzima de restrição corta ambas fitas de DNA duplex em um trecho de apenas umas bases. Os substratos para enzimas de restrição são sequências de DNA de fita dupla denominadas sítios/sequências de reconhecimento.

O comprimento de sítios de reconhecimento de restrição varia, dependendo da enzima de restrição que é usada. O comprimento da sequência de reconhecimento dita com que frequência a enzima cortará uma sequência de DNA.

A título de exemplo, inúmeras enzimas de restrição reconhecem uma sequência de DNA de 4 bp. As sequências e a enzima que reconhecem a sequência de DNA de 4 bp incluem, mas sem limitações, AATT (TspEI), ACGT (MaeII), AGCT (AluI), CATG (NlaIII), CCGG (HpaII), CGCG (FnuDII), CTAG (MaeI), GATC (DpnI, DpnII, Sau3AI & MboI), GCGC (HhaI), GGCC (HaeIII),

GTAC (RsaI), TCGA (TaqI), TGCA (CviRI), TTAA (MseI), CCCG (SthI321), CCGG (AciI) e CCTC (MnlI)

A título de exemplo adicional, inúmeras enzimas de restrição reconhecem uma sequência de DNA de 6 bp. As se-
 5 quências e a enzima que reconhecem a sequência bp par bases 6 de DNA incluem, mas sem limitações, AACGTT (AclI), AAGCTT (HindIII), AATATT (SspI), ACATGT (BspLU11I), ACCGGT (AgeI), ACGCGT (MluI), ACTAGT (SpeI), AGATCT (BglIII), AGCGCT (Eco47III), AGGCCT (StuI), AGTACT (ScaI), ATCGAT (ClaI), ATG-
 10 CAT (AvaIII), ATTAAT (VspI), CATATG (MfeI), CACGTG (PmaCI), CAGCTG (PvuII), CATATG (NdeI), CCATGG (NcoI), CCCGGG (SmaI), CCGCGG (SacII), CCTAGII), CTCGAG (XhoI), CTGCAG (PstI), CTTA-AG (AfiIII), GAATTC (EcoRI), GACGTC (AatII), GAGCTC (SacI), GATATC (EcoRV), GCATGC (SphI), GCCGGC (NaeI), GCCGGC (Bse-
 15 PI), GCTAGC (NheI), GGATCC (BamHI), GGCGCC (NarI), GGGCCC (ApaI), GGTACC (KpnI), GTATAC (SnaI), GTCGAC (SaII), GTCGAC (ApaLI), GTTAAC (HpaI), TACGTA (SnaBI), TCATGA (BspHI), TCCGGA (BspMII), TCGCGA (NruI), TCTAGA (XbaI), TGATCA (BclI), TGCACA (MstII), TGGCCA (BaiI), TGTACA (Bsp1407I), TTATAA
 20 (PsiI), TTCGAA (AsuII) e TTTAAA (AhaIII).

A título de exemplo adicional, inúmeras enzimas de restrição reconhecem um 7 bp sequência de DNA. As sequências e a enzima que reconhecem a sequência de 7 bp de DNA incluem, mas sem limitações, CCTNAGG (SauI), GCTNAGC (EspI), GGT-
 25 NACC (BstEII) e TCCNGGA (PfoI).

A título de exemplo adicional, inúmeras enzimas de restrição reconhecem uma sequência de 8 bp de DNA. As sequências e a enzima que reconhecem a sequência de 8 bp de

DNA incluem, mas sem limitações, ATTTAAAT (SwaI), CCTGCAGG (Sse8387I), CGCGGGCG (Sse232I), CGTCGACG (SgrDI), GCCCGGGC (SrfI), GCGATCGC (SgfI), GCGGCCGC (NotI), GGCCGGCC (FseI), GGCGCGCC (AscI), GTTTAAAC (PmeI) e TTAATTAA (PacI).

5 Inúmeras dessas enzimas contêm a sequência CG que podem ser metiladas *in vivo*. Inúmeras enzimas de restrição são sensíveis a esta metilação e não clivarão a sequência metilada, por exemplo, HpaII não clivará a sequência CC^mGG, ao passo que seu isosquizômero MspI é insensível a esta mo-
10 dificação e clivará a sequência metilada. Desta maneira, em alguns casos, as enzimas sensíveis a metilação eucariótica não são usadas.

Em uma modalidade, um sítio de reconhecimento é um sítio de digestão.

15 Em uma modalidade, um sítio de reconhecimento de enzima de restrição é um sítio de digestão de enzima de restrição.

Circularização

De acordo com uma modalidade da presente invenção,
20 o material para 4C é preparado criando círculos de DNA digerindo o padrão 3C com uma enzima de restrição secundária, seguido por ligação.

Preferivelmente, é escolhida uma enzima de restrição secundária que corta a primeira sequência de nucleotí-
25 deos (visada) em mais que cerca de 350 bp (por exemplo, 350-400 bp) do sítio de restrição primário. Vantajosamente, isto minimiza a propensão de formação de círculo por causa das restrições topológicas (Rippe *et al.* (2001) *Trends in Bio-*

chem. Sciences 26, 733-40).

Preferivelmente, a enzima de restrição secundária é um cortador freqüente que reconhece um sítio de reconhecimento de enzima de restrição de 4 ou um 5 bp. Assim é possível obter fragmentos de restrição menores para eficiências iguais de amplificação de todos os fragmentos ligados durante a amplificação.

Antes da ligação e digestão da enzima de restrição secundária, o padrão de DNA compreenderá um sítio de reconhecimento de enzima secundária na primeira seqüência de nucleotídeos (visada) localizada a mais que cerca de 350-400 bp do sítio de restrição primário e um outro sítio de reconhecimento de enzima secundária localizado na seqüência de nucleotídeos que foi ligada (isto é, na segunda seqüência de nucleotídeos).

Preferivelmente, a etapa de digestão da enzima de restrição secundária é realizada por mais que 1 hora por toda a noite e seguida por inativação a quente da enzima.

Preferivelmente, o DNA nesta mistura da reação é purificado usando métodos/estojos convencionais que são conhecidos na tecnologia.

Após a etapa de digestão de enzima de restrição secundária, um sítio de enzima de restrição secundária será localizado a mais que 350-400 bp do sítio de restrição primário na primeira seqüência de nucleotídeos (visada), e um outro sítio de enzima de restrição secundária será localizado na seqüência de nucleotídeos ligada (isto é, a segunda seqüência de nucleotídeos). Uma vez que as extremidades da

primeira seqüência de nucleotídeos (visada) e da seqüência de nucleotídeos ligada têm extremidades compatíveis, as seqüências podem ser ligadas a fim de circularizar o DNA.

A etapa de digestão é então seguida por ligação em
5 condições diluídas que favorecem interações intramoleculares e circularização do DNA por meio das extremidades compatíveis.

Preferivelmente, a reação da ligação é realizada em uma concentração de DNA de cerca de 1-5 ng/ μ L.

10 Preferivelmente, a reação da ligação é realizada por mais que 1 hora (por exemplo, 2, 3, 4 ou mais horas) a cerca de 16-25 °C.

Desta maneira, após a reação da ligação, DNA circularizado pode ser preparado. O DNA circularizado compreenderá os sítios de reconhecimento pelo menos para a enzima de restrição secundária, ou as enzimas de restrição primárias e as secundárias. Em DNA circularizado contendo a primeira seqüência de nucleotídeos (visada), a sítio de reconhecimento da enzima de restrição primária e os sítios reconhecimento
15 de enzima de restrição secundária definirão as extremidades da primeira seqüência de nucleotídeos (visada) e a seqüência de nucleotídeos ligada (isto é, a segunda seqüência de nucleotídeos). Desta maneira a primeira seqüência de nucleotídeos (visada) e a seqüência de nucleotídeos ligada são sepa-
20 radas (por exemplo, divididas) pelo sítio de reconhecimento da enzima de restrição primária e o sítio de reconhecimento de enzima de restrição secundária.
25

Amplificação

Uma ou mais reações de amplificação podem ser realizadas a fim de amplificar os padrões 4C de DNA.

Amplificação de DNA pode ser realizada usando inúmeros métodos diferentes que são conhecidos na tecnologia.

5 Por exemplo, DNA pode ser amplificado usando a reação em cadeia da polimerase (Saiki *et al.*, 1988); PCR mediada por ligação, amplificação da replicase Qb (Cahill, Foster and Mallyan, 1991; Chetverin and Spirin, 1995; Katanaev, Kurnasov and Spirin, 1995); a reação da cadeia ligase (LCR) (Landegren *et al.*, 1988; Barany, 1991); o sistema de replicação de
10 sequência auto-sustentado. (Fahy, Kwok e Gingeras, 1991) e amplificação de deslocamento de fita (Walker *et al.*, 1992).

Preferivelmente, DNA é amplificado usando PCR. "PCR" refere-se ao método de K. B. Mullis, patentes U.S.
15 4.683.195, 4.683.202 e 4.65.188, que descrevem um método para aumentar a concentração de um segmento de uma sequência de nucleotídeos em uma mistura de DNA genômico sem clonagem ou purificação.

Em uma modalidade, é usado PCR inverso. PCR inverso
20 (IPCR) (descrito por Ochman *et al.* (1988) *Genetics* 120(3), 621-3) é um método para a amplificação rápida *in vitro* de sequências de DNA que flanqueiam uma região de sequência conhecida. O método usa a reação da cadeia polimerase (PCR), mas ele tem os iniciadores orientados na direção
25 contrária à orientação usual. O padrão para os iniciadores reversos é um fragmento de restrição que foi ligado sobre si próprio para formar um círculo. PCR inverso tem muitas aplicações em genética molecular, por exemplo, a amplificação e

identificação de elementos reversíveis para flanquear seqüenciadores. Para aumentar a eficiência e reprodutibilidade da amplificação, prefere-se que os círculos de DNA sejam linearizados antes da amplificação usando uma enzima de restrição terciária. Preferivelmente, é usada uma enzima de restrição terciária que é um cortador 6 bp ou mais. Preferivelmente, a enzima de restrição terciária corta a primeira (visada) seqüência de nucleotídeos entre o sítio da enzima de restrição primária e secundária.

10 A digestão do padrão 3C com a enzima de restrição secundária, circularização opcional, ligação (por exemplo, ligação em condições diluídas) e linearização opcional de primeira seqüência de nucleotídeos (visada) contendo círculos produz um padrão de DNA para amplificação ("padrão 4C de
15 DNA").

 Para a etapa de amplificação, pelo menos dois oligonucleotídeos iniciadores são usados, nos quais cada iniciador hibridiza para uma seqüência de DNA que flanqueia as seqüências de nucleotídeos de interesse. Em uma modalidade
20 preferida, pelo menos dois oligonucleotídeos iniciadores são usados, nos quais cada iniciador hibridiza para a seqüência visada que flanqueia as seqüências de nucleotídeos de interesse.

 Em uma modalidade, o termo "flanquear" no contexto
25 de hibridização do iniciador significa que pelo menos um iniciador hibridiza para uma seqüência de DNA adjacente a uma extremidade (por exemplo, a extremidade 5') da seqüência de nucleotídeos de interesse e pelo menos um iniciador hibridi-

za para uma sequência de DNA nas outras extremidades (por exemplo, a extremidade 3') da sequência de nucleotídeos de interesse. Preferivelmente, pelo menos um iniciador direto hibridiza para uma sequência de DNA adjacente a uma extremidade (por exemplo, a extremidade 5') da sequência de nucleotídeos de interesse e pelo menos um iniciador reverso hibridiza para uma sequência de DNA nas outras extremidades (por exemplo, a extremidade 3') da sequência de nucleotídeos de interesse.

Em uma modalidade preferida, o termo "flanquear" no contexto de hibridização do iniciador significa que pelo menos um iniciador hibridiza para uma sequência visada adjacente a uma extremidade (por exemplo, a extremidade 5') da sequência de nucleotídeos de interesse e pelo menos um iniciador hibridiza para uma sequência visada nas outras extremidades (por exemplo, a extremidade 3') da sequência de nucleotídeos de interesse. Preferivelmente, pelo menos um iniciador direto hibridiza para uma sequência visada adjacente a uma extremidade (por exemplo, a extremidade 5') da sequência de nucleotídeos de interesse, e pelo menos um iniciador reverso hibridiza para uma sequência visada nas outras extremidades (por exemplo, a extremidade 3') da sequência de nucleotídeos de interesse.

Da maneira aqui usada, o termo "iniciador" refere-se a um oligonucleotídeo, quer de ocorrência natural como em uma digestão de restrição purificada quanto produzido sinteticamente, que é capaz de agir como um ponto de iniciação de síntese, quando colocado sob condições em que a síntese de

um produto de extensão do iniciador que é complementar a uma fita de ácido nucléico é induzida (isto é, a presença de nucleotídeos e um agente de indução tal como polimerase de DNA e em uma temperatura e pH adequados). O iniciador é preferi-
5 velmente fita única para eficiência máxima em amplificação, mas pode ser fita dupla. Caso seja fita dupla, o iniciador é primeiro tratado para separar suas fitas antes usadas para preparar produtos de extensão. Preferivelmente, o iniciador é um oligodeoxiribonucleotídeo. O iniciador deve ser sufici-
10 entemente grande para iniciar a síntese de produtos de extensão na presença do agente de indução. Os comprimentos exatos dos iniciadores dependerão de muitos fatores, incluindo temperatura, fonte de iniciador e o uso do método.

Adequadamente, os iniciadores terão pelo menos 15,
15 preferivelmente pelo menos 20, por exemplo pelo menos 25, 30 ou 40 nucleotídeos em comprimento. Preferivelmente, os iniciadores de amplificação têm de 16 a 30 nucleotídeos em comprimento.

Preferivelmente, os iniciadores são projetados o
20 mais perto possível dos sítios de reconhecimento de enzima de restrição primária e secundária que separam o primeiro (visado) sequência de nucleotídeos e a segunda sequência de nucleotídeos. Os iniciadores podem ser projetados de maneira que eles estejam dentro de cerca de 100 nucleotídeos - tal
25 como cerca de 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2 ou 1 nucleotídeos) fora do sítio de reconhecimento de enzima de restrição primária e secundária.

Adequadamente, os iniciadores de amplificação são

projetados de maneira tal que suas extremidades 3' fiquem voltadas para fora em direção aos sítios de reconhecimento de enzima de restrição primária e secundária para que a extensão siga imediatamente através dos sítios de restrição para a segunda sequência de nucleotídeos.

Se o método de amplificação que é usado é PCR inverso, então prefere-se que as reações de amplificação sejam realizadas a cerca de 100-400 ng de DNA do padrão 4C (por cerca de 50 µL da mistura de reação de PCR) ou outras quantidades de DNA para as quais reações PCR em replicata dá resultado produtivo (ver Figura 1) e inclui um número máximo de eventos de ligação por reação de PCR.

Preferivelmente, a reação de amplificação de PCR inverso é realizada usando o sistema Expand Long Template PCR (Roche), usando tampão 1 de acordo com as instruções do fabricante.

Amostra

O termo "amostra" da maneira aqui usada, tem significado natural. Uma amostra pode ser qualquer entidade física que compreende DNA que é ou pode ser reticulada. A amostra pode ser de material biológico, ou pode ser derivada dele.

A amostra pode ser de uma ou mais entidades, ou pode ser derivada delas - tal como uma ou mais células, um ou mais núcleos, ou uma ou mais amostras de tecido. As entidades podem ser de qualquer entidade em que DNA - tal como cromatina - está presente, ou podem ser deriváveis delas. A amostra pode ser de uma ou mais células isoladas ou uma ou

mais amostras de tecido isoladas, ou um ou mais núcleos isolados ou podem ser derivada deles.

A amostra pode ser de células vivas e/ou células mortas e/ou lisatos nucleares e/ou cromatina isolada, ou pode ser derivada deles.

A amostra pode ser de sujeito doente e/ou não doente, ou pode ser derivada dele.

A amostra pode ser de um sujeito que é suspeito de sofrer de uma doença, ou pode ser derivada dele.

A amostra pode ser de um sujeito para ser testado com relação à probabilidade de que ele sofrerá de uma doença no futuro, ou pode ser derivada dele.

A amostra pode ser de material de paciente viável ou não viável, ou pode ser derivada dele.

A fixação de células e tecidos par o uso para preparar o padrão 3C é descrito com detalhe em Splinter *et al.*, (2004) *Methods Enzymol.* 375, 493-507

Marcação

Preferivelmente, as seqüências de nucleotídeos (por exemplo, padrões de DNA 4C amplificados, iniciadores ou sondas *etc.*) são marcadas a fim de assistir em suas aplicações à jusante - tal como hibridização de arranjo. A título de exemplo, os padrões de DNA 4C podem ser marcados usando iniciação aleatória ou tradução por corte.

Uma ampla variedade de marcações (por exemplo, reportadores) pode ser usada para marcar as seqüências de nucleotídeos aqui descritas, particularmente durante a etapa de amplificação. Marcações adequadas incluem agentes radio-

nuclídeos, enzimas, fluorescente, quimioluminescente, ou cromogênicos bem como substratos, cofatores, inibidores, partículas magnéticas e similares. Patentes que preceituam o uso de tais marcações incluem US-A-3817837, US-A-3850752, 5 US-A-3939350, US-A-3996345, US-A-4277437, US-A-4275149 e US-A-366241.

Marcações adicionais incluem, mas sem limitações, P-galactosidase, invertase, proteína fluorescente verde, luciferase, cloramfenicol, acetiltransferase, β -glucuronidase, 10 exo-glucanase e glucoamilase. Marcações fluorescentes podem também ser usadas, bem como reagentes fluorescentes especificamente sintetizados com propriedades químicas particulares. Uma ampla variedade de meios para medir fluorescência é disponível. Por exemplo, algumas marcações fluorescentes a- 15 presentam uma mudança no espectro de excitação ou emissão, algumas apresentam transferência de energia de ressonância onde um reportador fluorescente perde fluorescência, ao passo que uma segunda ganha em fluorescência, algumas apresentam uma perda (finalização) ou aparência de fluorescência, 20 ao passo que algumas reportam movimentos rotacionais.

A fim de obter material suficiente para marcar, ampliações múltiplas podem ser ligadas, em vez de aumentar o número de ciclos de amplificação por reação. Alternativamente, nucleotídeos da reação de amplificação (por exemplo, 30 ciclos de PCR (sem marcação) + 10 ciclos de PCR 25 (mais marcação)).

Arranjo

Em uma modalidade particularmente vantajosa, os

padrões de DNA 4C que são preparados de acordo com os métodos aqui descritos podem ser hibridizados em um arranjo. Desta maneira, tecnologia de arranjo (por exemplo, microarranjo) pode ser usado para identificar seqüências de nucleotídeos - tais como fragmentos genômicos - que freqüentemente compartilham um sítio nuclear com uma primeira (visada) seqüência de nucleotídeos.

Arranjos existentes - tal como arranjos de expressão e genômicos - podem ser usados de acordo com a presente invenção. Entretanto, a presente invenção também procura fornecer arranjos inéditos (por exemplo, arranjos de DNA) como aqui descrito.

Um "arranjo" é uma coleção de ácidos nucléicos intencionalmente criada que pode ser preparada tanto sinteticamente quanto bio-sinteticamente e classificada pela atividade biológica em uma variedade de diferente formatos (por exemplo, bibliotecas de moléculas solúveis; e bibliotecas de oligos presos a contas de resina, chipes de sílica, ou outros suportes sólidos). Adicionalmente, o termo "arranjo" inclui as bibliotecas de ácidos nucléicos que podem ser preparadas localizando pontos de ácidos nucléicos essencialmente de qualquer comprimento (por exemplo, de 1 a cerca de 1.000 monômeros de nucleotídeos em comprimento) em um subtrato.

Tecnologia de arranjo e as várias técnicas e aplicações associadas com ela estão descritas de maneira geral em inúmeros livros texto e documentos. Isto inclui Lemieux *et al*, 1998, *Molecular Breeding* 4, 277-289, Schema and

Davis. *Parallel Analysis with Biological Chips*. in *PCR Methods Manual* (eds. M. Tunis, D. Gelfand, J. Sninsky), Schena and Davis, 1999, *Genes, Genomas artd Chips*. In *DNA Mcroarrays: A Practical Approach* (ed. M. Schema), Oxford University Press, Oxford, UK, 1999), *The Chipping Forecast* (Nature Genetics special issue; January 1999 Supplement), Mark Schema *Microarray Biochip Technology*, (Eaton Publishing Company), Comes, 2000, *The Scientist* 14(17):25, Gwynne and Page, *Microarray analysis: the next revolution in molecular biology*, *Science*, 1999 August 6; and Eakins and Chu, 1999, *Trends in Biotechnology*, 17, 217-218.

Tecnologia de arranjo supera as desvantagens com métodos tradicionais em biologia molecular, que de maneira geral funcionam em uma base "um gene em um experimento", resultando em baixa produção e o incapacidade para perceber a "imagem total" da função do gene. Atualmente, as principais aplicações para tecnologia de arranjo incluem a identificação de seqüência (mutação gene/gene) e a determinação de nível de expressão (abundância) de genes. Perfilamento de expressão do gene pode fazer uso de tecnologia de arranjo, opcionalmente em combinação com técnicas proteômicas (Cells et al., 2000, *FEBS Lett*, 480(1):2-16; Lockhart and Wier, 2000, *Nature* 405(6788):827-836; Khm et at., 1999, 20(2):223-9). Outras aplicações de tecnologia de arranjo são também conhecidas na tecnologia; por exemplo, descoberta do gene, pesquisa do câncer (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), análise SNP (Wang et

al., 1998, *Science*, 280(5366): 1077-82), descoberta de medicamento, farmacogenômicos, diagnóstico de doença (por exemplo, utilizando dispositivos microfluídicos: *Chemical & Engineering News*, February 22, 1999, 77(8):27-36), toxicology (Rockett and Dix (2000), *Xenobiotica*, 30(2): 155-77; Afshari et al., 1999, *Cancer Res*; 59(19): 4759-60) e toxicogenômicos (um híbrido de genômicos funcionais e toxicologia molecular).

De um modo geral, qualquer biblioteca pode ser ar-
10 ranjada de uma maneira ordenada em um arranjo, separando espacialmente os elementos da biblioteca. Exemplos de bibliotecas adequadas para arranjo incluem bibliotecas de ácido nucléico (incluindo bibliotecas de DNA, cDNA, oligonucleotídeo, etc.), peptídeo, polipeptídeo e bibliotecas de proteína,
15 na, bem como bibliotecas que compreendem qualquer molécula, tal como bibliotecas ligantes, entre outras.

As amostras (por exemplo, elementos de uma biblioteca) são de maneira geral fixadas ou imobilizadas em uma fase sólida, preferivelmente um substrato sólido, para limitar difusão e mistura das amostras. Em uma modalidade preferida, bibliotecas de ligantes de ligação de DNA podem ser preparadas. Em particular, as bibliotecas podem ser imobilizadas em uma fase sólida substancialmente planar, incluindo membranas e substratos não porosos tais como plástico e vi-
20 dro. Além do mais, as amostras são preferivelmente arranjadas de uma maneira tal que indexação (isto é, referência ou acesso a uma amostra particular) é facilitada. Tipicamente, as amostras são aplicadas como pontos em uma formação de ma-

lha. Sistemas de ensaio comuns podem ser adaptados com este propósito. Por exemplo, um arranjo pode ser imobilizado na superfície de uma microplaca, tanto com múltiplas amostras em um poço, quanto com uma única amostra em cada poço. Além
5 do mais, o substrato sólido pode ser uma membrana, tais como um nitrocelulose ou membrana de náilon (por exemplo, membranas usadas em experimentos tipo blot). Substratos alternativos incluem vidro, ou substratos a base de sílica. Assim, as amostras são imobilizadas por qualquer método adequado conhecido na tecnologia, por exemplo, por interações de carga,
10 ou por acoplamento químico nas paredes ou base dos poços, ou o superfície da membrana. Outros meios de arranjar e fixar podem ser usados, por exemplo, pipetação, gota e toque, dispositivo piezoelétrico, tecnologia de jato de tinta e jato
15 de bolha, aplicação eletrostática, etc. No caso de chipes a base de silicone, fotolitografia pode ser utilizada para arranjar e fixar as amostras no chipe.

As amostras podem ser arranjadas "localizadas" no substrato sólido; isto pode ser feito à mão ou fazendo uso
20 de robótica para depositar a amostra. De um modo geral, arranjos podem ser descritos como macroarranjos ou microarranjos, a diferença sendo o tamanho dos pontos de amostra. Macroarranjos tipicamente contêm tamanhos da localização pontual da amostra de cerca de 300 microns ou mais e podem ser
25 facilmente imageados por dispositivos de varredura de gel e blot existentes. Os tamanhos da localização pontual da amostra em microarranjos são tipicamente menos que 200 microns de diâmetro e estes arranjos freqüentemente contêm milhares

de pontos. Assim, microarranjos podem requerer robótica e equipamento de imageamento especializados, que podem precisar ser feitos sob medida. Instrumentação é descrita de maneira geral em uma revisão por Cortese, 2000, *The Scientist* 14[11]:26.

Técnicas para produzir bibliotecas imobilizadas de moléculas de DNA foram descritas na tecnologia. De maneira geral, métodos da tecnologia mais antiga descreveram como sintetizar bibliotecas de molécula de ácido nucléico de fita única, usando, por exemplo, técnicas de mascaramento para construir várias mutações de seqüências nas várias posições discretas no substrato sólido. Patente U.S. 5.837.832 descreve um método melhorado para produzir arranjos de DNA imobilizados com substratos de silicone com base na tecnologia de integração em escala muito grande. Em particular, Patente U.S. 5.837.832 descreve uma estratégia denominada "azulejamento" para sintetizar conjuntos específicos de sondas em locais espacialmente definidos em um substrato que pode ser usado para produzir as bibliotecas de DNA imobilizadas da presente invenção. A patente U.S. 5837832 também fornece referências para técnicas anteriores que podem também ser usadas.

Arranjos podem também ser construídos usando química de fotodeposição.

Arranjos de peptídeos (ou peptidomiméticos) podem também ser sintetizados em uma superfície de uma maneira que coloca cada elemento da biblioteca distinto (por exemplo, seqüência de peptídeo única) em uma localização discreta,

pré-definida no arranjo. A identidade de cada elemento da biblioteca é determinada por sua localização espacial no arranjo. Os locais no arranjo onde interações de ligação entre uma molécula pré-determinada (por exemplo, uma visada ou sonda) e elementos da biblioteca reativos ocorrem é determinado, identificado assim as seqüências dos elementos da biblioteca reativos na base de localização espacial. Estes métodos são descrito em Patente U.S. 5.143.854; W090115070 e W092110092; Fodor et al. (1991) *Science*, 251: 767; Dower and Fodor (1991) *Ann. Rep. Med. Chem.*, 26: 271.

Para ajudar a detecção, marcações são tipicamente usadas (da maneira supradiscutida) - tal como qualquer reportador facilmente detectável, por exemplo, um reportador fluorescente, bioluminescente, fosforescente, radioativo, etc. Tais reportadores, sua detecção, acoplamento em alvos/sondas, etc. são discutidos em outro lugar neste documento. Marcação de sondas e alvos é também revelada em Shalton et al., 1996, *Genoma Res* 6(7):639-45.

Exemplos específicos de arranjos de DNA são como a seguir:

Formato I: sonda cDNA (500 ~ 5.000 bases de comprimento) é imobilizada em uma superfície sólida tal como vidro usando localização por robô e exposta a um conjunto de alvos tanto separadamente quanto em uma mistura. Este método é amplamente considerado tendo sido desenvolvido na Stanford University (Ekins and Chu, 1999, *Trends in Biotechnology*, 1999, 17, 217-218).

Formato II: um arranjo de oligonucleotídeos (20-

25-mer oligos, preferivelmente, 40-60 mer oligos) ou ácido nucléico de peptídeo (PNA) as sondas são sintetizadas tanto *in situ* (em-chipe) quanto por síntese convencional seguido por imobilização em-chipe. O arranjo é exposto no DNA da amostra marcado, hibridizado, e a identidade/abundância de seqüências complementares são determinadas. Chipe de DNA como este é vendido pela Affymetrix, Inc., sob a marca registrada GeneChip®. Agilent e Nimblegen também fornecem arranjos adequados (por exemplo, arranjos de azulejamento genômico).

Exemplos de alguns formatos de microarranjo comercialmente disponíveis são apresentados na Tabela 1 a seguir (ver também Marshall e Hodgson, 1998, *Nature Biotechnology*, 16(1), 27-31).

Empresa	Nome do produto	Método do arranjo	Etapa de hibridização	Leitura
Affymetrix Inc. Santa Clara California	GepeChip®	Síntese <i>in situ</i> (no chipe) fotolitográfica de oligos de -20-25-mer em pastilha de silício, que são cortadas em chips de 1,25 cm ² ou 5,25 cm ²	10.000-260.000 características de oligo sondadas com fragmentos de 30-40 nucleotídeos marcados de amostra de cDNA ou RNA anti-sentido	Fluorescência

Brax, Cambridge, UK		oligo sintético pe- queno sintetizado for a do chipe	1.000 oligos em um "chipe universal" sondados com áci- do nucléico alvo	Espectome- tria ed de massa
<u>Gene Logic,</u> <u>Inc.,</u> Columbia, Maryland	READS™			
<u>Genometrix</u> <u>Inc., The</u> Woodlands, Texas	Universal Arrays™			
<u>GENSET</u> <u>Paris,</u> <u>França</u>				
<u>Hyseg inc.,</u> Sunnyvale, California	HyChip™	amostras de DNA de 500-2.000 nt impres- sas em membranas de 0,6 cm ² (HyGnostics) ou de --18 cm ² (Des- coberta do Gene) Fabricou oligos de 5- mer impressos como arranjos de 1,15 cm em vidro (HyChiP)	64 pontos de cDNA de amostra sonda- dos com oligos de 8.000 de 7 mer (HyGnostics) ou <=55,000 pontos de amostra de cDNA sondadas com 300 oligo de 7- mer (Descoberta do Gene)	Radioisóto- po Fluorescên- cia

			1024 pontos oligo universais sonda-ram cDNA de amostras de 10 kb, oligos de 5-mer, marcadas e ligase	
<u>Incyte</u> <u>pharmaceu</u> <u>tic ls.Inc.</u> Palo Alto California	GEM	Impressão piezoelétrica para fragmentos de localização de PCR e síntese no chip de oligos	<=1000 (eventualmente 10.000) pontos de fragmento oligo/PCR sondados com RNA marcado	Fluorescência e radioisóto
<u>Molecular</u> <u>dinamics,</u> <u>Inc.,</u> Sunnyvale, California	Storm® FluorImager ®	500-5.000 nt cDNAs impressos por caneta em -10 cm ² em lamina de vidro	10.000 pontos de cDNA sondados com 200-400 ht de cDNA de amostras marcados	Fluorescência
<u>Nanogen,</u> San Diego, california	Semiconduc tor Microchip	oligos de -20-mer pré-fabricados, capturados em pontos eletroativos em pastilha de silício, que são cortadas em chips de <=1 cm ²	25, 64, 400 (e eventualmente 10.000) pontos de oligo polarizados para melhorar a hibridização de 200-400 nt de cDNAs de amostras marcadas	Fluorescência

<u>Protogene</u> <u>Laborato-</u> <u>ries</u> Palo Alto california		Síntese em chipe de oligos de 40-50 mer em 9 cm ² de chipe de vidro por meio de impressão para um arranjo de tensão superficial	<-8000 pontos de oligo sondados com 200- 400 nt de amostra de ácido nucléico marcada	Fluorescên- cia
<u>Sequenom.</u> Hamburg, Germany, and San Diego California	MassArray Spectro- Chip	Impressão off-set de arranjo; em torno de oligos de 20-25-mer	250 localizações por SpectroChip in- terrogados por desorção a laser e espectometria de massa	espectome- tria de massa
Synteni, Inc., Fremont, California	UniGEM™	500-5.000 nt de cDNAs impressos por ponta no chipe de vidro de - 4cm ²	<=10.000 pontos de cDNA sondados com 200- 400 nt de cDNA de amostras marcadas	Fluorescên- cia
Nimblegen Systems Inc., Madison	Homo sapiens Whole- Genome 60 mer Microarray	38.000 transcritos com 5 sondas por gene de 17,4 mm x 13 mm		plataforma de varredu- ra de 5-micron

		Prototípico macrochipe PNA com síntese no chipe de sondas usando químicas cf-moc ou t-moc	Em torno de 1.000 pontos em um chipe de 8 x 12 cm	Fluorescência/espectometria de massa
--	--	---	---	--------------------------------------

Tabela 1: Exemplos de formatos de microarranjos de hibridização atualmente disponíveis

A fim de gerar dados de ensaios baseados em arranjo, é detectado um sinal que significa a presença de ou ausência de hibridização entre uma sonda e uma sequência de nucleotídeos. A presente invenção contempla adicionalmente técnicas de marcação direta e indireta. Por exemplo, marcação direta incorpora corantes fluorescentes diretamente nas sequências de nucleotídeos que hibridizam para o arranjo associado a sondas (por exemplo, corantes são incorporados na sequência de nucleotídeos por síntese enzimática na presença de nucleotídeos ou iniciadores de PCR marcados). Esquemas de marcação direta produzem sinais de hibridização fortes, tipicamente usando famílias de corantes fluorescentes com estruturas químicas e características similares, e são simples de implementar. Em modalidades preferidas que compreendem marcação direta de ácidos nucleicos, cianina ou análogos alexa são utilizados em análise de arranjo comparativo de multi-flúor. Em outras modalidades, esquemas de marcação indireta podem ser utilizados para incorporar epítomos nos ácidos nucleicos, tanto antes de quanto após a hibridização para as sondas de microarranjo. Um ou mais manchamentos, procedimentos e reagentes são usados para marcar o complexo

hibridizado (por exemplo, uma molécula fluorescente que liga aos epítomos, fornecendo assim um sinal fluorescente em virtude da conjugação de molécula do corante com o epítomo da espécie hibridizada).

5 Análise de dados é também uma importante parte de um experimento envolvendo arranjos. Os dados brutos de um experimento do arranjo tipicamente são imagens, que precisam ser transformadas em matrizes - tabelas, onde fileiras representam, por exemplo, genes, colunas representam, por exemplo, várias amostras tais como tecidos ou condições experimentais, e números em cada célula, por exemplo, caracterizam a expressão de uma sequência particular (preferivelmente, uma segunda sequência que foi ligada com a primeira (visada) sequência de nucleotídeos) na amostra particular. Estas matrizes têm de ser analisadas ainda mais, se qualquer conhecimento a cerca dos processos biológicos subjacente tiver que ser extraído. Métodos de análise de dados (incluindo análise de dados supervisionada e não supervisionada bem como abordagens bioinformáticas) são revelados em Brazma and
10
15
20 Vilo J (2000) FEES Lett 480(1): 17-24.

Na forma aqui descrita, uma ou mais sequências de nucleotídeos (por exemplo, o padrão de DNA) que são marcadas e subsequentemente hibridizadas para um arranjo compreendem uma sequência de nucleotídeos que é enriquecida por pequenos
25 trechos de sequências com uma assinatura distinta, isto é, cobrindo a sequência de nucleotídeos entre o sítio de reconhecimento da enzima de restrição primária que foi ligada durante o procedimento 3C com a primeira (visada) sequência

quência de nucleotídeos, e seu respectivo sítio de reconhecimento de enzima de restrição secundária vizinho.

Um arranjo simples pode compreender múltiplas (por exemplo, duas ou mais) seqüências baits.

5

Sondas

Da maneira aqui usada, o termo "sonda" refere-se a uma molécula (por exemplo, um oligonucleotídeo, quer de ocorrência natural como em uma digestão de restrição purificada quer produzido sinteticamente, recombinantemente ou por amplificação de PCR), que é capaz de hibridizar com uma outra molécula de interesse (por exemplo, um outro oligonucleotídeo). Quando as sondas são oligonucleotídeos, elas podem ser fita única ou fita dupla. As sondas são usadas na detecção, identificação e isolamento de alvos particulares (por exemplo, seqüências de gene). Aqui descrito, contempla-se que sondas usadas na presente invenção podem ser marcadas com uma marcação para que seja detectável em qualquer sistema de detecção, incluindo, mas sem limitações, sistemas de enzima (por exemplo, ELISA, bem como enzima com base em ensaios histoquímicos), fluorescente, radioativo, e luminescente. Com relação a arranjos e microarranjos, o termo "sonda" é usado para referir-se a qualquer material hibridizável que é afixado com o arranjo com o propósito de detectar uma seqüência de nucleotídeos que foi hibridizado com a dita sonda. Preferivelmente, estas sondas têm 25-60 mers ou mais.

25

Estratégias para desenho de sonda são descritas em W095/11995, EP 717,113 e W097129212.

Uma vez que 4C permite uma pesquisa de genoma não

predisposta ampla para interações, é vantajoso preparar um arranjo com sondas interrogando cada sítio de reconhecimento da enzima de restrição primária no genoma possível (por exemplo, único/não repetitivo).

5 Assim, desenho do arranjo somente depende da escolha de enzima de restrição primária e não das seqüências de nucleotídeos primária ou secundária atual.

10 Embora arranjos existentes possam ser usados de acordo com a presente invenção, prefere-se usar configurações alternativas.

15 Em uma configuração, uma ou mais sondas no arranjo são projetadas de maneira que elas possam hibridizar perto dos sítios que são digeridos pela enzima de restrição primária. Mais preferivelmente, a(s) sonda(s) é(são) em cerca de 20 bp do sítio de reconhecimento de enzima de restrição primária. Mais preferivelmente, a(s) sonda(s) é(são) em cerca de 50 bp do sítio de reconhecimento de enzima de restrição primária.

20 Adequadamente, a(s) sonda(s) é(são) dentro de cerca de 100 bp (por exemplo, cerca de 0-100 bp, cerca de 20.400 bp) do sítio de reconhecimento de enzima de restrição primária.

25 Em uma configuração preferida, uma sonda simples, única, é projetada em 100 bp a cada lado dos sítios que são digeridos pela enzima de restrição primária.

 Em uma outra configuração preferida, as posições os sítios digerido pela enzima de restrição secundária com relação às posições dos sítios digeridos pelos sítios de

restrição primária são levadas em consideração. Nesta configuração, uma sonda simples, única, é projetada somente a cada lado dos sítios digeridos pela enzima de restrição primária que têm o sítio de reconhecimento de enzima de restrição secundária perto a uma distância bastante grande por uma sonda de um dado comprimento a ser projetada entre o sítio de reconhecimento de enzima de restrição primária e secundária. Nesta configuração, por exemplo, nenhuma sonda é projetada ao lado de um sítio de reconhecimento da enzima de restrição primária particular que tem um sítio de reconhecimento da enzima de restrição secundária dentro de 10 bp desse mesmo lado.

Em uma outra configuração, as sondas no arranjo são projetadas de maneira que elas possam hibridizar em ambos os lados dos sítios que são digeridos pela enzima de restrição primária. Adequadamente, uma sonda simples a cada lado do sítio de reconhecimento da enzima de restrição primária pode ser usada.

Ainda em uma outra configuração, duas ou mais sondas (por exemplo, 3, 4, 5, 6, 7 ou 8 ou mais) podem ser projetadas a cada lado do sítio de reconhecimento de enzima de restrição primária, que podem em seguida ser usadas para investigar o mesmo evento de ligação. Para o número e posição de sondas com relação a cada sítio de reconhecimento de enzima de restrição primária, a localização genômica exata de seu sítio de reconhecimento de enzima de restrição secundária vizinho pode ser levada em consideração.

Ainda em uma outra configuração, duas ou mais son-

das (por exemplo, 3, 4, 5, 6, 7 ou 8 ou mais) pode ser projetadas próximo a cada sítio de reconhecimento da enzima de restrição primária independente do sítio de reconhecimento de enzima de restrição secundária mais próxima. Nesta configuração, todas sondas devem estar perto dos sítios de reconhecimento da enzima de restrição primária (preferivelmente dentro de 300 bp do sítio de restrição).

Vantajosamente, o último desenho e também o desenho que usa 1 sonda por (lado) sítio de reconhecimento de enzima de restrição primária, permite o uso de diferentes enzimas de restrição secundária em combinação com uma dada enzima de restrição primária.

Vantajosamente, o uso de múltiplas (por exemplo, 2, 3, 4, 5, 6, 7 ou 8 ou mais) sondas por sítio de reconhecimento da enzima de restrição primária pode minimizar o problema de obter resultados negativos falsos por causa do fraco desempenho de sondas individuais. Além do mais, pode também aumentar a confiabilidade de dados obtidos com um experimento chipe simples e reduzir o número de arranjos requeridos para tirar conclusões estatisticamente sólidas.

As sondas para o uso no arranjo podem ser maiores que 40 nucleotídeos em comprimento e podem ser isotérmicas.

Preferivelmente, sondas contendo seqüências de DNA repetitiva são excluídas.

Supõe-se que diagnóstico de sondas para os sítios de restrição que flanqueiam diretamente ou que estão próximas da primeira seqüência de nucleotídeos dêem sinais de hibridização muito fortes e podem também ser excluídas do de-

senho da sonda.

O arranjo pode tratar qualquer genoma incluindo mamífero (por exemplo, humano, camundongo (por exemplo, cromossomo 7)), vertebrado (por exemplo, zebrafish)), ou não
5 vertebrado (por exemplo, bacteriano, levedura, fúngico ou inseto (por exemplo, *Drosófila*)) genomas.

Em uma modalidade preferida adicional, o arranjo contém 2-6 sondas em torno de cada sítio de restrição primária único e o mais próximo possível do sítio de digestão de
10 enzima de restrição.

Preferivelmente, a distância máxima do sítio de digestão de enzima de restrição é cerca de 300 bp.

Em um modalidade preferida adicional da presente invenção, são fornecidos arranjos para enzimas de restrição
15 - tal como *HindIII*, *EcoRI*, *BglII* e *NotI* - que tratam os genomas mamífero ou não mamífero. Vantajosamente, o desenho dos arranjos aqui descritos evita a necessidade de redesenhar arranjos para cada sequência visada, análise fornecida é realizada na mesma espécie.

20 Conjuntos de sondas

Da maneira aqui usada, o termo "conjunto de sondas" refere-se a uma suíte ou uma coleção de sondas que hibridiza para cada um dos sítios de reconhecimento da enzima de restrição primária para uma enzima de restrição primária
25 em um genoma.

Desta maneira, é fornecido, em um aspecto adicional, um conjunto de sondas complementares em sequência à sequência de ácido nucléico adjacente a cada uma dos sítios de

reconhecimento de enzima de restrição primária para uma enzima de restrição primária no DNA genômico.

Adequadamente, o conjunto das sondas é complementar em sequência aos primeiros 25-60 (por exemplo, 35-60, 45-60, ou 50-60) ou mais nucleotídeos que são adjacentes a cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico. O conjunto de sondas pode ser complementar em sequência com um lado (por exemplo, qualquer) ou ambos os lados do sítio de reconhecimento de enzima de restrição primária. Desta maneira, as sondas podem ser complementares em sequência à sequência de ácido nucléico adjacente a cada lado de cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico.

É também possível definir uma janela (por exemplo, 300 bp ou menos - tal como 250 bp, 200 bp, 150 bp ou 100 bp - do sítio de reconhecimento de enzima de restrição primária) em que uma ou mais sondas para o conjunto podem ser projetadas. Tais fatores que são importantes na definição da janela na qual redesenha as sondas são, por exemplo, teor de GC, ausência de sequências palindrômicas que podem formar estruturas do grampo, tamanho máximo para distâncias de um tipo de nucleotídeo simples. Desta maneira, o conjunto de sondas pode ser complementar em sequência à sequência de ácido nucléico que está menos que 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico.

É também possível definir uma janela de cerca de 100 bp do sítio de reconhecimento da enzima de restrição

primária a fim de identificar sondas ideais próximas a cada sítio de restrição.

Em modalidades adicionais da presente invenção, o conjunto das sondas é complementar à sequência que está me-
5 nos que 300 bp de cada um dos sítios de reconhecimento de enzima de restrição primária no DNA genômico, complementares à sequência que é entre 200 e 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária no DNA ge-
nômico e/ou complementares à sequência que está entre 100 e
10 200 bp de cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico.

Em modalidades adicionais da presente invenção, o conjunto das sondas é complementar à sequência que é de 0 a 300 bp de cada um dos sítios de reconhecimento da enzima de
15 restrição primária no DNA genômico, complementar à sequência que está entre 0 a 200 bp de cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico e/ou complementar à sequência que está entre 0 a 100 bp de cada um dos sítios de reconhecimento da enzima de restrição pri-
20 mária em DNA genômico (por exemplo, cerca de 10, 20, 30, 40, 50, 60, 70, 80 ou 90 bp de cada um dos sítios de reconheci-
mento da enzima de restrição primária no DNA genômico).

Duas ou mais sondas pode ainda ser projetadas que são capazes de hibridizar para a sequência adjacente a cada
25 sítio de reconhecimento da enzima de restrição primária no DNA genômico.

As sondas podem sobrepor-se, ou sobrepõem-se parcialmente. Se as sondas ficarem sobrepostas, então a sobre-

posição é preferivelmente menos que 10 nucleotídeos.

Podem também ser usados fragmentos de PCR representando o primeiro 1-300 nucleotídeos (por exemplo, 1-20, 1-40, 1-60, 1-80, 1-100, 1-120, 1-140, 1-160, 1-180, 1-200, 5 1-220, 1-240, 1-260 ou 1-280 nucleotídeos) que flanqueiam cada sítio de reconhecimento da enzima de restrição primária.

Fragmentos de PCR podem também ser usados como sondas que correspondem exatamente a cada sítio genômico que 10 é flanqueado pelo sítio de reconhecimento da enzima de restrição primária e o primeiro sítio de reconhecimento de enzima de restrição vizinho ao segundo. Desta maneira, a sequência de sondas pode corresponder a toda ou parte da sequência entre cada um dos sítios de reconhecimento da enzima 15 de restrição primária e cada um dos sítios de reconhecimento de enzima de restrição primários vizinhos do secundário.

Tipicamente, as sondas, arranjos de sondas ou conjuntos de sondas serão imobilizados em um suporte. Suportes (por exemplo, suportes sólido) podem ser feitos de uma variedade de materiais - tais como vidro, sílica, plástico, náilon ou nitrocelulose. Suportes são preferivelmente rígidos e têm uma superfície plana. Suportes tipicamente têm de cerca 20 de 1-10.000.000 regiões espacialmente endereçáveis discretas, ou células. Suportes tendo cerca de 10-1.000.000 ou 25 cerca de 100-100.000 ou cerca de 1000-100.000 células são comuns. A densidade de células é tipicamente pelo menos cerca de 1.000, 10.000, 100.000 ou 1.000.000 células dentro de um centímetro quadrado. Em alguns suportes, todas as células

são ocupadas por misturas agrupadas de sondas ou um conjunto de sondas. Em outros suportes, algumas células são ocupadas por misturas agrupadas de sondas ou um conjunto de sondas, e outras células são ocupadas pelo menos até o grau de pureza
5 obtenível pelos métodos da síntese, por um tipo simples de oligonucleotídeo.

Preferivelmente, o arranjo aqui descrito compreende mais que uma sonda por sítio de reconhecimento de enzima de restrição primária, que, no caso de uma enzima de restrição
10 ção de corte 6 bp, ocorre, por exemplo, aproximadamente 750.000 vezes por genoma humano ou camundongo.

Para uma enzima de restrição que reconhece uma sequência de reconhecimento >6 bp, um arranjo simples de cerca de 2 x 750.000 sondas pode ser usado para cobrir, por exemplo,
15 plo, o genoma humano ou camundongo completo, com 1 sonda a cada lado de cada sítio de restrição.

Em um desenho de arranjo preferido, o número total de moléculas de sonda de uma dada sequência de nucleotídeos presente no arranjo é em grande excesso em relação a fragmentos homólogos presentes na amostra 4C a ser hibridizada
20 com tal arranjo. Dada a natureza de tecnologia 4C, fragmentos representando regiões genômicas perto da sequência de nucleotídeos analisada no padrão de cromatina linear será em grande excesso na amostra de hibridização 4C (como descrito
25 na Figura 2). Para obter informação quantitativa a cerca de eficiência de hibridização de tais fragmentos abundantes, pode ser necessário reduzir a quantidade de amostra a ser hibridizada e/ou aumentar o número de moléculas de uma dada

dada sonda de sequência de oligonucleotídeos no arranjo.

Assim, para a detecção de elementos reguladores de DNA que fazem contato freqüentemente, por exemplo, um elemento promotor de gene, pode ser necessário usar um arranjo
5 com sondas que representa somente a região genômica selecionada (por exemplo, cerca de 0,540 Mb), mas com cada sonda única presente em múltiplas posições no arranjo (por exemplo, cerca de 100, 200, 1000). Tais desenhos podem também ser preferidos com o propósito de diagnóstico para detectar
10 local (por exemplo, em cerca de 10 Mb) ré-arranjos genômicos - tal como deleções, inversões, duplicações, etc. - em torno um sítio (por exemplo, gene de interesse).

O arranjo pode compreender cerca de 3 x 750.000 sondas, 4 x 750.000 sondas, 5 x 750.000 sondas ou, preferi-
15 velmente, 6 x 750.000 sondas. Mais preferivelmente, o arranjo compreende 6 x 750.000 sondas com 2, 3, 4, 5, 6, 7 ou 8 ou mais sondas a cada lado de cada sítio de restrição. Mais preferivelmente, o arranjo compreende 6 x 750.000 sondas, com 3 sondas em cada lado de cada sítio de restrição.

20 Arranjos de sondas ou conjuntos de sondas podem ser sintetizados etapa por etapa em um suporte, ou podem ser anexados de forma pré-sintetizada. Um método de síntese VL-SIPS.TM. (como descrito em US 5.143.854 e EP 476.014), que exige o uso de luz para direcionar a síntese de sondas de
25 oligonucleotídeo em arranjos miniaturizados de alta densidade. Algoritmos para desenho de máscaras para reduzir o número de ciclos de síntese são descritos em US 5.571.639 e US. 5.593.839. Arranjos podem também ser sintetizados de um modo

combinatório, distribuindo monômeros às células de um suporte por caminhos de fluxo mecanicamente restritos, descritos em EP 624.059. Arranjos podem também ser sintetizados por reagentes de localização em um suporte usando uma impressora
5 jato de tinta (ver, por exemplo, EP 728.520).

No contexto da presente invenção, os termos "substancialmente um conjunto de sondas" "substancialmente o arranjo de sondas" significa que o conjunto ou o arranjo de sondas compreende pelo menos cerca de 50, 60, 70, 80, 90,
10 95, 96, 97, 98 ou 99 % do conjunto ou arranjo de sondas total ou completo. Preferivelmente, o conjunto ou o arranjo de sondas é um conjunto de sondas total ou completo (isto é, 100 %).

Em uma modalidade preferida, o arranjo compreende
15 uma sonda simples única por lado de cada sítio de reconhecimento enzima de restrição primária que é presente em um dado genoma. Se este número de sondas exceder o número de sondas que pode ser contido por um arranjo simples, o arranjo pode preferivelmente ainda conter uma representação do genoma
20 completo de uma dada espécie, mas em baixa resolução, com por exemplo uma a cada 2, 3, 4, 5, 6, 7, 8, 9, 10, 10^2 , 10^3 ou 10^4 etc. sondas ordenadas no padrão do cromossomo linear presente no arranjo. Tais arranjos que tratam o genoma humano completo, ou outros, em resolução subideal podem ser pre-
25 feridos com relação a arranjos de alta resolução que tratam parte do mesmo genoma, por exemplo, em casos onde devem ser encontrados parceiros translocação.

Preferivelmente, a representação do genoma comple-

to de uma dada espécie em baixa resolução é obtida por sondas no arranjo em que cada uma representa um fragmento de restrição simples obtido após digestão com uma enzima de restrição primária. Preferivelmente, isto é obtido ignorando
5 cada segunda, terceira, quarta, quinta, sexta, sétima, oitava, nona, décima, vigésima, trigésima, quadragésima, quinquagésima, sexagésima, septuagésima, octagésima, nonagésima ou centésima, etc. sonda que hibridiza para o mesmo fragmento de restrição.

10 Preferivelmente, a representação do genoma completo de uma dada espécie em baixa resolução compreende sondas que são distribuídas igualmente juntamente com os padrões do cromossomo lineares. Preferivelmente, isto é obtido ignorando uma ou mais sondas nas regiões genômicas que apresentam
15 densidade alta de sonda.

Hibridização

O termo "hibridização" da maneira aqui usada incluirá "o processo pelo qual uma fita de ácido nucléico une-se a uma fita complementar por meio de pareamento de base"
20 bem como o processo de amplificação realizado por exemplo, em tecnologias de reação da cadeia de polimerase (PCR).

Seqüências de nucleotídeos capazes de hibridização seletiva serão feitas de maneira geral pelo menos-75 %, preferivelmente pelo menos 85 ou 90 % e mais preferivelmente
25 pelo menos 95 % ou 98 % homólogo a seqüência de nucleotídeos complementar correspondente em uma região de pelo menos 20, preferivelmente pelo menos 25 ou 30, por exemplo, pelo menos 40, 60 ou 100 ou mais nucleotídeos contíguos.

"Hibridização específica" refere-se a ligação, duplexação, ou hibridização de uma molécula somente com uma sequência de nucleotídeos particular sob condições rigorosas (por exemplo, 65 °C e 0,1 x SSC {1 x SSC = NaCl 0,15 M, Na-citrato 0,015 M pH 7,0}). Condições rigorosas são condições sob as quais uma sonda hibridizará com sua sequência visada, mas com nenhuma outra sequência. Condições rigorosas dependem da sequência, e são diferentes em circunstâncias diferentes. Sequências maiores hibridizam especificamente em altas temperaturas. De maneira geral, condições rigorosas são selecionadas para ser cerca de 5 °C menor que os pontos de fusão (T_m) para a sequência específica em um comprimento e pH iônicos definidos. A T_m é a temperatura (sob intensidade iônicas, pH, e concentração de ácido nucléico definidos) em que 50 % das sondas complementares em uma sequência visada hibridizam para a sequência visada em equilíbrio. (Como as sequências visadas estão de maneira geral presentes em excesso, a T_m, 50 % das sondas são ocupadas em equilíbrio). Tipicamente, condições rigorosas incluem uma concentração de sal de pelo menos cerca de 0,01 a 1,0 M de concentração iônica Na (ou outros sais) a pH 7,0 a 8,3 e a temperatura é pelo menos cerca de 30 °C para sondas pequenas. Condições rigorosas podem também ser alcançadas com a adição de agentes desestabilizantes - tais como formamida ou sais de tetralquila amônio.

Como é de conhecimento dos versados na técnica, uma hibridização de máxima severidade pode ser usada para identificar ou detectar sequências de nucleotídeos idênticas

ao passo que uma hibridização intermediária severa (ou baixa) pode ser usada para identificar ou detectar seqüências de polinucleotídeos similares ou relacionadas.

São também descritos métodos para a hibridização de arranjos de sondas com seqüências de nucleotídeos marcadas ou não marcadas. As condições de reação de hibridização particulares podem ser controladas para alterar a hibridização (por exemplo, aumentar ou diminuir a severidade de ligação da sonda/alvo). Por exemplo, temperatura da reação, concentrações de ânions e cátions, adição de detergentes, e similares podem todos alterar as características de hibridização de sondas de arranjos e moléculas visadas.

Frequência de interação

Frequências de ligação de quantificação de fragmentos de restrição dão uma medida de suas frequências reticuladas. Adequadamente, isto pode ser obtido usando PCR usado em tecnologia 3C convencional descrita por Splinter et al. (2004) (*supra*). Resumidamente, a formação de produtos PCR pode ser medida varrendo as intensidades de sinal após a separação em brometo de etídio manchado em gel de agarose, usando um imageador Typhoon 9200 (Molecular Dynamics, Sunnyvale, CA). Adequadamente, diversos controles são usados para a interpretação correta de dados também descritos em Splinter et al. (2004) (*supra*).

Uma vez que a tecnologia 4C aqui descrita fornece para a análise de alta produção da frequência de interação de duas ou mais seqüências de nucleotídeos no espaço nuclear, prefere-se que as frequências de ligação de fragmentos

de restrição sejam quantificadas usando os arranjos aqui descritos.

Para quantificação, sinais obtidos por uma amostra 4C podem ser normalizados com sinais obtidos por uma amostra de controle. Amostra 4C e amostras(s) de controle(s) serão marcadas com marcações diferentes e discerníveis (por exemplo, corantes) e serão simultaneamente hibridizadas com o arranjo. Amostra(s) de controle(s) conterá(ão) tipicamente todos fragmentos de DNA (isto é, todas segundas seqüências de nucleotídeos potenciais que foram ligadas com a primeira seqüência de nucleotídeos (visada)) em quantidades equimolares e, para excluir uma propensão na eficiência de hibridização, elas devem ser similares em tamanho à(s) segunda(s) seqüência(s) de nucleotídeos. Assim, o padrão de controle conterá tipicamente DNA genômico (do mesmo fundo genético ao usado para obter o padrão 4C), digerido tanto com a enzima de restrição primária quanto secundária e marcado pelo mesmo método (por exemplo, iniciação aleatória) como o padrão 4C. Tal padrão de controle possibilita corrigir, sonda-
sonda, diferenças na eficiência de hibridização. A normalização de sinais de arranjo 4C para controlar sinais de arranjo possibilita expressar o resultado em termos de enriquecimento em relação a eventos aleatórios.

O padrão 4C marcado pode ainda ser hibridizado com um arranjo com ou sem uma amostra de controle diferencialmente marcada e com ou sem um ou mais outros padrões 4C diferencialmente marcados. Outros padrões 4C podem ser não relacionados a este padrão 4C, por exemplo, ele podem ser

obtidos de diferente tecido e/ou obtidos com um conjunto diferente de iniciadores inversos de PCR. Por exemplo, o primeiro padrão 4C pode ser material do paciente e o segundo padrão 4C pode ser obtido de um sujeito saudável ou uma amostra de controle.

Por causa dos rigorosos padrões de hibridização que espera-se encontrar para ré-arranjos genéticos, não será sempre necessário comparar sujeitos doentes com sujeitos saudáveis. Desta maneira, múltiplos (por exemplo, dois ou mais) padrões 4C, cada qual interrogando um locus diferente do mesmo paciente ou sujeito, podem ser hibridizados para um (por exemplo, um ou mais) arranjo.

Os padrões 4C podem ser diferencialmente marcados (por exemplo, com hibridização de duas ou múltiplas cores) e/ou podem ser identicamente marcados no caso de tal loci normalmente residir em diferentes cromossomos ou no mesmo cromossomo a uma distância grande o bastante para sobreposição mínima entre interação de sinais de DNA-DNA. Como um exemplo, material de um sujeito com leucemia de célula T pode ser processado para obter padrões 4C para *TCR α/δ* (marcados em uma cor, a fim de detectar translocações), e *MLL, TAL1, HOX11* e *LMO2* (cada um marcado na mesma segunda cor, a fim de detectar outros ré-arranjos genéticos). Estes cinco padrões 4C podem ser hibridizados para um arranjo, que permitirá a análise simultânea em múltiplos loci por um ré-arranjo genômico associado com a doença.

Para quantificação de frequências de interação, intensidades de sinal absoluto ou razões em relação à amos-

tra de controle podem também ser consideradas. Além do mais, sinais de sondas adjacentes no padrão do cromossomo linear podem ser usados para identificar regiões de interação cromossomais. Tal informação posicional é preferivelmente analisada ordenando as sondas no padrão do cromossomo linear e analisando as intensidades absolutas de sinal, ou razões com relação a sinais de padrão de controle, por abordagens de janela deslizando, usando, por exemplo, abordagens de média corrida ou mediana corrida.

10 Método de ensaio

Em um aspecto adicional da presente invenção, é fornecido um método de arranjo para identificar um ou mais agentes que modulam uma interação de DNA-DNA.

Da maneira aqui usada, o termo "modular" refere-se a prevenir, diminuir, suprimir, restaurar, elevar, aumentar ou de outra forma afetar a interação de DNA-DNA.

Em alguns casos, pode ser desejável avaliar dois ou mais agentes juntos para o uso em modular a interação de DNA-DNA. Nestes casos, ensaios podem ser facilmente modificados adicionando tal(s) agente(s) adicional(s) tanto simultaneamente quanto subsequentemente ao primeiro agente.

O método da presente invenção pode também ser uma seleção, por meio do que inúmeros agentes são testados para modular a atividade da interação de DNA-DNA.

25 Espera-se que os métodos de ensaio da presente invenção sejam adequados tanto para seleção em pequena escala quanto em grande escala de agentes, bem como em ensaios quantitativos.

Usos médicos de tais agentes terapêuticos estão no escopo da presente invenção, bem como os programas de desenvolvimento dos medicamento em si e composições farmacêuticas que compreendem tais agentes. Um programa de desenvolvimento
5 de medicamento pode, por exemplo, envolver tomar um agente identificado ou identificável pelos métodos aqui descritos, opcionalmente modificando-o (por exemplo, modificar sua estrutura e/ou fornecer uma composição inédita que compreende a dita fração) e desenvolver estudos adicionais (por exemplo, estudos de toxicidade e /ou estudos em atividade, estrutura ou função). Experiências podem ser realizadas em animais não humanos e podem eventualmente ser realizadas em seres humanos. Tais experiências incluirão de maneira geral determinar o(s) efeito(s) de níveis de dosagem diferentes.
10 Programas de desenvolvimento de medicamento podem utilizar computadores para analisar frações identificadas pela seleção (por exemplo, para prever estrutura e/ou função, para identificar possíveis agonistas ou antagonistas, para pesquisar outras frações que podem ter estruturas ou funções
15 similares, etc.).

Teste de diagnóstico

Atualmente, várias ré-arranjos genômicos continuam difíceis de detectar por técnicas citogenéticas moleculares disponíveis. Embora o arranjo técnica de hibridização genômica comparativa (arranjo-CGII) seja uma técnica recém-
25 desenvolvida para a detecção de amplificação e/ou deleções cromossomais com uma resolução de 35-300 Kb, esta técnica não é adequada para detectar translocações balanceadas e in-

versões cromossomais. Por outro lado, cariotipagem espectral (SKY) ou cariotipagem convencional é freqüentemente realizada em material do paciente para a detecção de translocações cromossomais bem como mudanças numéricas, mas a resolução para definir pontos de ruptura de translocação é baixa, freqüentemente 10-50 Mb e 540 Mb, respectivamente. Conseqüentemente, resultados obtidos por ambos métodos e especialmente SKY levarão a experimentos validações demorados e de muita mão-de-obra como hibridização *in situ* de fluorescência (FISH) e de estratégias de clonagem de ponto de ruptura molecular.

Tecnologia 4C envolve um procedimento que pode detectar qualquer ré-arranjo cromossomal com base nas frequências da interação entre seqüências alteradas de DNA fisicamente ligadas. Tecnologia 4C é, portanto, usada para a identificação de (recorrente) ré-arranjos cromossomais para a maioria dos males humanos / más formações congênitas múltiplas ou retardo mental. Uma vantagem importante da tecnologia 4C é que ela permite o mapeamento muito preciso do ponto de ruptura para uma região de somente diversos milhares de pares de base. Uma outra vantagem de tecnologia 4C é que nenhum conhecimento anterior é requerido na posição exata do ponto de ruptura, uma vez que pontos de ruptura serão detectáveis mesmo quando a seqüência 4C bait estiver localizada 1-5 Mb fora do ponto de ruptura. Isto tem também a vantagem que a mesma seqüência bait pode ser usada para a detecção de ré-arranjos cromossomais específicos cobrindo grandes áreas de ponto de ruptura. O mapeamento preciso de ré-arranjos ge-

nômicos pela tecnologia 4C facilitará bastante a identificação de gene(s) aberrantemente expresso(s) que dão suporte a doenças ou desordens genéticas que contribuirão importante-
mente para um melhor entendimento das correlações genótipo-
5 fenótipo, ajuda na tomada de decisão do tratamento e que adiciona informação importante de prognóstico.

Em uma modalidade da presente invenção, a fim de fornecer uma base para o diagnóstico ou prognóstico de doença, são estabelecidos valores normais ou padrões de um sujeito. Isto pode ser obtido testando amostras tomadas de sujeitos normais - tal como animais ou humanos. A frequência da interação de DNA-DNA pode ser quantificada para compará-la com uma série de diluição de controles positivos. Em seguida, valores de padrão obtidos de amostras normais podem
10 ser comparados com valores obtidos de amostras de sujeitos afetados ou potencialmente afetados por uma doença ou uma desordem. O desvio entre valores padrões e do sujeito estabelece a presença do estado da doença.

Tais ensaios de diagnóstico podem ser feitos sob
20 medida para avaliar a eficácia de um regime de tratamento terapêutico particular e podem ser usados em estudos animais, em experiências clínicas, ou em monitoramento de tratamento de um paciente individual. A fim de fornecer uma base para o diagnóstico de doença, pode-se estabelecer um perfil normal ou padrão para a interação de DNA-DNA. Valores de
25 padrão obtidos de amostras normais podem ser comparados com valores obtidos de amostras de sujeitos potencialmente afetados por uma desordem ou doença. O desvio entre valores pa-

drões e sujeito estabelece a presença do estado da doença. Se doença for estabelecida, um agente terapêutico existente pode ser administrado, e o perfil do tratamento ou valores podem ser gerados. Finalmente, o método pode ser repetido em
5 uma base regular para avaliar se os valores progridem a favor ou contra o modelo normal ou padronizado. Perfis de tratamento sucessivos podem ser usados para mostrar a eficácia do tratamento por um período de diversos dias ou diversos meses.

10 A tecnologia 4C detecta precisamente pelo menos 5 Mb de DNA genômico ligado em cis com a sequência de nucleotídeos que é analisada (ver Figura 2-3 e 5). Vantajosamente, a tecnologia 4C pode ser usada para detectar qualquer aberração genômica que é acompanhada por uma mudança em separação
15 ção do sítio genômico entre sequências ré-arranjadas e uma sequência 4C (bait) de escolha. Tal mudança pode ser, por exemplo, um aumento ou diminuição na separação do sítio genômico, ou pode ser uma subrepresentação (como em deleção(s)), ou sobrerrepresentação (como em duplicações) de sequências proximais (por exemplo, de até 15 Mb ou mais) com a
20 sequência 4C (bait). Tipicamente, tais aberrações genômicas ou ré-arranjos são uma causa ou estão associados com doenças - tais como câncer (por exemplo, leucemia) - e outras doenças genéticas ou congênitas aqui descritas.

25 Aberrações genéticas (por exemplo, aberrações genômicas ou cromossomais - tais como aberrações genômicas ou cromossomais balanceadas e/ou ou não balanceadas) incluem, mas sem limitações, ré-arranjos, translocações, inversões,

inserções, deleções e outras mutações de ácido nucléico (por exemplo, cromossomos) e também perdas ou ganhos de parte ou todos os cromossomos. Elas são a causa principal de desordens genéticas ou doenças, incluindo desordens congênitas e
5 doenças adquiridas - tal como males. Em muitos ré-arranjos, dois diferentes cromossomos são envolvidos. Desta maneira, genes (ou fragmentos de genes) são removidos do contexto fisiológico normal de um cromossomo particular e são localizados em um cromossomo recipiente, adjacente a genes não relacionados ou fragmentos de genes (freqüentemente oncogenes ou
10 proto-oncogenes).

Males podem incluir leucemias agudas, linfomas malignos e tumores sólidos. Exemplos não limitantes de alterações são t(14;18) que ocorrem freqüentemente em NHL;
15 t(12;21) que é freqüentemente encontrado em infância precursor-B-ALL; e a presença de 11q23 (MLL (leucemia linfóide mielóica ou leucemia de linhagem misturada) gene) aberrações em leucemias agudas.

O gene MLL na região do cromossomo 11q23 é envolvido em diversas translocações tanto em leucemias ALL quanto mielóide aguda (AML). Até hoje, pelo menos dez genes parceiros foram identificados. Algumas dessas translocações, -
tais como t(4;11) (q21; q23), t(11;19) (q23; p13) e t(1;11) (p32; q23), ocorrem predominantemente em ALL, onde outros,
25 como t(1;11) (q21; q23), t(2;11) (p21; q23), t(6;11) (q27; q23) e t(9;11) 10 (p22; q23) são mais freqüentemente observados em AML. Ré-arranjos envolvendo a região 11q23 ocorrem muito freqüentemente em leucemias agudas infantis (em torno

60-70 %), e em uma extensão muito menor em leucemias na infância e adulta (cada qual em torno 5 %).

Ré-arranjos em males linfóides freqüentemente envolvem genes Ig ou TCR. Exemplos incluem os três tipos de translocações (t(8;14), t(2;8), e t(8;22)) que são encontrados em linfomas de Burldtt, em que o gene MYC é acoplado à cadeia pesada Ig (IGH), kappa Ig (IGK), ou lambda Ig (IGL) segmentos do gene, respectivamente. Um outro tipo comum de translocação nesta categoria é t(14;18) (q32 ;q21) que é observado em cerca de 90 % de linfomas foliculares, um dos tipos NHL principais. Nesta translocação gene BCL2 é ré-arranjado em regiões dentro do locus IGH dentro ou adjacente aos segmentos do gene TH. O resultado dessa aberração do cromossomo é o sobreexpressão da proteína BCL2, que desempenha um papel como um fator de sobrevivência no controle do crescimento, inibindo a morte de célula programada.

O gene BCL2 consiste em três exons, mas esses são dispersos em uma grande área. Desses, o último exon codifica uma grande região não traduzida de 3' (3' UTR). Esta UTR 3' é uma das duas regiões em que muitos pontos de ruptura t(14; 18) são agrupados e é denominado a "região do ponto de ruptura principal"; a outra região de ponto de ruptura envolvida em translocações t(14; 18) fica localizada 20-30 kb à jusante do locus BCL2 e é denominada a "menor região de agrupamento". Uma terceira área de ponto de ruptura, BCL2 o VCR (região de agrupamento variante), é localizada no lado de 5' do locus BCL2 e está entre outras coisas envolvida em translocações variantes, isto é, t(2;18) e t(18;22), em que seg-

mentos de gene IGK e IGL são os genes parceiro.

Assim, a título de exemplo, tecnologia 4C pode ser aplicada à seleção de material do paciente para aberrações genéticas próximas ou em loci que foram escolhidos com base na sua associação freqüente com um dado fenótipo clínico. Exemplos não limitantes adicionais de tal loci são *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, loci de imonoglobulina, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCR α/δ* , *TCR/ δ* , *HOX* e outros loci em várias leucemias linfoblásticas.

Vantajosamente, se uma aberração genética é suspeita, a tecnologia 4C pode ser aplicada como a primeira e única seleção para verificar e mapear a presença da aberração da maneira aqui explicada.

Deteção de ré-arranjos genômicos

Em uma modalidade particularmente preferida da presente invenção, os métodos aqui descritos podem ser usados para a detecção de ré-arranjos genômicos.

Atualmente, ré-arranjos genômicos - tais como pontos de ruptura de translocação - são muito difíceis de detectar. Por exemplo, microarranjos de hibridização genômica comparativos (CGH) podem detectar diversos tipos de ré-arranjos, mas não conseguem detectar translocações. Se translocação é suspeita em um paciente, mas parceiros cromossomo são desconhecidos, cariotipagem espectral (SKY) pode ser realizado para encontrar parceiros de translocação e obter uma estimativa aproximada dos locais de ponto de ruptura. Entretanto, a resolução é muito fraca (freqüentemente não maior que ~50 Mb) e é freqüentemente necessário o mapea-

mento fino adicional (que é tanto demorado quanto caro). Isto é normalmente feito usando hibridização *in situ* de fluorescência (FISH), que novamente fornece resolução limitada. Usando FISH, pontos de ruptura podem ser localizados até +/-
5 50 kb da região em resolução máxima.

Frequências da interação de DNA-DNA basicamente são função da separação do sítio genômico, isto é, as frequências de interação DNA-DNA são inversamente proporcionais à distância linear (em quilobases) entre dois loci de DNA
10 presentes no mesmo padrão de DNA físico (Dekker et al., 2002). Assim, uma translocação, que cria um ou mais novos padrões DNA físicos, é acompanhada por interações de DNA-DNA alteradas próximo aos pontos de ruptura, e isto pode ser medido pela tecnologia 4C. Doenças com base nas translocações
15 são tipicamente causadas por interações de DNA-DNA aberrantes, já que as translocações são o resultado da ligação física (interação) de braços de cromossomo quebrados (DNA).

Desta maneira, para a detecção de translocações, tecnologia 4C pode ser usada para identificar as interações
20 de DNA-DNA que são diferentes entre sujeitos doentes e não doentes.

A título de exemplo, tecnologia 4C pode ser aplicada à seleção de material do paciente para translocações próximas aos loci que foram escolhidos com base nas suas associações frequentes com um dado fenótipo clínico aqui
25 descrito.

Se houver suspeita de translocação em um paciente,

mas os parceiros cromossomos forem desconhecidos, um mapeamento inicial pode ser realizado usando métodos atualmente disponíveis como cariotipagem espectral (SKY). Isto pode identificar os parceiros de translocação e fornecer uma estimativa bastante grosseira dos locais de ponto de ruptura (frequentemente não maior que ~50 Mb resolução). Tecnologia 4C pode em seguida ser aplicada, usando seqüências 'bait' nesta região localizada, por exemplo, a cada 2 Mb, 5 Mb, 10 Mb, 20 Mb (ou outros intervalos aqui descritos) para mapear finamente o ponto de ruptura e identificar por exemplo o(s) gene(s) que é(são) mal-expressos em decorrência da translocação.

Tipicamente uma translocação será identificada por meio de uma transição abrupta de freqüências da interação baixa para alta em um cromossomo sem ser o que contém a seqüência bait 4C, ou em qualquer lugar nesse mesmo cromossomo.

Em uma modalidade preferida, a amostra do sujeito é em um estado pré-maligno.

Em uma modalidade preferida, a amostra do sujeito consiste em líquidos amnióticos cultivados ou não cultivados obtidos por amniocentesis para diagnóstico pré-natal.

Em um desenho do arranjo preferido, sondas presentes em um arranjo simples representam o genoma completo de uma dada espécie em máxima resolução. Assim, arranjos para detectar translocações e similares pela tecnologia 4C contêm sondas aqui descritas complementares, de cada lado de cada sítio de reconhecimento da enzima de restrição primária no

genoma de uma dada espécie (por exemplo, humana).

Em um outro desenho preferido, sondas presentes em um arranjo simples representam o genoma completo de uma dada espécie, mas não em máxima resolução. Assim, arranjos para
5 detectar translocações e similares pela tecnologia 4C contêm sondas aqui descritas que são complementares a somente um lado de cada sítio de reconhecimento da enzima de restrição primária no genoma de uma dada espécie (por exemplo, humana).

10 Em um outro desenho preferido, sondas presentes em um arranjo simples representam o genoma completo de uma dada espécie; mas não em máxima resolução. Assim, arranjos para detectar translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos pela tecnologia 4C contêm son-
15 das aqui descritas que são complementares a um lado de cada outro sítio de reconhecimento da enzima de restrição primária ordenadas junto com o padrão linear do genoma de uma dada espécie (por exemplo, humana).

Assim, arranjos para detectar translocações, dele-
20 ções, inversões, duplicações e outros ré-arranjos genômicos pela tecnologia 4C contêm sondas aqui descritas que representam cada qual um fragmento de restrição simples, obtido após a digestão com uma enzima de restrição primária. Preferivelmente, isto é obtido ignorando cada segunda, terceira,
25 quarta, quinta, sexta, sétima, oitava, nona, décima, vigésima, trigésima, quadragésima, quinquagésima, sexagésima, septuagésima, octagésima, nonagésima ou centésima etc., sonda que hibridiza para o mesmo fragmento de restrição. Arranjos

para detectar translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos pela tecnologia 4C podem conter sondas aqui descritas que são distribuídas igualmente junto com os padrões do cromossomo lineares. Preferivelmente, isto é obtido ignorando uma ou mais sondas nas regiões genômicas que apresentam a mais alta densidade de sonda.

Em um outro desenho preferido, sondas presentes em um arranjo simples representam o genoma completo de uma dada espécie, mas não em resolução máxima. Assim, arranjos para detectar translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos pela tecnologia 4C contêm sondas aqui descritas complementares a um lado de cada terceiro, quarto, quinto, sexto, sétimo, oitavo, nono, décimo, vigésimo, trigésimo, quadragésimo, quinquagésimo, sexagésimo, septuagésimo, octagésimo, nonagésimo, ou um centésimo etc., sítio de reconhecimento da enzima de restrição primária ordenadas junto com o padrão linear do genoma de uma dada espécie (por exemplo, humana). Arranjos para detectar translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos pela tecnologia 4C pode conter sondas aqui descritas, que representam o genoma completo, mas com uma sonda simples cada 100 quilobases. Arranjos para detectar translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos pela tecnologia 4C pode contêm sondas aqui descritas que representam cada sítio de reconhecimento de enzima de restrição primária simples no genoma que pode ser representado por uma seqüência de sonda única.

Em um outro desenho de arranjo preferido, sondas

aqui descritas em um arranjo simples representam regiões genômicas de um dado tamanho - tal como cerca de 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb ou 10 Mb (por exemplo, cerca de 50 kb-10 Mb) em torno de todo loci conhecido a ser envolvido em translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos.

Em um outro desenho de arranjo preferido, sondas aqui descritas em um arranjo simples representam regiões genômicas de um dado tamanho - tal como cerca de 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb ou 10 Mb (por exemplo, cerca de 50 kb-10 Mb) em torno um seleção de loci conhecidos envolvidos em translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos. Seleções podem ser feitas em critérios educados, por exemplo, eles podem representar somente os loci que são implicados em um dado tipo de doença.

Em um outro desenho de arranjo preferido, sondas aqui descritas em um arranjo simples representam uma região genômica de interesse de, por exemplo, 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, ou 100 Mb (por exemplo, 100 kb-10 Mb) (parte de) um cromossomo ou múltiplos cromossomos, com cada sonda sendo representada múltiplas (por exemplo, 10, 100, 1000) vezes para permitir medições quantitativas de intensidades de sinal de hibridização em cada sequência de sonda.

Em um desenho experimental preferido, a sequência 4C (bait) é em cerca de 0 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb 10 Mb, 11 Mb, 12 Mb, 13 Mb, 14 Mb ou 15 Mb (por exemplo, cerca de 0-15 Mb) ou mais da sequência ré-arranjada atual (isto é, ponto de ruptura no caso de uma translocação).

Em uma hibridização preferida, dois padrões 4C marcados diferencialmente obtidos com uma sequência (4C bait) de um sujeito doente e não doente são hibridizados simultaneamente com o mesmo arranjo. Diferenças nas interações de DNA-DNA permitem a detecção do ponto de ruptura em *cis* (no mesmo cromossomo como o 4C bait) e em *trans* (no parceiro de translocação).

Em uma hibridização preferida, múltiplos padrões 4C diferencialmente marcados obtidos com uma sequência (4C bait) de sujeitos doentes e não doentes são hibridizados simultaneamente com o mesmo arranjo. Diferenças nas interações de DNA-DNA permitem a detecção do ponto de ruptura em *cis* (no mesmo cromossomo como o 4C bait) e em *trans* (no parceiro de translocação).

Vantajosamente, análise multicolor, em vez de duas cores, em microarranjos pode ser utilizada, permitindo a hibridização simultânea de mais que duas amostras com um arranjo simples. Desta maneira, hibridização multicolor pode ser usada em tecnologia 4C.

Em uma hibridização preferida, múltiplos padrões 4C diferencialmente marcados obtidos com uma sequência (4C

bait) de sujeito doentes e um padrão 4C diferencialmente marcado de um sujeito não doente são hibridizados simultaneamente com o mesmo arranjo. Diferenças nas interações de DNA-DNA permitem a detecção do ponto de ruptura em cis (no
5 mesmo cromossomo como o 4C bait) e em *trans* (no parceiro de translocação).

Em uma outra hibridização preferida, dois padrões 4C diferencialmente marcados do mesmo sujeito não doente, obtidos com duas seqüências diferentes (4C baits) que cada
10 representa um outro parceiro de translocação possível, são hibridizados simultaneamente com o mesmo arranjo. Agrupamentos de sinais de hibridização fortes observados no padrão linear dos cromossomos não relacionados com o cromossomo que leva a seqüência de interesse (4C-bait) identificarão o cromossomo
15 parceiro de translocação e o ponto de ruptura no parceiro de translocação.

Em uma outra hibridização preferida, múltiplos padrões 4C diferencialmente marcados do mesmo sujeito não doente, obtidos com múltiplas diferentes seqüências (4C-baits)
20 que representam cada qual um outro parceiro de translocação possível, são hibridizados simultaneamente com o mesmo arranjo. Agrupamentos de sinais de hibridização fortes observados no padrão linear dos cromossomos não relacionadas com o cromossomo que leva a seqüência de interesse (4C-bait) i-
25 dentificarão o cromossomo parceiro de translocação e seu ponto de ruptura para a seqüência de interesse.

Material usado para a detecção de translocações, deleções, inversões, duplicações e outros ré-arranjos genô-

micos pela tecnologia 4C pode ser obtido por reticulação (e processo adicional, descrito) de células vivas e/ou células mortas e/ou lisatos nucleares e/ou cromatina isolada etc. (aqui descrito) de sujeitos doentes e/ou não doentes.

5 *Detecção de Inversões*

Inversões (por exemplo, inversões balanceadas) não podem ser detectadas por métodos - tal como técnica de hibridização genômica comparativa - mas podem ser detectados pela tecnologia 4C particularmente quando a inversão (balanceada) está perto (por exemplo, até cerca de 1-15 Mb ou mais) da sequência 4C (bait).

Detecção de inversões (balanceadas) é baseada na identificação das interações de DNA-DNA que foram diferentes entre sujeitos doentes e não doentes. Inversões mudarão a posição relativa (em quilobases) no padrão de DNA físico de todas sequências da região ré-arranjada (mas o mais centralmente localizado) medida contra uma sequência próxima no mesmo cromossomo que é tomado como sequência 4C (bait). Uma vez que frequências da interação de DNA-DNA são inversamente relacionadas com a separação do sítio genômico, sujeitos doentes darão parceiros inversos de intensidades de hibridização para todas as sondas localizadas na região genômica ré-arranjada, comparado com um sujeito não doente. Assim, tecnologia 4C permite a identificação de posição e tamanho de inversões (balanceadas).

De acordo com este aspecto da presente invenção, um desenho do arranjo dedicado preferido compreende sondas em um arranjo simples representando regiões genômicas de um

dado tamanho - tais como cerca de 50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb, 3 Mb, 4 Mb, 5 Mb, 6 Mb, 7 Mb, 8 Mb, 9 Mb ou 10 Mb) (por exemplo, 50 kb-10 Mb) em torno do locus no qual a inversão ou outro ré-arranjo é suspeita.

5 Em um outro desenho do arranjo dedicado preferido, sondas em um arranjo simples representam regiões genômicas de um dado tamanho (50 kb, 100 kb, 200 kb, 300 kb, 400 kb, 500 kb, 1 Mb, 2 Mb etc.) em torno do locus no qual a inversão ou outro ré-arranjo é suspeita. Para análise quantitativa confiável de intensidades de sinal a quantidade de sonda
10 presente no arranjo é tipicamente em grande excesso para a quantidade de fragmentos cognatos que são hibridizados com o arranjo. Portanto, pode ser necessário ter cada sonda presente múltiplas vezes no arranjo (por exemplo, 10, 20, 50,
15 100, 1.000 vezes etc.). Além do mais, pode ser necessário para titular a quantidade de padrão que deve ser hibridizado com o arranjo.

Detecção de deleções

A detecção de deleções é baseada na identificação
20 das interações de DNA-DNA que foram diferentes entre sujeitos doentes e não doentes. Deleções resultarão na ausência de interações de DNA com uma sequência 4C (bait) localizada próximo (por exemplo, cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ou 15 Mb ou mais) na região deletada. Isto
25 pode resultar na ausência completa de sinais de hibridização para todas sondas localizadas na região ré-arranjada, se a deleção estiver presente em ambos alelos (homozigotos), ou uma redução para sujeitos doentes em função de não doen-

tes de intensidades de sinal, se a deleção estiver presente somente em um alelo (heterozigoto). Deleção coloca seqüências distais em proximidade imediata no padrão de DNA físico com a seqüência 4C analisada (bait), que resultará em sinais de hibridização maiores para sondas localizadas diretamente além da região deletada.

Deteção de duplicação(s)

Deteção de duplicação é tipicamente baseada na identificação das interações de DNA-DNA que são diferentes entre sujeitos doentes e não doentes. Sondas na região duplicada apresentarão maiores sinais de hibridização com uma seqüência 4C (bait) localizada próximo da região ré-arranjada (por exemplo, cerca de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 ou 15 Mb ou mais), comparado com sinais de um sujeito não doente de controle. Sondas além da região duplicada são adicionalmente separadas umas das outras da seqüência 4C e, conseqüentemente, apresentarão menores sinais de hibridização, comparados com sinais de um sujeito não doente de controle.

Preferivelmente, um aumento ou uma diminuição na freqüência de interação DNA-DNA para a amostra do sujeito comparada com o controle é indicativo de uma duplicação ou inserção.

Preferivelmente, um aumento na freqüência de interação de DNA-DNA para a amostra do sujeito comparada com o controle e/ou uma redução na freqüência de interação de DNA-DNA para regiões mais distantes é indicativo de uma duplicação ou inserção.

Diagnóstico Pré-natal

Vantajosamente, a tecnologia 4C pode também ser usada em diagnóstico pré-natal.

Ácido nucléico pode ser obtido de um feto usando
5 vários métodos que são conhecidos na tecnologia. A título de exemplo, amniocentesis pode ser usada para obter fluido amniótico do qual células fetais em suspensão são extraídas e cultivadas por diversos dias (Mercier & Bresson (1995) *Am. Gnt.*, 38, 151-157). Ácido nucléico das células pode ser em
10 seguida extraído. A coleção de chorial villi pode possibilitar a dispensa da etapa de cultura e evitar a coleta de fluido amniótico. Estas técnicas podem ser aplicadas previamente (até 7 semanas de gestação para a coleção de chorial villi e 13-14 semanas for amniocentesis), mas com um risco
15 de aborto ligeiramente maior.

Uma coleta direta de sangue fetal no nível do cordão umbilical pode também ser usada para obter ácido nucléico, mas tipicamente requer uma equipe de médicos de especializada nesta técnica (Donner et al. (1995) *Fetal Diagn.*
20 *Ther.*, 10, 192-199).

Vantajosamente, aberrações genéticas (por exemplo, aberrações genômicas ou cromossomais) - tais como rearranjos, translocações, inversões, inserções, deleções e outras mutações em cromossomos e ácido nucléico - podem ser
25 detectadas neste estágio.

Preferivelmente, aberrações genéticas (por exemplo, aberrações genômicas ou cromossomais) - tais como rearranjos, translocações, inversões, inserções, deleções e

outras mutações em cromossomos 21, 18, 13, X ou Y e também perdas ou ganhos de parte ou todos os cromossomos 21, 18, 13, X ou Y podem ser detectadas, uma vez que esses são os cromossomos em que ocorre a maioria de aberrações no feto.

5 *Determinação de sítios integração genômica*

A tecnologia 4C também permite a determinação de sítios de integração genômicos de vírus e transgênicos, etc., também quando múltiplas cópias são inseridas em diferentes posições no genoma (descrito na Figura 4).

10 *Predisposição Determinante para adquirir certas translocações*

Vantajosamente, a tecnologia 4C pode também ser aplicada aos sujeitos não doentes para medir o ambiente genômico de loci freqüentemente envolvido em aberrações genéticas. Desta maneira, é possível determinar a predisposição do sujeito em adquirir certas aberrações genéticas.

Assim, além dos usos médicos aqui descritos, a presente invenção pode ser usada em diagnóstico.

Sujeito

20 O termo "sujeito" inclui mamíferos - tais como animais e seres humanos.

Agente

O agente pode ser um composto orgânico ou outro composto químico. O agente pode ser um composto, que é obtível ou produzido por qualquer fonte adequada, seja natural ou artificial. O agente pode ser uma molécula de aminoácido, um polipeptídeo, ou um derivado químico destes, ou uma combinação destes. O agente pode ainda ser uma molécula de po-

linucleotídeo - que pode ser uma molécula sentido ou anti-sentido, ou um anticorpo, por exemplo, um anticorpo policlonal, um anticorpo monoclonal ou um anticorpo humanizado monoclonal.

5 Várias estratégias foram desenvolvidas para produzir anticorpos monoclonais com caracteres humano, que elimina a necessidade de uma linha celular humana de produção de anticorpo. Por exemplo, anticorpos monoclonais de camundongo usados foram "humanizados" ligando regiões variáveis de roe-
10 dor e regiões constantes de ser humano (Winter, G. and Milstein., C. (1991) *Nature* 349, 293-299). Isto reduz a imunogenicidade humano anticamundongo do anticorpo, mas imunogenicidade residual é retida em virtude da estrutura da região V estranha. Além do mais, a especificidade de ligação antígeno
15 é essencialmente que de o murino doador. Enxerto de CDR e estrutura de manipulação (EP 0239400) tem melhorado e refinado a manipulação de anticorpo até o ponto onde é possível produzir anticorpos murino humanizados que são aceitáveis para uso terapêutico em humanos. Anticorpos humanizados po-
20 dem ser obtidos usando outros métodos bem conhecidos na tecnologia (por exemplo, descrito em US-A-239400).

Os agentes podem ser anexados a uma entidade (por exemplo, uma molécula orgânica) por um ligante que pode ser um ligante bifuncional hidrolisável.

25 A entidade pode ser projetada ou obtida de uma biblioteca de compostos, que pode compreender peptídeos, bem como outros compostos, tais como pequenas molécula orgânicas.

A título de exemplo, a entidade pode ser uma substância natural, uma macromolécula biológica, ou um extrato feito de materiais biológicos tais como bactéria, fungo, ou animal (particularmente mamífero) células ou tecidos, uma
5 molécula orgânica ou um inorgânica, um agente sintético, um agente semi-sintético, um mimético estrutural ou funcional, um peptídeo, um peptidomimético, um peptídeo clivado de uma proteína total, ou um peptídeo sintetizado sinteticamente (tal como, a título de exemplo, tanto usando um sintetizador
10 de peptídeo quanto por técnicas recombinantes ou suas combinações, um agente recombinante, um anticorpo, um agente natural ou um não natural, um proteína de fusão ou equivalente destes e mutantes, derivados ou suas combinações.

Tipicamente, a entidade será um composto orgânico.
15 Para alguns casos, os compostos orgânicos compreenderão dois ou mais grupos hidrocarbila. Aqui, o termo "grupo hidrocarbila" significa um grupo que compreende pelo menos C e H e pode opcionalmente compreender um ou mais outros substituintes adequados. Exemplos de tais substituintes podem incluir
20 halo-, alcóxi-, nitro-, um grupo alquila, um grupo cíclico etc. Além da possibilidade de o substituinte ser um grupo cíclico, uma combinação de substituintes pode formar um grupo cíclico. Se o grupo hidrocarbila compreender mais que um C, então os carbonos não precisam necessariamente ser ligados
25 dos uns aos outros. Por exemplo, pelo menos dois dos carbonos podem ser ligados por meio de um elemento ou grupo adequado. Assim, o grupo hidrocarbila pode conter heteroátomos. Heteroátomos adequados serão aparentes aos versados na tec-

nologia e incluem, por exemplo, enxofre, nitrogênio e oxigênio. Para algumas aplicações, preferivelmente a entidade compreende pelo menos um grupo cíclico. O grupo cíclico pode ser um grupo policíclico, tal como um grupo policíclico não fundido. Para algumas aplicações, a entidade compreende pelo menos um dos ditos grupos cíclicos ligados a um outro grupo hidrocarbila.

A entidade pode conter grupos halo - tal como grupos de flúor, cloro, bromo ou iodo.

10 A entidade pode conter um ou mais dos grupos alquila, alcóxi, alquenila, alquilenos e alquênilos - que podem ser cadeia não ramificada ou ramificada.

Promedicamento

Versados na tecnologia percebem que a entidade pode ser derivada de um promedicamento. Exemplos de promedicamentos incluem certos grupos protegidos que podem não possuir atividade farmacológica como tal, mas podem, em certos exemplos, ser administrados (tal como oralmente ou parenteralmente) e depois metabolizados no corpo para formar uma entidade que é farmacologicamente ativa.

Promedicamentos adequados podem incluir, mas sem limitações, Doxorubicina, Mitomicina, Mustarda Fenol, Metotraxato, Antifolatos, Cloramfenicol, Camptotecina, 5-Fluoruracila Cianeto, Quinina, Dipiridamol e Paclitaxel.

25 Será mais apreciado que certas frações conhecidas como "pro frações", por exemplo, como descrito em "Desenho de Promedicamentos" por H. Bundgaard, Elsevier, 1985, podem ser colocados em funcionalidades apropriadas dos agentes.

Tais promedicamentos são também incluídos no escopo da invenção.

O agente pode ser na forma de um sal farmaceuticamente aceitável - tal como um sal de adição ácida ou um sal de base - ou um solvato destes, incluindo um hidrato deste. Para uma revisão em sais adequados ver Berge et al, J. Pharm. Sci., 1977, 66, 1-19.

O agente pode ser capaz de revelar outras propriedades terapêuticas.

10 O agente pode ser usado em combinação com um ou mais outros agentes farmaceuticamente ativos.

Se forem administradas combinações de agentes ativos, então as combinações de agentes ativos podem ser administradas simultaneamente, separadamente ou seqüencialmente.

15 Isômeros estéreos e geométricos

A entidade pode existir como esteroisômeros e/ou isômeros geométricos - por exemplo, a entidade pode possuir um ou mais centros assimétricos e/ou geométricos e assim pode existir em duas ou mais formas esteroisoméricas e/ou geométricas. A presente invenção contempla o uso de todos os esteroisômeros individuais e isômeros geométricos dessas entidades, e misturas destes.

Sal farmacêutico

25 O agente pode ser administrado na forma de um sal farmaceuticamente aceitável.

Sais farmaceuticamente aceitáveis são bem conhecidos pelos versados na tecnologia e, por exemplo, incluem os mencionados por Berge et al, em J. Pharm. Sci., 66, 1-19

(1977): Sais de adição ácida adequados são formados de ácidos que formam sais não tóxicos e incluem os sais cloridrato, bromidrato, iodidrato, nitrato, sulfato, bissulfato, fosfato, hidrogenfosfato, acetato, trifluoracetato, gluconato, lactato, salicilato, citrato, tartrato, ascorbato, succinato, maleato, fumarato, gluconato, formato, benzoato, metanossulfonato, etanossulfonato, benzenossulfonato e p-toluenossulfonato.

Quando uma ou mais frações acídicas são presentes, sais de adição de base farmacêuticamente aceitáveis adequados podem ser formados de bases que formam sais não tóxicos e incluem os sais alumínio, cálcio, lítio, magnésio, potássio, sódio, zinco, e aminas farmacêuticamente ativas tal como dietanolamina.

Um sal farmacêuticamente aceitável de um agente pode ser facilmente preparado misturando soluções do agente e o ácido ou base desejado, da maneira apropriada. O sal pode precipitar a partir da solução e ser coletado por filtração ou pode ser recuperado por evaporação do solvente.

O agente pode existir em forma polimórfica.

O agente pode conter um ou mais assimétricos carbono átomos e, portanto, existe em duas ou mais formas estereoisoméricas. Onde um agente contém um grupo alquênico ou alquênico, isomerismo *cis* (E) e *trans* (Z) pode também ocorrer. A presente invenção inclui o estereoisômeros do agente individual e, onde apropriado, as formas tautoméricas individual destes, juntos com misturas destes.

Separação de diaestereoisômeros ou isômeros *cis* e

trans pode ser obtida por técnicas convencionais, por exemplo, por cristalização fracional, cromatografia ou H.P.L.C. de uma mistura estereoisomérica do agente ou um sal adequado ou seu derivado. Um enantiômero individual do agente pode também ser preparado por resolução, tal como por H.P.L.C. do racemato correspondente usando um suporte quiral adequado ou por cristalização fracional dos sais diaestereoisoméricos formados pela reação do racemato correspondente com um ácido ou base opticamente ativos adequados, da maneira apropriada.

O agente pode também incluir todas as variações isotópicas adequadas do agente ou um sal farmacologicamente aceitável destes. Uma variação isotópica de um agente ou um sal farmacologicamente aceitável destes é definida como um no qual pelo menos um átomo é substituído por um átomo tendo o mesmo número atômico mas uma massa atômica diferente da massa atômica frequentemente encontrada na natureza. Exemplos de isótopos que podem ser incorporados no agente e sais farmacologicamente aceitáveis destes incluem isótopos de hidrogênio, carbono, nitrogênio, oxigênio, fósforo, enxofre, flúor e cloro tal como ^2H , ^3H , ^{13}C , ^{14}C , ^{15}N , ^{17}O , ^{18}O , ^{31}P , ^{32}P , ^{35}S , ^{18}F e ^{36}Cl , respectivamente.

Certas variações isotópicas do agente e sais farmacologicamente aceitáveis destes, por exemplo, aquelas em que um isótopo radioativo tal como ^3H ou ^{14}C é incorporado, são usadas em medicamento e/ou estudos de distribuição de tecido do substrato. Isótopos titulados, isto é, ^3H , e carbono-14, isto é, ^{14}C , são particularmente preferidos por sua facilidade de preparação e detectabilidade. Adicionalmente,

substituição com isótopos tal como deutério, isto é, $2H$, pode disponibilizar certas vantagens terapêuticas resultando em maior estabilidade metabólica, por exemplo, maior meia vida *in vivo* ou exigências dosagem reduzida e, portanto, podem ser preferidos em algumas circunstâncias. Variações isotópicas do agente e sais farmacologicamente aceitáveis desta invenção podem de maneira geral ser preparados por procedimentos convencionais usando variações isotópicas apropriadas de reagentes adequados.

Sal farmacologicamente ativo

O agente pode ser administrado como um sal farmacologicamente aceitável. Tipicamente, um sal farmacologicamente aceitável pode ser facilmente preparado usando um ácido ou base desejados, da maneira apropriada. O sal pode precipitar a partir da solução e ser coletado por filtração ou pode ser recuperado por evaporação do solvente.

Métodos de síntese química

O agente pode ser preparado por técnicas de síntese química.

Fica aparente aos versados na tecnologia que grupos funcionais sensíveis podem precisar ser protegidos e desprotegidos durante síntese de um composto da invenção. Isto pode ser obtido por técnicas convencionais, por exemplo, como descrito em "Protective Groups in Organic Synthesis" by T W Greene and P G M Wuts, John Wiley and Sons Inc. (1991), e por P.J.Kocienski, em "Protecting Groups", Georg Thieme Verlag (1994).

É possível, durante algumas das reações, que

quaisquer estereocentros presentes possam, em certas condições, ser racemizados, por exemplo, se for usada uma base em uma reação com um substrato tendo um centro ótico que compreende um grupo sensível a base. Isto é possível durante, 5 por exemplo, uma etapa de guanilação. Deve ser possível evitar problemas potenciais tal como este pela escolha de regimes de sequência de reação, condições, reagentes, de proteção/desproteção, etc. como é bem conhecido na tecnologia.

Os compostos e sais podem ser separados e purifi- 10 cados por métodos convencionais.

Separação de diastereômeros pode ser obtida por técnicas convencionais, por exemplo, por cristalização fracional, cromatografia ou H.P.L.C. de uma mistura esteroisomérica de um composto de fórmula (I) ou um sal adequado ou 15 seus derivados. Enantiômero individual de um composto de fórmula (I) pode também ser preparado de um intermediário opticamente puro correspondente ou por resolução, tal como por H.P.L.C. do racemato correspondente usando um suporte quiral adequado ou por cristalização fracional dos sais diastereoméricos formados pela reação do racemato correspon- 20 dente com um ácido ou base de forma adequada opticamente ativos.

O agente pode ser produzido usando métodos químicos para sintetizar o agente no todo ou em parte. Por exemplo, se o agente compreender um peptídeo, então o peptídeo 25 pode ser sintetizado por técnicas de fase sólida, clivado da resina, e purificado por cromatografia líquida preparativa de alto desempenho (por exemplo, Creighton (1983) Proteins

Structures And Molecular Principles, WH Freeman and Co, New York NY). A composição dos peptídeos sintéticos pode ser confirmada por análise ou seqüenciamento de aminoácido (por exemplo, o procedimento de degradação de Edman; Creighton, 5 *supra*).

Síntese de agentes inibidores de peptídeo (ou variantes, homólogos, derivados, fragmentos ou miméticos destes) pode ser realizada usando várias técnicas de fase sólida (Roberge J Y et al (1995) Science 269: 202-204), e síntese 10 se automatizada pode ser obtida, por exemplo, usando o Sintetizador de Peptídeo ABI 431 A (Perkin Elmer) de acordo com as instruções fornecidas pelo fabricante. Adicionalmente, as seqüências de aminoácido compreendendo o agente podem ser alteradas durante síntese direta e/ou combinadas usando 15 métodos químicos com uma seqüência de outras subunidades, ou qualquer parte destes, para produzir um agente variante.

Derivado químico

O termo "derivado" ou "derivatizado" da maneira aqui usada inclui modificação química de um agente. Ilustrativo de tais modificações químicas seria a substituição de 20 hidrogênio por um grupo halo, um grupo alquila, um grupo acila ou um grupo amino.

Modificação química

O agente pode ser um agente modificado - tal como, 25 mas sem limitações, um agente quimicamente modificado.

A modificação química de um agente pode tanto aumentar quanto reduzir a interação de ligação de hidrogênio, interação de carga, interação hidrofóbica, interação Van Der

Walls ou interação dipolo.

Em um aspecto, o agente pode agir como um modelo para o desenvolvimento de outros compostos (por exemplo, um padrão).

5 Composições farmacêuticas

Em um aspecto adicional, é fornecida uma composição farmacêutica que compreende um agente identificado pelo método de arranjo aqui descrito misturado com um veículo farmacêuticamente aceitável, diluente, excipiente ou adjuvante e/ou suas combinações.

Em um aspecto adicional, é fornecida uma composição de vacina que compreende um agente.

Em um aspecto adicional, é fornecido um processo de preparar uma composição farmacêutica que compreende misturar um agente identificado pelo ensaio com um diluente, veículo, excipiente ou adjuvante farmacêuticamente aceitável e/ou suas combinações.

Em um aspecto adicional, é fornecido um método de prevenir e/ou tratar uma doença que compreende administrar a um agente ou uma composição farmacêutica ou uma vacina para um sujeito.

As composições farmacêuticas podem ser para uso humano ou animal em medicina humana e veterinária e compreenderá tipicamente qualquer um ou mais de um diluente, veículo, ou excipiente farmacêuticamente aceitável. Veículos ou diluentes aceitáveis para uso terapêutico são bem conhecidos na tecnologia farmacêutica, e são descritos, por exemplo, em Remington's Pharmaceutical Sciences, Mack Publishing Co. (A.

R. Gennaro edit 1985). A escolha de veículo, excipiente ou diluente farmacêutico pode ser selecionada no sentido de entender a via de administração e prática farmacêutica padrão. As composições farmacêuticas podem compreender como o veículo, o excipiente ou diluente qualquer adequado ou além deles ligante(s), lubrificante(s), agente(s) de suspensão, agente(s) de revestimento, agente(s) solubilizante(s).

Conservantes, estabilizadores, corantes e ainda agentes flavorizantes podem ser fornecidos na composição farmacêutica. Exemplos de conservantes incluem benzoato de sódio, ácido sórbico e ésteres de ácido p-hidroxibenzóico. Antioxidantes e agente de suspensões podem ser também usados.

Pode haver diferentes exigências de composição/formulação dependendo dos diferentes sistemas de distribuição. A título de exemplo, a composição farmacêutica da presente invenção pode ser formulada para ser administrada usando uma minibomba ou por uma via mucosal, por exemplo, como uma aspersão nasal ou aerossol para inalação ou solução de ingestão, ou parenteralmente em que a composição é formulada por uma forma injetável com ar, para distribuição, por exemplo, por uma via intravenosa, intramuscular ou subcutânea. Alternativamente, a formulação pode ser projetada para ser administrada por inúmeras vias.

Se o agente tiver que ser administrado mucosalmente por meio da mucosa gastrintestinal, ele tem que ficar estável durante o trânsito por todo o trato gastrintestinal; por exemplo, ele deve ser resistente a degradação proteolítica.

tica, estável a pH ácido e resistente aos efeitos detergentes da biliar.

Quando apropriado, as composições farmacêuticas podem ser administradas por inalação, na forma de um supositório ou pessário, topicamente na forma de uma loção, solução, creme, unguento ou pó de talco, pelo uso de um adesivo de pele, oralmente na forma de comprimidos contendo excipientes tal como amido ou lactose, ou em cápsulas ou óvulos tanto sozinhos como em misturas com excipientes, ou na forma de elixires, soluções ou suspensões contendo agentes flavorizantes ou corantes, ou as composições farmacêuticas podem ser injetadas parenteralmente, por exemplo, intravenosamente, intramuscularmente ou subcutaneamente. Para administração parenteral, as composições podem ser bem usadas na forma de uma solução aquosa estéril que pode conter outras substâncias, por exemplo, sais ou monossacarídeos suficientes para tornar a solução isotônica com sangue. Para além do mais bucal ou sublingual, as composições podem ser administradas na forma de comprimidos ou pastilhas que podem ser formulados de uma maneira convencional.

Os agentes podem ser usados em combinação com uma ciclodextrina. Ciclodextrinas são conhecidas por formar complexos de inclusão e não inclusão com moléculas de medicamento. A formação de um complexo medicamento-ciclodextrina pode modificar a solubilidade, taxa de dissolução, propriedade de biodisponibilidade e/ou estabilidade de uma molécula de medicamento. Complexos de medicamento-ciclodextrina são de maneira geral usadas para a maioria das formas de dosagem

e vias de administração. Como uma alternativa a complexação direta com o medicamento, a ciclodextrina pode ser usada como um aditivo auxiliar, por exemplo, as um veículo, diluente ou solubilizante. Ciclodextrinas alfa, beta e gama são exemplos mais normalmente usados e adequados e estão descritos em WO-A-91/11172, WO-A-94/02518 e WO-A-98155148.

Se o agente for uma proteína, então a dita proteína pode ser preparada *in situ* no sujeito que está sendo tratado. A este respeito, seqüências de nucleotídeos que codificam a dita proteína podem ser distribuídas pelo uso de técnicas não virais (por exemplo, pelo uso de lipossomas) e/ou técnicas virais (por exemplo, pelo uso de vetores retrovirais) de maneira tal que a dita proteína seja expressa a partir da dita seqüência de nucleotídeos.

As composições farmacêuticas da presente invenção podem também ser usadas em combinação com tratamentos convencionais.

Administração

O termo "administrado" inclui distribuição por técnicas viral ou não viral. Mecanismos de distribuição viral incluem, mas sem limitações, vetores adenovirais, vetores virais adeno associados (AAV), vetores virais de herpes, vetores retrovirais, vetores lentivirais, e vetores baculovirais. Mecanismos de administração não viral incluem transfecção mediada por lipídeos, lipossomas, imunolipossomas, anfifilos faciais catiônicos (CFAs) e suas combinações.

Os componentes podem ser administrados sozinhos, mas de maneira geral serão administrados como uma composição

farmacêutica - por exemplo, quando os componentes estão misturados com um excipiente, diluente ou veículo farmacêuticamente adequado selecionado com relação à via de administração visada e prática farmacêutica padrão.

5 Por exemplo, os componentes podem ser administrados na forma de comprimidos, cápsulas, óvulos, elixires, soluções ou suspensões, que podem conter agentes flavorizantes ou corantes, para aplicações de liberação imediata, atrasada, modificada, pulsada ou controlada.

10 Se o produto farmacêutico for um comprimido, então os comprimidos podem conter excipientes tais como celulose microcristalina, lactose, citrato de sódio, carbonato de cálcio, fosfato de cálcio dibásico e glicina, desintegrantes tal como amido (preferivelmente amido de milho, batata ou
15 tapioca), glicolato de amido de sódio, croscarmellose sódica e certos silicatos complexos, e aglutinantes de granulação, tais como polivinilpirrolidona, hidroxipropilmetilcelulose (EIPMC), hidroxipropilcelulose (CPC), sacarose, gelatina e acácia. Adicionalmente, agentes de lubrificação tais como
20 estearato de magnésio, ácido esteárico, beenato de glicerila e talco podem ser incluídos.

 Composições sólidas de um tipo similar podem também ser empregadas como cargas em cápsulas de gelatina. Excipientes preferidos a este respeito incluem lactose, amido,
25 uma celulose, açúcar de leite ou polietileno glicol de alto peso molecular. Para suspensões aquosas e/ou elixires, o agente pode ser combinado com vários agentes adoçantes ou flavorizantes, matéria de cor ou corantes, com agentes de

emulsificação e/ou suspensão e com diluentes tais como água, etanol, propileno glicol e glicerina e suas combinações.

As vias de administração (distribuição) podem incluir, mas sem limitações, um ou mais de oral (por exemplo, como um comprimido, cápsula, ou como uma solução de ingestão), tópica, mucosal (por exemplo, como uma aspersão nasal ou aerossol para inalação), nasal, parenteral (por exemplo, por uma forma injetável), gastrintestinal, intraespinhal, intraporitoneal, intramuscular, intravenosa, intrauterina, intraocular, intradérmica, intracraniana, intratraqueal, intravaginal, intracerebroventricular, intracerebral, subcutânea, oftálmica (incluindo intravitreal ou infracameral), transdérmica, retal, bucal, vaginal, epidural, sublingual.

Níveis de dose

Tipicamente, um médico determinará a real dosagem que será mais adequada para um sujeito individual. O nível de dose específica e frequência de dosagem para qualquer paciente particular podem variar e dependerão de uma variedade de fatores, incluindo a atividade do composto específico empregado, da estabilidade metabólica e extensão de ação desse composto, da idade, peso corpóreo, saúde geral, sexo, dieta, modo e tempo de administração, taxa de excreção, combinação de medicamento, da severidade da condição particular, e da terapia individual em andamento.

Formulação

O(s) componente(s) pode(m) ser formulado(s) em uma composição farmacêutica, tal como misturando com um ou mais de um veículo, diluente ou excipiente adequado, usando téc-

nicas que são conhecidas na tecnologia

Doença

Aspectos da presente invenção podem ser usados para o tratamento e/ou prevenção e/ou diagnóstico e/ou prognóstico de uma doença - tais como as listadas em WO-A-98/09985.

Para facilidade de referência, parte dessa lista é agora fornecida: atividade inibitória de macrófago e/ou inibitória de célula T e, assim, atividade antiinflamatória; atividade antiimune, isto é, efeitos inibitórios contra uma resposta imune celular e/ou humoral, incluindo uma resposta não associada com inflamação; doenças associadas com vírus e/ou outros patógenos intracelulares; inibição da capacidade de macrófagos e células T aderirem nos componentes da matriz extracelular e fibronectina, bem como expressão de receptor fas supra-regulado em células T; inibição de reação imune e inflamação indesejadas, incluindo artrite reumatóide, inflamação associada com hipersensibilidade, reações alérgicas, asma, lupus eritomatoso sistêmico, doenças do colágeno e outras doenças autoimunes, inflamação associada com aterosclerose, arteriosclerose, doença do coração aterosclerótico, lesão por reperfusão, parada cardíaca, infarto do miocárdio, desordens inflamatórias vasculares, síndrome do desconforto respiratório ou outras doenças cardiopulmonares, inflamação associada com úlcera péptica, colite ulcerativa e outras doenças do trato gastrointestinal, fibrose hepática, cirrose hepática ou outras doenças hepáticas, tireoidite ou outras doenças glandulares, glomerulonefrite ou outras doenças re-

nais e urológicas, otite ou outras doenças otorrinolaringicas, dermatite ou outras doenças dérmicas, doenças periodontais ou outras doenças dentárias, orquite ou orquiepididimite, infertilidade, trauma orquidal ou outras doenças testiculares imune relacionadas, disfunção placentária, insuficiência placentária, aborto habitual, eclampsia, pré-eclampsia e outras doenças ginecológicas imune e/ou inflamatória relacionadas, uveíte posterior, uveíte intermediária, uveíte anterior, conjuntivite, coriorretinite, uveroretinite, neurite ótica, inflamação intraocular, por exemplo, retinite ou edema macular cistóide, oftalmia simpatética, esclerite, retinite pigmentosa, componentes imune e autoimune de doença de fondus degenerativa, componentes inflamatórios de trauma ocular, inflamação ocular causada por infecção, vítreoretinopatias proliferativa, neuropatia ótica isquêmica aguda, cicatriz excessiva, por exemplo, após operação de filtração de glaucoma, reação imune e/ou inflamação contra implantes oculares e outras doenças oftálmicas imune e inflamatórias relacionadas, inflamação associada com doenças ou condições ou desordens autoimunes onde tanto no sistema nervoso central (CNS) quanto em qualquer outro órgão, supressão imune e/ou de inflamação seria benéfica, doença de Parkinson, complicação e/ou efeitos colaterais do tratamento de doença de Parkinson, demência relacionada a AIDS, encefalopatia relacionada a HIV, doença de Devic, Sydenham chorea, doença de Alzheimer e outras doenças degenerativas, condições ou desordens do CNS, componentes inflamatórios de acidentes vasculares, síndrome pós-pólio, componentes imune e

inflamatórios de desordens psiquiátricas, mielite, encefalite, pan-encefalite esclerosante subaguda, encefalomielite, neuropatia aguda, neuropatia subaguda, neuropatia crônica, síndrome de Guillain-Arre, Sydenham chorea, miastenia grave, pseudotumor cerebral, síndrome de Down, doença de Huntington, esclerose lateral amiotrófica, componentes inflamatórios de compressão CNS ou trauma CNS ou infecções do CNS, componentes inflamatórios de atrofias e distrofias musculares, e doenças, condições ou desordens imune e inflamatória relacionadas do sistema nervoso central e periférico, inflamação pós-traumática, choque séptico, doenças infecciosas, complicações inflamatórias ou efeitos colaterais de cirurgia, complicações de transplante de medula óssea ou outras complicações e/ou efeitos colaterais de transplante, complicações inflamatória e/ou imune e efeitos colaterais de terapia genética, por exemplo, por causa de infecção com um portador viral, ou inflamação associada com AIDS, para suprimir ou inibir uma resposta imune humoral e/ou celular, para tratar ou atenuar doenças proliferativas de monócitos ou leucócitos, ex., leucemia, reduzindo a quantidade de monócitos ou linfócitos, para a prevenção e/ou tratamento de rejeição de enxerto em casos de transplante de células, tecido e órgãos naturais ou artificiais tais como córnea, medula óssea, órgãos, lentes, marcapassos, tecido de pele natural ou artificial. Desordens específicas relacionadas a câncer incluem, mas sem limitações: tumores sólidos; tumores levados pelo sangue tais com leucemias; metástase de tumor; tumores benignos, por exemplo, hemangiomas, neuromas acústicos, neuro-

fibromas, tracomias, e granulomas piogênicos; artrite reumatóide; psoríase; doenças angiongênicas oculares, por exemplo, retinopatia diabética, retinopatia de prematuridade, degeneração macular, rejeição de enxerto de córnea, glaucoma
 5 neovascular, fibroplastia retrolental, rubeose; Síndrome de Osler-Webber; angiogênese miocárdica; neovascularização de placas; telangiectasia; juntas hemofílicas; angiofibroma; granulação de ferida; colaterais coronarianas; colaterais cerebrais; más formações arteriovenosas; angiogênese do mem-
 10 bro isquêmico; glaucoma neovascular; fibroplasia retrolental; neovascularização diabética; doenças relacionadas a *Helicobacter*, fraturas, vasculogêneses, hematopoiese, ovulação, menstruação e placentação.

Preferivelmente, a doença é câncer, tais como leu-
 15 cemia linfocítica aguda (ALL), leucemia mielóide aguda (AML), câncer adrenocortical, câncer anal, câncer da bexiga, câncer do sangue, câncer dos ossos, tumor no cérebro, câncer de mama, câncer sistema genital feminino, câncer sistema genital masculino, linfoma do sistema nervoso central, câncer
 20 cervical, rabdomiosarcoma da infância, sarcoma da infância, leucemia linfocítica crônica (CLL), leucemia mielóide crônica (CML), câncer do cólon e retal, câncer do cólon, câncer do endométrio, sarcoma do endométrio, câncer do esôfago, câncer dos olhos, câncer da vesícula biliar, câncer gástri-
 25 co, câncer do trato gastrointestinal, leucemia da célula capilar, câncer da cabeça e pescoço, câncer hepatocelular, doença de Hodgkin, câncer hipofaríngeo, sarcoma de Kaposi, câncer do rim, câncer da laringe, leucemia, câncer do fígado.

do, câncer do pulmão, histiocitoma fibrosa maligna, timoma maligna, melanoma, mesotelioma, mieloma múltipla, mieloma, câncer da cavidade nasal e seios paranasais, câncer nasofaríngeal, câncer do sistema nervoso, neuroblastoma, linfoma de não Hodgkin, câncer da cavidade oral, câncer orofaríngeal, osteosarcoma, câncer do ovário, câncer do pâncreas, câncer de paratireóide, câncer do pênis, câncer da faringe, tumor da hipófise, neoplasma da célula plasma, linfoma primária CNS; câncer de próstata, câncer retal, sistema respiratório, retinoblastoma, câncer de glândula salivar, câncer de pele, câncer do intestino delgado, sarcoma do tecido macio, câncer do estômago, câncer do estômago, câncer do testículo, câncer da tireóide, câncer sistema urinário, sarcoma uterina, câncer vaginal, sistema vascular, macroglobulinemia de Waldenstrom e tumor de Wilms.

Estojo

Os materiais para uso nos métodos da presente invenção são idealmente adequados para preparação de estojos.

Um estojo como este pode compreender recipientes, com cada um ou mais dos várias reagentes (tipicamente em forma concentrada) usados nos métodos aqui descritos, incluindo, for exemplo, uma enzima de restrição primária, uma enzima de restrição secundária, um agente de reticulação, uma enzima de ligação (por exemplo, um ligase) e um agente para reverter a reticulação (por exemplo, proteinase K).

Oligonucleotídeos podem também ser fornecidos em recipientes que podem ser em qualquer forma, por exemplo, liofilizada, ou em solução (por exemplo, uma água destilada

ou solução tamponada), etc.

Em um aspecto preferido da presente invenção, é fornecido um estojo que compreende um conjunto de sondas aqui descritas, um arranjo e opcionalmente um ou mais marcadores.

Um conjunto de instruções será também tipicamente incluído.

Usos

Vantajosamente, a presente invenção pode ser usada a fim de obter informação a cerca da organização espacial de seqüência de nucleotídeos - tais como loci genômicos *in vitro* ou *in vivo*.

A título de exemplo, a tecnologia 4C pode ser usada para estudar a organização tridimensional de um ou mais loci do gene. De maneira particular, esta tecnologia pode ser usada para estudar o papel de um ou mais fatores de transcrição na organização tridimensional de um ou mais loci do gene.

A título de exemplo adicional, a tecnologia 4C pode ser usada para estudar o papel de fatores de elementos de DNA *trans*-atuantes e *cis*-reguladores.

A título de exemplo adicional, a tecnologia 4C pode ser usada para estudar regulamento gene de longo alcance *in vitro* ou *in vivo*.

A título de exemplo adicional, a tecnologia 4C pode ser usada para estudar proximidade e interação infra cromossomal.

A título de exemplo adicional, a tecnologia 4C po-

de ser usada para estudar proximidade e interação inter-cromossomal.

A título de exemplo adicional, a tecnologia 4C pode ser usada para identificar seqüências de nucleotídeos que
5 funcionam com um promotor, intensificador, silenciador, isolante, região de controle de locus, origem de replicação, MAR, SAR, centrômero, telômero ou qualquer outra seqüência de interesse em uma rede regulatória.

A título de exemplo adicional, a tecnologia 4C pode
10 de ser usada para identificar genes responsáveis por um fenótipo (doença) em casos onde acontece de uma mutação e/ou deleção afetar um elemento regulador distante e seu mapeamento portanto não consegue fornecer tal informação.

A título de exemplo adicional, a tecnologia 4C pode
15 de ser usada para reconstruir eventualmente a conformação espacial de loci do gene, grandes regiões genômicas ou mesmo cromossomos completos.

A título de exemplo adicional, a tecnologia 4C pode
20 de ser usada para definir seqüências âncora potenciais que sustenta certos cromossomos juntos no espaço nuclear.

A título de exemplo adicional, a tecnologia 4C pode ser usada para reconstruir eventualmente o posicionamento dos cromossomos em alta resolução uns com relação aos outros.

25 A título de exemplo adicional, a tecnologia 4C pode ser usada em diagnóstico (por exemplo, diagnóstico pré-natal) para detectar ou identificar ré-arranjos genômicos e/ou aberrações - tais como translocações, deleções, inver-

sões, duplicações.

Técnicas gerais de metodologia de dna recombinante

A presente invenção emprega, a menos que de outra forma indicada, técnicas convencionais de química, biologia molecular, microbiologia, DNA recombinante e imunologia, que estão de acordo com a capacidade de um versado na tecnologia. Tais técnicas são explicadas na literatura. Ver, por exemplo, J. Sambrook, E. F. Fritsch, and T. Maniatis, 1989, *Molecular cloning: A Laboratory Manual*, Second Edition, Books 1-3, Cold Spring Harbor Laboratory Press; Ausubel, F. M. et al. (1995 and periodic supplements; *Current Protocols in Molecular Biology*, ch. 9, 13, and 16, John Wiley & Sons, New York, N.Y.); B. Roe, J. Crabtree, e A. Kahn, 1996, *DNA Isolation and Sequencing: Essential Technics*, John Wiley & Sons; M. J. Gait (Editor), 1984, *Oligonucleotide Synthesis: A Practical approach*, Irl Press; and, D. M. J. Lilley and J. E. Dahlberg, 1992, *Methods of Enzymology: DNA Structure Part A: Synthesis and Physical Analysis of DNA* Methods in Enzymology, Academic Press. Cada um desses textos gerais é aqui incorporado pela referência.

A invenção será agora adicionalmente descrita a título de exemplo, que deve servir para ajudar versados na tecnologia realizar a invenção e não devem ser de maneira nenhuma para limitar o escopo da invenção.

25 Exemplo 1

Materiais & Métodos

Tecnologia 4C

As etapas iniciais do procedimento da tecnologia

3C foram realizadas da maneira previamente descrita (Splinter *et al.* (2004). *Methods Enzymol* 375, 493-507 (2004), produzindo produtos de ligação entre fragmentos HindIII. Este padrão 3C ligado a HindIII (-50 µg) foi digerido toda a noite a 100 ng/µL com 50 U de uma enzima de restrição de corte freqüente secundária, sendo tanto DpnII (HS2, Rad23A) quanto NlaIII (β principal). Para evitar restrições em formação de círculo de DNA (Rippe *et al.* (1995) *Trends Biochem Sci* 20, 500-6), foi tomado cuidado para escolher uma enzima de restrição secundária que não corte dentro de cerca de 350-400 bp do sítio de restrição HindIII que demarca o fragmento de restrição de interesse (isto é, o 'bait'). Após a digestão da enzima de restrição secundária, DNA foi fenol extraído, etanol precipitado e subsequentemente ligado a baixa concentração (amostra de 50 µg em 14 mL usando 200 U ligase (Roche), 4 horas a 16 °C) para promover DpnII ou formação de círculo DpnII. Produtos de ligação foram fenol extraído e etanol precipitado, usando glicogeno (Roche) como um veículo (20 µg/mL). Os círculos de interesse foram linearizados digerindo toda a noite com um 50 U de uma enzima de restrição terciária que corta o bait entre o sítio de reconhecimento de enzima de restrição primária e secundária, usando as seguintes enzimas de restrição: Spel (HS2), PstI (Rad23A) e PflmI (β principal). Esta etapa de linearização foi realizada para facilitar hibridização subsequente do iniciador durante a primeira rodada de amplificação de PCR. Produtos digeridos foram purificados usando uma coluna de remoção de nucleotídeos (250) QIAquick (Qiagen).

Reações de PCR foram realizadas usando o sistema PCR de Expand Long Template (Roche), usando condições cuidadosamente otimizadas para assegurar amplificação linear de fragmentos medidos em até 1,2 kb (80 % de fragmentos de PCR 4C são menores que 600 bp). Condições do PCR foram como a seguir: 94 °C por 2 minutos, 30 ciclos de 94 °C por 15 segundos, 55 °C por 1 minuto e 68 °C por 3 minutos, seguido por uma etapa final de 68 °C por 7 minutos. A quantidade máxima de padrão que ainda apresenta faixa linear de amplificação foi determinada. Para isto, diluições seriadas de padrão foram adicionadas às reações de PCR, material do DNA amplificado foi corrido em um gel de agarose e produtos PCR foram quantificados usando software ImageQuant. Tipicamente, 100-200 ng de padrão por 50 µL da reação de PCR deu produtos na faixa linear de amplificação. 16 a 32 reações de PCR foram agrupadas e purificaram este padrão 4C usando o sistema de remoção de nucleotídeos QIAquick (250) (Qiagen). Padrão 4C purificado foi marcado e hibridizado com arranjos de acordo com padrão dos protocolos ChIP-chip (Nimblegen Systems of Iceland, LLC). DNA genômico diferencialmente marcado que foi digerido com a enzima primária e secundária usado no procedimento 4C, serviu como um padrão de controle para corrigir as diferenças nas eficiências de hibridização. Para cada experimento duas amostras independentemente processadas foram marcadas com orientações de corante alternados.

Seqüências de iniciadores 4C usadas:

HS2:5'- ACTTCCTACACATTAACGAGCC-3',

5'- GCTGTTATCCCTTTCTCTTCTAC-3'

Rad23A:5'- TCACACGCGAAGTAGGCC-3',

5'- CCTTCCTCCACCATGATGA-3'

β principal:5' -AACGCATTTGCTCAATCAACTACTG-3',

5' - GTTGCTCCTCACAITTTGCTTCTGAC-3'

Arranjos 4C

Arranjos e análise foram baseados em m34 construídos por NCBI. Sondas (60-mer) foram selecionadas das seqüências 100 bp à montante - e à jusante dos sítios HindIII. O teor de CG foi otimizado em 50 %, para sinais de hibridização uniformes. Para prevenir hibridização cruzada sondas que tiveram qualquer similaridade com repetições altamente abundantes (RepBase 10,09) foram removidas do conjunto de sonda. Além do mais, sondas que deram mais que dois bons êxitos de BLAST no genoma foram também removidas do conjunto de sonda. Alinhamentos de seqüência foram realizados usando MegaBLAST (Zhang et al. (2000) *J Comput Biol* 7, 203-14) usando a configuração padrão. Um bom êxito definido como um alinhamento de 30 nt ou mais.

20 Análise de dados 4C

A razão de sinal amostra-4C/genômico DNA foi calculado para cada sonda e os dados foram visualizados com software SignalMap fornecido por Nimblegen Systems. Dados foram analisados usando o pacote R (<http://www.r-project.org>), Spotfire and Excel. Razões de hibridização não processada apresentou agrupamentos de 20-50 sinais 4C positivos junto com o padrão do cromossomo. Para definir estes agrupamentos, foi aplicada uma média corrida. Vários tamanhos de

janela foram usados, variando de 9-39 sondas, dos quais todas identificaram os mesmos agrupamentos. Os resultados apresentados foram baseados em um tamanho de janela de 29 sondas (em média 60 kb) e foram comparados com a média corrida realizada em dados randomizados. Isto foi feito para cada arranjo separadamente. Conseqüentemente, todas as medições foram apreciadas com relação a amplitude e ruído para esse arranjo específico. A Taxa de falsa descoberta (FDR), definida como (número positivos falsos) / (número de positivos falsos + número de positivos verdadeiros) foi determinada como a seguir: (número de positivos no conjunto randomizado) / (número de positivos nos dados). O nível do patamar foi determinado usando uma abordagem de cima para baixo para o valor mínimo estabelecido para o qual: $FDR < 0,05$.

Em seguida, experimentos biológicos duplicados foram comparados. Janelas que atenderam o patamar em ambas duplicatas foram consideradas positivas. Comparando dados randomizados, nenhuma janela ficou acima do patamar em ambas duplicatas. Janelas positivas diretamente adjacentes ao padrão do cromossomo foram ligadas (não é permitida nenhuma folga), criando áreas positivas.

Análise de expressão

Para cada tecido, três microarranjos independentes foram realizados de acordo com o protocolo Affymetrix (ca-mundongo 430 - 2 arranjos). Dados foram normalizados usando RMA ca-tools; www.bioconductor.org) e para cada conjunto de sonda foi calculada a média das medições dos três microarranjos. Além do mais, quando múltiplos conjuntos de sonda

representaram o mesmo gene, foi também calculada a média. Mas5calls (Affy library: www.bioconductor.org) foi usado para estabelecer campos calls "presente", "ausente" e "marginal". Genes com um campo call "presente" em todos os três

5 arranjos e um valor de expressão maior que 50 foram denominados expresso. "Genes específicos de Fígado fetal" foram classificados como genes que atenderam nossos critérios de ser expresso em fígado fetal e tiveram valores de expressão mais que cinco vezes superiores comparados com cérebro fe-

10 tal. Para fornecer uma medida de atividade transcricional geral em torno de cada gene, foi aplicada uma soma corrida. Para isto, nós usamos valores de expressão log-transformados. For cada gene nós calculamos a soma da expressão de todos genes encontrados em uma janela 100 kb á

15 montante do início e 100 kb à jusante da extremidade do gene, incluindo o próprio gene. Valores resultantes para genes ativos encontrados nas regiões 4C positivas (n = 124, 123 e 208 respectivamente para HS2 em fígado, Rad23A em cérebro e Rad23A em fígado) foram comparados com os valores obtido pa-

20 ra genes ativos fora áreas 4C positivas (n = 153, 301 e 186, respectivamente, onde n = 153 corresponde ao número de genes não interagentes ativos, presentes entre a região mais interagente centromérica e o telômero do cromossomo 7); os dois grupos foram comparados usando um teste Wilcoxon de soma de

25 poços unilaterais.

Sondas FISH

Os clones BAC seguintes (BACPAC Resources Centre) foram usado; RP23-370E12 para Hbb-1, RP23-317H16 para chr.7

a 80,1 Mb (ou agrupamento de gene), RP23-334E9 para Uros, RP23-32C19 para chr.7 a 118,3 Mb, RP23-143F10 para chr.7 a 130,1, RP23-470N5 para chr.7 a 73,1 Mb, RP23-247L11 para chr.7 a 135,0 Mb (ou agrupamento de gene), RP23-136A15 para Rad23A, RP23-307P24 para chr.8 at 21,8 Mb e RP23-460F21 para chr.8 a 122,4 Mb. Para uma sonda específica do centrômero do cromossomo 7 nós usamos clone PI 5279 (Genome Ststems Inc.) que anela no segmento de DNA D7Mit21. Sondas marcadas de primo aleatório foram preparadas usando BioPrime Array CGH genomic Labeling System (Invitrogeno). Antes da marcação, DNA foi digerido com DpnII e purificado com um conjunto de DNA limpo e concentrador 5 (pesquisa Zymo). DNA digerido (300 ng) foi marcado com SpectrumGreen dUTP (Vysis) ou flúor Alexa 594 dUTP (sondas Moleculares) e purificado por meio um estojo de purificação de GFX PCR DNA e Gel Band (Amersham Biosciences) para remover nucleotídeos não incorporados. Especificidade de sondas marcadas foi testada em esfregaços de metáfase preparado de células ES murina.

Crio-FISH

Crio-FISH foi realizado da maneira descrita anteriormente ⁵. Resumidamente, fígado E14.5 e cérebro foram fixados por 20 minutos em 4 % de paraformaldeído/HEPES250 mM, pH 7,5 e cortados em pequenos blocos de tecido, seguido por uma outra etapa de fixação de 2 horas em 8 % de paraformaldeído a 4 °C. Blocos de tecido fixado foram imersos em 2,3 M de sacarose por 20 minutos a temperatura ambiente, montados em um suporte de corpo de prova e congelados rapidamente em nitrogênio líquido. Blocos de tecido foram armazenados em

nitrogênio líquido até o seccionamento. Criosseções de ultratrina de aproximadamente 200 nm foram cortadas usando um Reichert Ultramicrotome E equipado com crio-acessório (Leica). Usando um laço cheio com sacarose, seções foram transferidas para a lamínula e armazenadas a -20 °C. Para hibridização, seções foram lavadas com PBS para remover sacarose, tratadas com 250 ng/mL de RNase em 2 x SSC por 1 hora a 37 °C, incubadas por 10 minutos em 0,1 M de HCL, desidratadas em uma série de etanol e desnaturadas por 8 minutos a 80 °C em 70 % de formamida/2 x SSC, pH 7,5. Seções foram novamente desidratadas diretamente antes da hibridização da sonda. 500 ng de sonda marcada foi co-precipitada com 5 µg de DNA de camundongo Cot1 (Invitrogen) e dissolvida em hybmix (50 % de formamida, 10 % de sulfato dextrano, 2 x SSC, 50 mM de tampão de fosfato, pH 7,5). Sondas foram desnaturadas por 5 minutos a 95 °C, ré-aneladas por 30 minutos a 37 °C e hibridizadas por pelo menos 40 horas a 37 °C. Após lavagem pós-hibridização, núcleos foram contramanchados com 20 ng/mL DAPI (Sigma) em PBS/0,05 % Tween-20 e montados em reagente antidesbotamento Prolong Gold (Sondas Moleculares).

Imagens foram coletadas com um microscópio de epifluorescência Zeiss Axio Imageador ZI (x 100 plan apochromat, 1,4 oil objeativo), equipado com uma câmera CCD e software Isis FISH Imaging System (Metasystems). Um mínimo de 250 β-globinas ou alelos Rad23A foi analisado e pontuado como sobreposição ou não sobreposição com BACs localizados em outro lugar no genoma, por uma pessoa que não conhece a combinação da sonda aplicada às seções. Testes de ajuste em re-

plicata (estatística G) ⁶ foram realizados para avaliar significância de diferenças entre valores medidos para regiões positivas 4C contra negativas 4C. Revisão geral do resultado é fornecida na Tabela 2.

5 Embora tenhamos encontrado diferenças estatisticamente significativas entre frequências de interação de fundo (0,4-3,9 %) e reais (5-20,4 %), pode ficar claro que frequências medidas por crio-FISH são menores que as medidas por outros usando protocolos FISH diferentes. O seccionamen-
10 to pode separar alguns loci de interação e medições crio-FISH subestimarão ligeiramente portanto as verdadeiras frequências de interação. Por outro lado, procedimentos atuais 2D- e 3D FISH sobrestimarão estas porcentagens por causa da resolução limitada na direção z. No futuro, técnicas de mi-
15 croscopia melhoradas em combinação com sondas FISH mais específicas revelarão melhores frequências de interação verdadeiras.

Exemplo 2

O procedimento 3C (isto é, fixação de formaldeído, (primária) digestão de enzima de restrição, re-ligação de
20 fragmentos de DNA reticulados e purificação de DNA) é realizado essencialmente como descrito (Splinter et al., (2004) *Methods Enzymol.* 375: 493-507), produzindo uma mistura de DNA ('padrão 3C') contendo fragmentos de restrição que são
25 ligados em virtude de eles estarem originalmente perto do espaço nuclear.

PCR inverso é realizada para amplificar todos os fragmentos ligados a um dado fragmento de restrição ('bait';

escolhido em virtude de ele conter um promotor, intensificador, isolante, região de anexação de matriz, origem de replicação ou qualquer outra primeira (visada) sequência de nucleotídeos).

5 Para isto, círculos de DNA são criados digerindo o padrão 3C com uma enzima de restrição secundária (preferivelmente, um cortador freqüente reconhecendo seqüências de tetra- ou penta-nucleotídeos), seguido por ligação em condições diluídas de maneira que interações intra-moleculares
10 sejam favorecidas. Para minimizar uma propensão à formação de círculo por causa das restrições topológicas (Rippe et al, (2001) Trends in Biochem. Sciences 26, 733-40), deve ser escolhida uma enzima de restrição secundária que corta preferivelmente o bait a X350-400 bp do sítio de restrição primária. Para aumentar a eficiência e reprodutibilidade de amplificação de PCR inversa, círculos são mais bem linearizados antes da amplificação de PCR por uma enzima de restrição (por exemplo, um cortador de 6 ou mais bp) que corta o bait entre o sítio de restrição primário e secundária de diagnóstico.
15
20 tico.

Digestão do padrão 3C com a enzima de restrição secundária, circularização por meio de ligação em condições diluídas e linearização de círculos contendo bait são realizados sob condições padrão para tais manipulações de DNA
25 produzir um padrão de DNA para amplificação de PCR inverso ('padrão 4C').

Desta maneira, 10 µg de padrão 3C são digeridos em 100 µL com 20 U da enzima de restrição secundária (por

toda a noite), seguido por inativação de calor da enzima e purificação do DNA. Ligação é realizada em 10 mL. (1 ng/ μ L DNA) com 50 U T4 ligase (4 horas a 16 °C, 30 minutos a RT), seguido por purificação de DNA. Finalmente, a linearização dos círculos de interesse é feita em 100 μ L com 20 U de enzima de restrição (por toda a noite), seguido novamente por purificação de DNA.

Para PCR inversa, dois iniciadores específicos bait são projetados, cada qual o mais próximo possível do sítio de reconhecimento de enzima de restrição primária e secundária diretamente vizinha, respectivamente, e cada qual com sua extremidade 3' voltada para fora para que a extensão se dê imediatamente nos sítios de restrição em um fragmento ligado ao bait. PCR inversa com estes iniciadores é preferivelmente realizado em 100-400 ng do DNA de padrão 4C (por 50 μ L da reação da mistura de PCR), para incluir um número máximo de eventos de ligação por reação de PCR. Nós realizamos PCR inversa aplicando o Sistema PCR de Padrão Expand Long (Roche), usando tampão 1 de acordo com procedimentos do fabricante.

São realizados os ciclos de PCR seguintes:

1. 2 minutos 94 °C
2. 15 segundos 94 °C
3. 1 minuto 55 °C
4. 3 minutos 68 °C
5. etapa repetida 2-4 29 x (ou nada entre 25-40 x)
6. 7 minutos 68 °C
7. fim.

Gel eletroforese é realizado para analisar reprodutibilidade entre reações de PCR individuais. Tipicamente, parceiros produtos idênticos devem ser obtidos.

A fim de obter material suficiente para marcar por
5 iniciação aleatória e hibridização do arranjo, múltiplas reações de PCR (cada obtida após 30 ciclos de PCR) podem ser agrupadas, (em vez de aumentar o número de ciclos de PCR por reação). Como uma alternativa para marcação com primo aleatório, nucleotídeos marcados podem ser incorporados nos últimos
10 ciclos de PCR (por exemplo, 30 ciclos (sem marcação) + 10 ciclos (marcação)).

Exemplo 3

Detecção de translocação usando tecnologia 4C

Tecnologia 4C é usada para medir as frequências da
15 interação para a dada sequência X presente em um dado cromossomo A em células de um sujeito saudável e em células de um paciente que carrega uma única translocação recíproca, entre cromossomo A e B com o ponto de ruptura perto da sequência X (apresentado na Figura 9).

20 Em células normais, esta análise revela sinais de elevada hibridização (isto é, interações frequentes com X) para (quase) toda sonda localizada em 0,2-10 Mb da sequência X em cromossomo A (o tamanho atual da região cromossomal mostrando fortes sinais de reticulação depende na maioria
25 das vezes da complexidade da amostra que foi hibridizada com o arranjo). Em outro lugar no mesmo cromossomo A, bem como em outros cromossomos, não é observada nenhuma região grande como esta (no padrão linear de DNA) de sondas com eleva-

vados sinais de hibridização.

Em células de paciente, entretanto, sinais de hibridização com todas as sondas do cromossomo A localizadas no outro lado do ponto de ruptura são reduzidas para ~50 % (uma cópia do cromossomo A é ainda intacta e produzirá sinais normais), ao passo que uma única (isto é, não presente em células normais) concentração de sinais de elevada hibridização é observada para sondas vizinhas do ponto de ruptura em cromossomo B. De fato, a transição abrupta entre sondas mostrando no contra sinais de hibridização fortes em cromossomo B revela a localização do ponto de ruptura em cromossomo B.

Exemplo 4

Resultado de análise de tecnologia 4C

Tecnologia 4C foi usada para caracterizar o ambiente genômico da região de controle do locus β -globina do camundongo (LCR), focado em um fragmento de restrição contendo seu sítio 2 hipersensível (HS2). O LCR é um forte elemento regulador de transcrição específico de eritróide requerido para altos níveis de expressão do gene β -globina. O locus β -globina está presente em cromossomo 7 na posição 97 Mb, onde ele reside em um grande agrupamento 2,9 Mb, de genes do receptor olfativo que são transcritos somente em neurônios olfativos. Interações foram analisadas em dois tecidos: fígado fetal E14.5, onde o LCR é ativo e os genes β -globina são altamente transcritos, e cérebro fetal E14.5, onde o LCR é inativo e os genes globina são silenciosos. Em ambos tecidos, a grande maioria de interações foi encontrada

com seqüências em cromossomo 7 e muito poucas interações LCR foram detectadas com seis cromossomos não relacionados (8, 10, 11, 12, 13, 14) (Figura 13a). Os sinais mais fortes em cromossomo 7 foram encontrados em uma região 5-10 Mb centralizados em torno da posição cromossomal de β -globina, de acordo com a idéia de que freqüências da interação são inversamente proporcionais à distância (em pares de base) entre seqüências de DNA fisicamente ligadas. Não foi possível interpretar quantitativamente as interações nesta região. Nós inferimos que estas seqüências próximas foram juntamente com β -globina tão freqüentes que sua grande sobre-representação em nossas amostras de hibridização saturaram as sondas correspondente. Isto foi confirmado quando nós realizamos hibridização com amostras diluídas 1 : 10 e 1 : 100 e observamos que a intensidade do sinal foi reduzida em sondas fora e na borda, mas não nesta região (dados não apresentados).

O procedimento 4C forneceu dados altamente reprodutíveis. A figura 2b-c apresenta razões não processadas de sinais 4C com relação a sinais de hibridização de controle para duas regiões em cromossomo 7 1,5 Mb, grosseiramente 25 Mb e 80 Mb fora do gene β -globina. Neste nível de resolução, os resultados das amostras independentemente processadas foram quase idênticos. Tanto em fígado fetal quanto em cérebro, agrupamentos de sinais positivos foram identificados em cromossomo 7, freqüentemente em locais cromossomais dezenas de megabases fora de β -globina. Estes agrupamentos tipicamente consistem em minimamente 20-50 sondas com maiores razões de sinal justapostos no padrão do cromossomo (Figura

13b-c). Cada sonda na análise de arranjo um evento de ligação independente. Além do mais, somente duas cópias de fragmento de restrição HS2 estão presentes por célula, cada uma das quais pode somente ligar com um outro fragmento de restrição. Portanto, a detecção de eventos de ligação independentes com 20 ou mais fragmentos de restrição vizinhos indicam fortemente que o locus correspondente faz contato com o LCR de β -globina em múltiplas células.

Para determinar a significância estatística destes agrupamentos, dados de experimentos individuais foram ordenados em mapas cromossomais e analisados usando um algoritmo de média corrida com um tamanho de janela de aproximadamente 60 kb. A distribuição da média corrida de dados misturados aleatoriamente foi usada para estabelecer um valor de limite, permitindo uma falsa taxa de descoberta de 5 %. Esta análise identificou 66 agrupamentos em fígado fetal e 45 em cérebro que foram encontrados reprodutíveis em experimentos em duplicata (Figura. 13d-f). De fato, alta resolução FISH confirmou que tais agrupamentos verdadeiramente representam loci que interagem freqüentemente (ver a seguir).

Assim, a tecnologia 4C identifica loci de interação de longo alcance pela detecção de eventos de ligação independentes com múltiplos fragmentos de restrição agrupados em uma posição cromossomal.

Uma série de experimentos 4C completamente independentes foi realizada com um conjunto de iniciadores de PCR inversa diferente que investigou o ambiente genômico do gene β principal, localizado ~50 kb à jusante de HS2. Em um

fígado fetal, o gene β principal é altamente transcrito e freqüentemente recebe contato do LCR. Agrupamentos de interações de longo alcance quase idênticos com β principal como com HS2 foram encontrados, tanto em fígado fetal quanto em

5 cérebro, substanciando adicionalmente que estes loci fazem contato freqüentemente com o locus β -globina (Figura 17).

Exemplo 5

O locus β -globina ativo e inativo ocupa ambiente genômico distinto.

10 Uma Comparação entre o dois tecidos revelou que o locus β -globina ativamente transcrito em fígado fetal interage com um conjunto de loci completamente diferente de sua contraparte transcricionalmente silenciosa em cérebro ($r=0,03$; Correlação de Spearman's Rank) (Figura 13f). Isto

15 exclui que resultados sejam influenciados pela composição de seqüência das sondas. Em fígado fetal, os segmentos do DNA de interação foram localizados em uma região 70 Mb centralizada em torno do locus β -globina, com a maioria (40/66) localizada em direção ao telômero do cromossomo 7. Em fetal

20 cérebro, loci de interação foram encontrados em similares ou mesmo grandes distâncias de β -globina comparados ao fígado fetal e com a grande maioria de interações (43/45) localizada em direção ao centrômero do cromossomo 7. Estes dados demonstraram que o locus β -globina ativo e inativo faz contato

25 com diferentes partes do cromossomo 7.

Seis outros cromossomos (8, 10, 11, 12, 13 e 14) foram representados nos microarranjos. Sinais fortes de hibridização nestes cromossomos foram raros, aparecem tipica-

mente isolados no padrão de DNA linear e freqüentemente foram ausentes dos experimentos em duplicata. Também, níveis de média corrida através destes cromossomos nunca chegaram perto reprodutivelmente dos níveis pontuados para cromossomo 7 (Figura 19). Assim, outros dados mostraram que o locus β -globina basicamente faz contato com os loci em outro lugar no mesmo cromossomo, de acordo com a localização preferida deste locus dentro de seu próprio território do cromossomo. Nós notamos que o locus β -globina foi também presente no arranjo (cromossomo 11) e não pontuou positivo para interação com β -globina, de acordo com a demonstração recente por FISH que camundongo α e β -globina não se encontraram freqüentemente no espaço nuclear (Brown, J. M et al. (2006) *J Cell Mol* 172, 177-87).

A fim de um melhor entendimento da relevância das interações de longo alcance observadas no cromossomo 7, nós comparamos os loci de interação com as posições cromossomais de genes. Além do mais, análise de arranjo de expressão Affymetrix foi realizada para determinar atividade de transcrição nestas posições nos dois tecidos. Embora o tamanho médio de áreas de interação em fígado fetal e cérebro seja comparável (183 kb e 159 kb, respectivamente), diferenças drásticas foram observadas em seu teor e atividade de gene. Em fígado fetal, 80 % dos loci de interação de β -globina contiveram um ou mais genes ativamente transcritos, ao passo que em cérebro fetal a grande maioria (87 %) não demonstrou atividade de gene detectável (Figura 15). Assim, o locus β -globina é incorporado em um ambiente genômico muito diferen-

te nos dois tecidos. Em cérebro, onde o locus não é ativo, ele primeiramente faz contato com os loci transcrpcionais silenciosos localizados em direção ao centrômero do cromossomo 7. Em fígado fetal, onde o locus é altamente ativo, ele interage preferencialmente com regiões ativamente transcritas localizadas mais proeminentemente em direção ao lado telomérico do cromossomo 7. Importaneamente, tecnologia 4C identificou ambos Uros e Eraf, (~30 Mb distante de β -globina) como genes interagindo com o locus β -globina em fígado fetal, de acordo com observações prévias feitas por FISH (Osborne, C. S. et al. (2004) *Nat Genet* 36, 1065-71 (2004)). Interessantemente, em cérebro em contatos foram observados com dois outros agrupamentos do gene receptor olfativo presente em cromossomo 7 que foram localizados a cada lado de β -globina, e 17 e 37 Mb distante dela.

Nem todas regiões transcritas em cromossomo 7 interagem com o locus β -globina ativo em fígado fetal. Portanto, nós pesquisamos denominador que compartilha exclusivamente pelo loci de interação mas não por outras regiões ativas em fígado fetal. Os genes β -globina, Uros e Eraf são todos genes específicos de eritróide que podem ser regulados pelo mesmo conjunto de fatores de transcrição, e é uma idéia atrativa que esses fatores coordenam expressão de seus genes visados no espaço nuclear. Nós comparados dados arranjos de expressão Affymetrix de fígado fetal E14.5 com o de cérebro fetal para identificar genes preferencialmente expressos (>5 vezes mais) em fígado fetal. Foram classificados 28 % dos genes ativos como estes em cromossomo 7 como "fígado fetal

específico", dos quais 25 % foram encontrados em uma área co-localizada. Assim, nós não encontramos nenhum enriquecimento de genes de "fígado fetal específico" nas áreas co-localizadas. Mais importantemente, 49 de 66 (74 %) regiões de interação não continham um "fígado fetal-específico" e concluiu-se portanto que nossos dados não mostraram evidência para expressão coordenada de genes específicos do tecido no espaço nuclear. Os genes β -globina são transcritos em taxas excepcionalmente altas e questionou-se em seguida se o locus preferencialmente interage com outras regiões de atividade transcricional alta, sendo tanto genes altamente expressos quanto áreas com uma alta densidade de genes ativos. Usando contas Affymetrix como uma medida para atividade do gene, nós realizamos um algoritmo de soma de corrida para medir diversas atividades transcricionais nas regiões 200 kb em torno dos genes ativamente transcritos. Esta análise revelou que atividade transcricional em torno genes de interação não foi maior que em torno de genes de não de interação ativos em cromossomo 7 ($p = 0,9867$; Wilcoxon Rank sum).

Exemplo 6

O ambiente genômico de um gene zelador é amplamente conservado entre os tecidos

Em seguida, foi investigado se um gene que é similarmente expresso em ambos tecidos também muda seu ambiente genômico. Rad23A é um gene ubiqüitariamente expresso que reside em um denso agrupamento de genes basicamente de genes zeladores em cromossomo 8. Tanto em fígado fetal E14.5 quanto em cérebro, este gene e muitos de seus vizinhos diretos

são ativos. Análise 4C foi realizada e identificou muitas interações de longo alcance com loci de até 70 Mb distante de Rad23A. Importaneamente, interações com Rad23A foram altamente correlacionadas entre fígado fetal e cérebro ($\tau =$ 0,73; correção de Spearman's Rank) (Figura 15a). Uma marca compartilhada desses loci foi novamente que eles continham genes ativamente transcritos. Assim, em ambos tecidos grosseiramente 70 % contiveram pelo menos um gene ativo (Figura 15b-c). Regiões em torno de genes de interação revelam níveis mais altos estatisticamente significativos de atividade do gene comparados com genes ativos em outro lugar no cromossomo, como determinado por um algoritmo de soma de corrida ($p < 0,001$ para ambos tecidos). Assim, diferente do locus β -globina, o gene Rad23A que é localizado em uma região rica em gene preferencialmente interage em distância com outras regiões cromossômicas de maior atividade transcricional. Foi observado por FISH que a área cromossômica contendo Rad23A reside basicamente na borda de (90 %) ou fora (10 %) de seu território do cromossomo (não publicado, D. Noordermeer, M. Branco, A. Pombo and W. de Laat). Entretanto, a análise 4C revelou somente interações infra-cromossômicas e nenhuma área no cromossomo 7, 10, 11, 12, 13 ou 14 reprodutível atendeu nossos rigorosos critérios para interação. Assim, Rad23A é basicamente envolvido em interações intracromossômicas que são similares em dois tecidos muito diferentes. Se Rad23A tiver loci vizinhos preferidos nestes cromossomos não relacionados, eles não interagem frequentemente o bastante para ser detectados nas condições aqui usadas para a tecnologia

4C.

Exemplo 7

Validação de tecnologia 4C por microscopia de alta resolução

5 Para validar o resultado obtido pela tecnologia 4C, experimentos crio-FISH foram realizados. Crio-FISH é uma técnica de microscopia recentemente desenvolvida, que tem a vantagem sobre protocolos 3D-FISH atuais, que preserva melhor a ultra-estrutura nuclear oferecendo, portanto, melhor
10 resolução no eixo-z pela preparação de crio-seções ultrafinas (Branco, M. R.. & Pombo, A (2006). *PLoS Biol* 4, e138). Dados de 4C foram verificados medindo agora β -globina frequente ou alelos Rad23A (sempre $n > 250$) co-localizados com mais que 15 regiões cromossomais selecionadas em seções ultrafinas de 200 nm preparadas de fígado e cérebro E14.5. Importaneamente, todas as frequências da interação medidas por crio-FISH estavam em perfeito acordo com o resultado 4C (Figura 17). Por exemplo, regiões distantes que foram identificadas para interagir com β -globina pela tecnologia 4C co-localizada mais frequentemente do que as áreas intervenientes não detectadas por 4C (7,4 % e 9,7 %, contra 3,6 % e 3,5 %, respectivamente). Também, os dois agrupamentos do receptor de gene olfativo distantes identificados pela tecnologia 4C para interagir com β -globina em cérebro fetal, mas não em
20 frequências de colocação pontuadas no fígado respectivamente de 12,9 % e 7 % em cérebro, contra 3,6 % e 1,9 % em seções do fígado. Em resumo, frequências de colocação medidas por loci positivamente identificados pela tecnologia

4C foram todas significativamente maiores que as frequências medidas pelos loci de fundo G . Nós concluimos que tecnologia 4C identifica verdadeiramente loci de interação de DNA. Finalmente, nós usamos crio-FISH para demonstrar que loci identificados interagindo com β -globina também fazem contato freqüentemente uns com os outros. Isto foi verdadeiro para duas regiões ativas separadas por grande distância cromossomal em fígado fetal (Figura 19) bem como para dois agrupamentos de gene OR inativos separados longe um do outro no cromossomo em cérebro (Figura 17). Interessantemente, freqüentes contatos entre estes dois agrupamentos de gene OR distantes foram também encontrados em fígado fetal, onde eles não interagirão com o agrupamento de gene OR que contém o locus β -globina ativamente transcrito. Estes dados indicam que interações nucleares entre agrupamentos de gene OR distintos não foram uma peculiaridade do tecido do cérebro fetal analisado. Tentando especular que tais contatos espaciais facilitam a comunicação entre os muitos genes OR necessários para assegurar que é transcrito somente um alelo simples por neurônio olfativo (Shykind, B. (2005) *Hum Mol Genet* 14 Spec No 1, R33-9.

Exemplo 8

Organização nuclear de domínios de cromatina ativos e inativos

As observações aqui descritas demonstram que não somente regiões genômicas ativas, mas também inativas, formam regiões distintas no espaço nuclear que envolvem muitos contatos de longo alcance, sugerindo fortemente que cada

segmento de DNA tem seu próprio conjunto de interações preferido. Nossos dados sugerem que quando o locus β -globina é ligado, ele deixa um ambiente transcricional silencioso genômico e entra em uma área nuclear onde interações com domínios ativos são favorecidos. Presume-se que um reposicionamento drástico como este mediante ativação transcricional pode bem ser uma marca distinta somente de genes específicos de tecido que atingiram um certo nível de expressão e, mais importantemente, ficaram isolados dos outros genes ativos no padrão de cromossomo linear, como é o caso para β -globina. Propõe-se que a rede extensiva de interações de longo alcance que são identificadas tanto entre loci genômicos inativos quanto entre ativos, reflita diferenças célula-para-célula em conformações de cromossomo mais que sendo uma consequência de movimentos dinâmicos durante interfase (Chakalova et al. (2005) *Nat Rev Genet* 6, 669-77 (2005)). Presumidamente, diferentes graus de descondensação após a divisão celular impulsionam as regiões genômicas ativas para fora da cromatina inativa (Gilbert, N. et al. (2004) *Cell* 118, 555-66 (2004)) e contatos entre loci distante de composição de cromatina similar são estabilizados basicamente por meio das afinidades entre proteínas ligadas a cromatina. Justaposição espacial entre loci distantes pode ser funcional, mas pode também simplesmente ser a consequência dos parceiros de desdobramento de um cromossomo. Enquanto loci individuais podem mover em um volume nuclear restrito, a conformação geral de um cromossomo será amplamente mantida por todo o ciclo celular e exigindo divisão celular para restabelecimento. Esta

idéia está de acordo com estudos de imageamento de célula vivas que mostram movimentos restritos de loci de DNA alvos no interior nuclear (Chubb et al. (2002) *Curr Biol* 12, 439-45 (2002)) e se ajustam bem aos estudos que mostram que a
5 informação da posição de cromatina nuclear é frequentemente propagada durante a divisão celular sem ser conservado na população de células (Essers, J. et al. *Mol Biol Cell* 16, 769-75 (2005); Gerlich, D. et al.. *Cell* 112, 751-64 (2003)).

Aspectos adicionais 1

10 Aspectos adicionais da presente invenção são apresentados a seguir nos parágrafos numerados.

1. Um conjunto de sondas complementares a cada lado de cada sítio de reconhecimento da enzima de restrição primária no genoma de uma dada espécie (por exemplo, huma-
15 na).

2. Um conjunto de sondas complementares a somente um lado de cada sítio de reconhecimento da enzima de restrição primária no genoma de uma dada espécie (por exemplo, humana),

20 3. Um conjunto de sondas complementares a um lado de cada outro sítio de reconhecimento da enzima de restrição primária ordenado ao longo do padrão linear do genoma de uma dada espécie (por exemplo, humana).

4. Um conjunto de sondas complementares a um lado
25 de cada terceiro, quarto, quinto, sexto, sétimo, oitavo, nono, décimo, vigésimo, trigésimo, quadragésimo, quinquagésimo, sexagésimo, septuagésimo octagésimo, nonagésimo ou um centésimo sítio de reconhecimento da enzima de restrição

primária ordenadas junto com o padrão linear do genoma de uma dada espécie (por exemplo, humana).

5 5. Um conjunto de sondas representando regiões genômicas de um dado tamanho (por exemplo, cerca de 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 ou 10 Mb) (por exemplo, 50 kb-10 Mb) em torno todos loci os conhecidos para ser envolvidos em translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos.

10 6. Um conjunto de sondas representando regiões genômicas de um dado tamanho (por exemplo, cerca de 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 ou 10 Mb) (por exemplo, 50 kb-10 Mb) em torno de uma seleção de loci conhecido para
15 ser envolvida em translocações, deleções, inversões, duplicações e outros ré-arranjos genômicos.

 7. Preferivelmente, a sequência 4C (bait) está a cerca de 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, 10
20 Mb, 11 Mb, 12 Mb, 13 Mb, 14 Mb ou 15 Mb ou mais da sequência ré-arranjada atual (isto é, ponto de ruptura no caso de uma translocação).

 8. Um conjunto de sondas representando o genoma completo de uma dada espécie, com cada sonda representando
25 um fragmento de restrição simples obtido ou obtenível após digestão com uma enzima de restrição primária.

 9. Um conjunto de sondas representando o genoma completo de uma dada espécie, com sondas igualmente distri-

buidas junto com os padrões do cromossomo lineares.

10. Um arranjo que compreende o conjunto de sondas de acordo com qualquer uma dos parágrafos 1-10.

11. Um método para analisar a frequência de interação de uma sequência de nucleotídeos alvo com uma ou mais sequências de nucleotídeos (por exemplo, um ou mais loci genômicos) que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

12. Um método para identificar uma ou mais interações de DNA-DNA que são indicativas de um estado da doença particular que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

13. Um método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em um DNA-DNA que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

14. Um método de arranjo para identificar um ou mais agentes que modulam uma interação de DNA-DNA que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou uma sonda aqui descrita

15. Um método para detectar a localização de um ponto de ruptura (por exemplo, uma translocação) que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

16. Um método para detectar a localização de uma inversão que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

5 17. Um método para detectar a localização de uma deleção que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

10 18. Um método para detectar a localização de uma duplicação que compreende o uso de uma sequência de nucleotídeos ou um arranjo de sondas ou um conjunto de sondas ou um arranjo aqui descrito.

15 19. O uso de microarranjos em tecnologia 4C para identificar (todos) segmentos do DNA que estão perto proximidade da espacial de um segmento de DNA de escolha.

20 20. Um microarranjo contendo as sondas homólogas às sequências de DNA diretamente adjacente aos sítios de reconhecimento da enzima de restrição primária presente na região genômica que são incluídos na análise (que pode ser o genoma completo ou parte do genoma): cada sonda localiza preferivelmente dentro de 100 bp, ou no máximo a 300 bp de um sítio de reconhecimento de enzima de restrição primária única, ou alternativamente é projetada entre cada sítio de reconhecimento da enzima de restrição primária e seu sítio de reconhecimento de enzima de restrição secundária perto.

25

21. Um arranjo aqui descrito que compreende sondas complementares às sequências de loci selecionados, em que o dito arranjo é representativo do genoma completo de uma dada

espécie.

22. Um arranjo de acordo com o parágrafo 21, em que os loci são loci associados com uma ou mais doenças.

23. Um arranjo de acordo com o parágrafo 21 ou parágrafo 22, em que as seqüências de loci selecionados incluem as seqüências que estão até 20 Mb distantes do dito loci.

24. Um método para analisar a freqüência de interação de uma seqüência de nucleotídeos alvo com uma ou mais seqüências de nucleotídeos de interesse (por exemplo, um ou mais loci genômico) que compreende as etapas de:

- (a) fornecer uma amostra de DNA reticulado;
- (b) digerir o DNA reticulado com uma enzima de restrição primária;
- (c) ligar as seqüências de nucleotídeos reticuladas;
- (d) reverter a reticulação;
- (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;
- (f) ligar as seqüências de nucleotídeos;
- (g) amplificar uma ou mais seqüências de nucleotídeos de interesse que são ligadas com a seqüência de nucleotídeos alvo usando pelo menos dois iniciadores de oligonucleotídeo, em que cada iniciador hibridiza para uma seqüência conhecida de DNA que flanqueia as seqüências de nucleotídeos de interesse;
- (h) hibridizar as seqüências amplificadas para um arranjo; e
- (i) determinar a freqüência de interação entre as

seqüências de DNA.

Aspectos adicionais 2

Aspectos ainda adicionais da presente invenção são apresentados a seguir nos parágrafos numerados.

5 1. Uma seqüência de nucleotídeos circularizada que compreende uma primeira e uma segunda seqüência de nucleotídeos separadas por um sítio de reconhecimento de enzima de restrição primária e uma secundária, em que a dita primeira seqüência de nucleotídeos é uma seqüência de nucleotídeos
10 alvo e a dita segunda seqüência de nucleotídeos é obtenível por DNA genômico de reticulação .

 2. A seqüência de nucleotídeos circularizada de acordo com o parágrafo 1, em que a seqüência de nucleotídeos alvo é selecionada do grupo que consiste em um promotor, um
15 intensificador, um silenciador, um isolante, uma região de anexação de matriz, uma região de controle de locus, uma unidade de transcrição, uma origem de replicação, uma ponto quente de recombinação, um ponto de ruptura de translocação, um centrômero, um telômero, uma região densa em genes, uma
20 região fraca em gene, um elemento repetitivo e um (viral) sítio de integração.

 3. A seqüência de nucleotídeos circularizada de acordo com o parágrafo 1, em que a seqüência de nucleotídeos alvo é uma seqüência de nucleotídeos que é associada com uma
25 doença, ou que causa a mesma, ou está localizada a menos que 15 Mb em um padrão de DNA linear de um locus que é associado com uma doença, ou que causa a mesma.

 4. A seqüência de nucleotídeos circularizada de

acordo com qualquer um dos parágrafos 1-3, em que a sequência de nucleotídeos alvo é selecionada do grupo que consiste em: *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, *IGH*, *LYL1*, *TAL1*, *TAL1*, *LMO2*, *TCR α/β* *TCRβ* e *HOX* ou outros loci associados com doença como descrito em "Catalogue of Unbalanced Chromosome Aberrations in Man." 2nd edition. Albert Schinzel. Berlin: Walter de Gruyter, 2001. ISBN 3-11-011607-3.

5 5. A sequência de nucleotídeos circularizada de acordo com qualquer um dos parágrafos 1-4, em que o sítio de reconhecimento da enzima de restrição primária é um sítio de
10 reconhecimento de 6-8 bp, preferivelmente selecionado do grupo que consiste em *BgIII*, *HindIII*, *EcoRI*, *BamHI*, *SpeI*, *PstI* e *NdeI*.

15 6. A sequência de nucleotídeos circularizada de acordo com qualquer um dos parágrafos anteriores, em que o sítio de reconhecimento da enzima de restrição secundária é um sítio de reconhecimento sequência de nucleotídeos de 4 ou 5 bp.

20 7. A sequência de nucleotídeos circularizada de acordo com qualquer um dos parágrafos anteriores, em que o sítio de reconhecimento da enzima de restrição secundária está localizado a mais que cerca de 350 bp do sítio de restrição primária.

25 8. A sequência de nucleotídeos circularizada de acordo com qualquer um dos parágrafos anteriores, em que a sequência de nucleotídeos é marcada.

9. Uma sequência de nucleotídeos que compreende uma primeira e uma segunda sequência de nucleotídeos separa-

das por um sítio de reconhecimento de enzima de restrição primária e uma secundária, onde a dita primeira sequência de nucleotídeos é uma sequência de nucleotídeos alvo, a segunda sequência de nucleotídeos é obtível por DNA genômico de
5 reticulação em que a dita segunda sequência de nucleotídeos cruza a sequência de nucleotídeos alvo.

10. Um método para preparar uma sequência de nucleotídeos circularizada que compreende as etapas de

- (a) fornecer uma amostra de DNA reticulado;
- 10 (b) digerir o DNA reticulado com uma enzima de restrição primária;
- (c) ligar as sequências de nucleotídeos reticuladas;
- (d) reverter a reticulação
- 15 (e) digerir as sequências de nucleotídeos com uma enzima de restrição secundária; e
- (f) circular as sequências de nucleotídeos.

11. Um método para preparar uma sequência de nucleotídeos que compreende as etapas de:

- 20 (a) fornecer uma amostra de DNA reticulado;
- (b) digerir o DNA reticulado com uma enzima de restrição primária;
- (c) ligar as sequências de nucleotídeos reticuladas;
- 25 (d) reverter a reticulação;
- (e) digerir as sequências de nucleotídeos com uma enzima de restrição secundária;
- (f) circular as sequências de nucleotídeos; e

(g) amplificar uma ou mais seqüências de nucleotídeos ligadas à seqüência de nucleotídeos alvo.

12. Um método de acordo com o parágrafo 11, em que a seqüência de nucleotídeos circularizada visada é linearizada antes da amplificação.

13. Um método de acordo com o parágrafo 12, em que a seqüência de nucleotídeos visada circularizada é linearizada usando uma enzima de restrição que reconhece um sítio de reconhecimento de 6 bp ou mais.

14. Um método de acordo com qualquer um dos parágrafos 10-13, em que a seqüência reticulada de nucleotídeos é amplificada usando PCR

15. Um método de acordo com o parágrafo 14, em que a seqüência reticulada de nucleotídeos é amplificada usando PCR inversa

16. Um método de acordo com o parágrafo 14 ou parágrafo 15, em que é usado o Sistema PCR Padrão Expand Long (Roche).

17. Um método para analisar a freqüência de interação de uma seqüência de nucleotídeos alvo com uma ou mais seqüências de nucleotídeos (por exemplo, um ou mais loci genômico) que compreende o uso de uma seqüência de nucleotídeos de acordo com qualquer um dos parágrafos 1-9.

18. Um arranjo de sondas imobilizadas em um suporte que compreende uma ou mais sondas que hibridizam ou são capazes de hibridizar para uma seqüência de nucleotídeos de acordo com qualquer um dos parágrafos 1-9.

19. Um conjunto de sondas complementares em se-

quência com a sequência de ácido nucléico adjacente a cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima primária restrição no DNA genômico.

20. Um conjunto de sondas de acordo com o parágrafo 5 19, em que as sondas são complementares em sequência com a sequência de ácido nucléico adjacente a cada lado de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

21. Um conjunto de sondas de acordo com o parágrafo 10 19 ou parágrafo 20, em que as ditas sondas são complementares em sequência com a sequência de ácido nucléico que está menos que 300 pares base de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

15 22. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-21, em que as sondas são complementares à sequência que está a menos de 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária em DNA genômico.

20 23. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-22, em que as sondas são complementares à sequência que está entre 200 e 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

25 24. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-23, em que as sondas são complementares à sequência que está entre 100 e 200 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de

uma enzima de restrição primária no DNA genômico.

25. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-24, em que duas ou mais das sondas são projetadas que são capazes de hibridizar com a sequência ad-
5 jacente a cada sítio de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

26. Um conjunto de sondas de acordo com o parágrafo 25, em que as sondas sobrepõem-se, ou sobrepõem-se parcial-
10 almente.

27. Um conjunto de sondas do parágrafo 26, em que a sobreposição é menos que 10 nucleotídeos.

28. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-27, em que a sequência de sondas cor-
15 responde a toda ou parte da sequência entre cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária e cada um dos primeiros sítios de reconhecimento de enzima de restrição secundária vizinhos de uma enzima de restrição secundária.

20 29. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-28, em que cada sonda é pelo menos 25 mer.

30. Um conjunto de sondas de acordo com qualquer um dos parágrafos 19-29, em que cada uma das sondas é 25-60
25 mer.

31. Um processo para preparar um conjunto de sondas que compreende as etapas de:

(a) identificar cada um dos sítios de reconheci-

mento da enzima de restrição primária para uma enzima de restrição primária em DNA genômico;

(b) projetar sondas que são capazes de hibridizar com a sequência adjacente a cada um dos sítios de reconhecimento da enzima de restrição primária no DNA genômico;

(c) sintetizar as sondas; e

(d) combinar as sondas entre si para formar um conjunto de sondas ou substancialmente um conjunto de sondas.

10 32. Um processo de acordo com o parágrafo 31, em que as sondas são produtos de amplificação PCR.

33. Um conjunto de sondas, ou substancialmente um conjunto de sondas obtido ou obtenível pelo processo de acordo com o parágrafo 31 ou parágrafo 32.

15 34. Um arranjo compreendendo um arranjo de sondas de acordo com o parágrafo 18 ou substancialmente o conjunto de sondas de acordo com qualquer um dos parágrafos 19-30 ou 33.

20 35. Um arranjo que compreende o conjunto de sondas de acordo com qualquer um dos parágrafos 19-30 ou 33.

36. Um arranjo de acordo com o parágrafo 34 ou parágrafo 35, em que o arranjo compreende cerca de 300.000-400.000 sondas.

25 37. Um arranjo de acordo com qualquer um dos parágrafos 34-36, em que o arranjo compreende cerca de 385.000 ou mais sondas, preferivelmente, cerca de 750.000 sondas, mais preferivelmente, 6 x 750.000 sondas.

38. Um arranjo de acordo com qualquer um dos pará-

grafos 34-37, em que, se o número de sondas exceder o número de sondas que pode ser contido em um arranjo simples, então o arranjo compreenderá ou consistirá em uma representação do genoma completo de uma dada espécie em baixa resolução.

5 39. Um arranjo de acordo com o parágrafo 38, em que uma a cada 2, 3, 4, 5, 6, 7, 8, 9 ou 10 sondas ordenadas em um padrão de cromossomo linear é contida no arranjo.

 40. Um processo para preparar um arranjo que compreende a etapa de imobilizar em um suporte substancialmente
10 sólido o arranjo de sondas de acordo com o parágrafo 18 ou substancialmente o conjunto de sondas de acordo com qualquer um dos parágrafos 19-30 ou 33.

 41. Um processo para preparar um arranjo que compreende a etapa de imobilizar em um suporte sólido o arranjo
15 de sondas de acordo com o parágrafo 18 ou o conjunto de sondas de acordo com qualquer um dos parágrafos 19-30 ou 33.

 42. Um arranjo obtido ou obtenível pelo método de acordo com o parágrafo 40 ou parágrafo 41.

 43. Um método para analisar a frequência de interação de uma sequência de nucleotídeos visada com uma ou
20 mais sequências de nucleotídeos (por exemplo, um ou mais loci genômico) que compreende as etapas de:

- (a) fornecer uma amostra de DNA reticulado;
- (b) digerir o DNA reticulado com uma enzima de
25 restrição primária;
- (c) ligar as sequências de nucleotídeos reticuladas;
- (d) reverter a reticulação;

(e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

(f) circular as seqüências de nucleotídeos;

(g) amplificar uma ou mais seqüências de nucleotí-
5 deos que são ligadas à seqüência de nucleotídeos alvo;

(i) hibridizar opcionalmente as seqüências ampli-
ficadas para um arranjo; e

(h) determinar a freqüência de interação entre as
seqüências de DNA.

10 44. Um método para identificar uma ou mais intera-
ções de DNA-DNA que são indicativas de um estado da doença
particular que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado de uma
célula doente e um não doente;

- 15 (b) digerir o DNA reticulado em cada um das amos-
tras com uma enzima de restrição primária;

(c) ligar as seqüências de nucleotídeos reticula-
das;

(d) reverter a reticulação;

20 (e) digerir as seqüências de nucleotídeos com uma
enzima de restrição secundária;

(f) circular as seqüências de nucleotídeos;

(g) amplificar uma ou mais seqüências que são li-
gadas à seqüência de nucleotídeos alvo;

25 (h) hibridizar opcionalmente as seqüências de nu-
cleotídeos amplificadas para um arranjo; e

(i) determinar a freqüência de interação entre as
seqüências de DNA,

em que uma diferença entre o frequência de interação entre as seqüências de DNA das células doentes e não doentes indica que a interação de DNA-DNA é indicativo de um estado de doença particular.

5 45. Um método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado de um sujeito;

10 (b) digerir o DNA reticulado com uma enzima de restrição primária;

(c) ligar o seqüências de nucleotídeos reticuladas;

(d) reverter a reticulação;

15 (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

(f) circular as seqüências de nucleotídeos;

(g) amplificar uma ou mais seqüências que são ligadas à seqüência de nucleotídeos alvo;

20 (h) hibridizar opcionalmente as seqüências amplificadas de nucleotídeos para um arranjo;

(i) determinar a frequência de interação entre as seqüências de DNA; e

25 (j) comparar a frequência de interação entre as seqüências de DNA com aquela de um controle não afetado;

em que uma diferença entre o valor obtido do controle e o valor obtido do sujeito é indicativo que o sujeito sofre da doença ou síndrome ou é indicativo que o sujeito

sofrerá da doença ou síndrome.

46. Um método de acordo com o parágrafo 45, em que uma transição das frequências da interação baixa para alta é indicativa da localização de um ponto de ruptura.

5 47. Um método de acordo com o parágrafo 45 em que um padrão inverso de frequências de interação de DNA-DNA para a amostra do sujeito comparado com o controle é indicativo de uma inversão.

10 48. Um método de acordo com o parágrafo 45 em que uma redução na interação de frequência de DNA-DNA com a amostra do sujeito comparado com o controle, em combinação com um aumento na interação de frequência de DNA-DNA para regiões mais distantes, é indicativa de deleção.

15 49. Um método de acordo com o parágrafo 45, em que um aumento ou uma diminuição na interação de frequência de DNA-DNA com a amostra do sujeito comparado com o controle é indicativo de uma duplicação ou inserção.

20 50. Um método de acordo com qualquer um dos parágrafos 45-49, em que cariotipagem espectral e/ou FISH é usado antes de realizar o dito método.

51. Um método de acordo com qualquer um dos parágrafos 45-50, em que a doença é uma doença genética.

52. Um método de acordo com qualquer um dos parágrafos 45-51, em que a doença é câncer.

25 53. Um método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado de um

sujeito;

(b) digerir o DNA reticulado com uma enzima de restrição primária;

(c) ligar as seqüências de nucleotídeos reticuladas;

(d) reverter a reticulação;

(e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

(f) circular as seqüências de nucleotídeos;

(g) amplificar duas ou mais seqüências que são ligadas à seqüência de nucleotídeos alvo(s);

(h) marcação o duas ou mais seqüências amplificadas;

(i) hibridizar as seqüências de nucleotídeos em um arranjo;

(j) determinar a freqüência de interação entre as seqüências de DNA; e

(j) identificar um ou mais loci que foram submetidos a um ré-arranjo genômico que é associado com uma doença.

54. Um método de acordo com o parágrafo 53, em que as duas ou mais seqüências amplificadas são diferencialmente marcadas.

55. Um método de acordo com o parágrafo 54, em que as duas ou mais seqüências amplificadas são identicamente marcadas quando as seqüências residem em diferentes cromossomos.

56. Um método de acordo com o parágrafo 53, em que as duas ou mais seqüências amplificadas são identicamente

marcadas quando as seqüências residem no mesmo cromossomo em uma distância que é grande o bastante para sobreposição mínima entre interação de sinais de DNA-DNA.

57. Um método de arranjo para identificar um ou
5 mais agentes que modulam uma interação de DNA-DNA que compreende as etapas de:

(a) colocar em contato uma amostra com um ou mais agentes;

(b) fornecer DNA reticulado da amostra;

10 (c) digerir o DNA reticulado com uma enzima de restrição primária;

(d) ligar as seqüências de nucleotídeos reticuladas;

(e) reverter a reticulação;

- 15 (f) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

(g) circular as seqüências de nucleotídeos;

(h) amplificar uma ou mais seqüências de nucleotídeos que são ligadas à seqüência de nucleotídeos alvo;

20 (i) hibridizar opcionalmente as seqüências amplificadas de nucleotídeos para um arranjo; e

(j) determinar a freqüência de interação entre as seqüências de DNA,

em que uma diferença entre (i) a freqüência de interação entre as seqüências de DNA na presença do agente e
25 (ii) a freqüência de interação entre as seqüências de DNA na ausência do agente é indicativo de um agente que modula a interação de DNA-DNA.

58. Um método para detectar a localização de um ponto de ruptura (por exemplo, uma translocação) que compreende as etapas de:

- (a) fornecer uma amostra de DNA reticulado; (b) digerir o DNA reticulado com uma enzima de restrição primária;
 - (c) ligar as seqüências de nucleotídeos reticuladas;
 - (d) reverter a reticulação;
 - (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;
 - (f) circular as seqüências de nucleotídeos;
 - (g) amplificar uma ou mais seqüências que são ligadas à seqüência de nucleotídeos alvo;
 - (h) hibridizar opcionalmente as seqüências amplificadas de nucleotídeos para um arranjo;
 - (i) determinar a freqüência de interação entre as seqüências de DNA; e
 - (j) comparar a freqüência de interação entre as seqüências de DNA com aquela de um controle;
- em que uma transição de baixa a alta interação de freqüência de DNA-DNA na amostra comparado com o controle é indicativo da localização de um ponto de ruptura.

59. Um método para detectar a localização de uma inversão que compreende as etapas de:

- (a) fornecer uma amostra de DNA reticulado;
- (b) digerir o DNA reticulado com uma enzima de restrição primária;
- (c) ligar as seqüências de nucleotídeos reticula-

das;

(d) reverter a reticulação;

(e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

5 (f) circular as seqüências nucleotídeos;

(g) amplificar uma ou mais seqüências que são ligadas à seqüência de nucleotídeos alvo;

(h) hibridizar opcionalmente as seqüências amplificadas de nucleotídeos para um arranjo;

10 (i) determinar a freqüência de interação entre as seqüências de DNA; e

(j) comparar a freqüência de interação entre as seqüências de DNA com a de um controle;

em que um padrão inverso de freqüências da interação de DNA-DNA com a amostra comparado com o controle é indicativo de um inversão.

60. Um método para detectar a localização de uma deleção que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado;

20 (b) digerir o DNA reticulado com uma enzima de restrição primária;

(c) ligar as seqüências de nucleotídeos reticulada;

(d) reverter a reticulação;

25 (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

(f) circular as seqüências de nucleotídeos;

(g) amplificar uma ou mais seqüências que são li-

gadas à sequência de nucleotídeos alvo;

(h) hibridizar opcionalmente as sequências amplificadas de nucleotídeos com um arranjo;

(i) determinar a frequência de interação entre as
5 sequências de DNA; e

(j) comparar a frequência de interação entre as sequências de DNA com a de um controle;

em que uma redução na interação de frequência de DNA-DNA com a amostra comparado com o controle é indicativo
10 de deleção.

61. Um método para detectar a localização de uma duplicação que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado;

(b) digerir o DNA reticulado com uma enzima de
15 restrição primária;

(c) ligar as sequências de nucleotídeos reticuladas;

(d) reverter a reticulação,

(e) digerir as sequências de nucleotídeos com uma
20 enzima de restrição secundária;

(F) circular as sequências de nucleotídeos;

(g) amplificar uma ou mais sequências que são ligadas à sequência de nucleotídeos alvo;

(h) hibridizar opcionalmente as sequências amplificadas de nucleotídeos para um arranjo;
25

(i) determinar a frequência de interação entre as sequências de DNA; e

(j) comparar a frequência de interação entre as

seqüências de DNA com a de um controle;

em que um aumento ou uma diminuição em frequência de interação de DNA-DNA com a amostra do sujeito comparado com o controle é indicativo de uma duplicação ou inserção.

5 62. Um agente obtido ou obtenível pelo método de arranjo de acordo com o parágrafo 57.

63. Uso de uma seqüência de nucleotídeos de acordo com qualquer um dos parágrafos 1-9 para identificar uma ou mais interações de DNA-DNA em uma amostra.

10 64. Uso de uma seqüência de nucleotídeos de acordo com qualquer um dos parágrafos 1-9 para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

15 65. Uso de um arranjo de sondas de acordo com o parágrafo 18 ou o conjunto de sondas de acordo com qualquer um dos parágrafos 19-30 ou 33 para identificar uma ou mais interações de DNA-DNA em uma amostra.

20 66. Uso de um arranjo de sondas de acordo com o parágrafo 18 ou o conjunto de sondas de acordo com qualquer um dos parágrafos 19-30 ou 33 para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

25 67. Uso de um arranjo, de acordo com qualquer um dos parágrafos 34-39 ou 42 para identificar uma ou mais interações de DNA-DNA em uma amostra.

68. Uso de um arranjo de acordo com qualquer um dos parágrafos 34-39 ou 42 para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mu-

dança em uma interação de DNA-DNA.

69. Uso de acordo com qualquer um dos parágrafos 64, 66 ou 68, em que o diagnóstico ou prognóstico é diagnóstico ou prognóstico pré-natal.

5 70. Um método substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

71. Um arranjo de sondas substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

10 72. Um conjunto de sondas substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

73. Um processo substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

15 74. Um arranjo substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

75. Um método de arranjo substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

20 76. Um agente substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

77. Um uso substancialmente aqui descrito e com referência a qualquer dos Exemplos ou Figuras.

Tabela 2

Interação	Em 4C	N	% de sobreposição	em Crio- FISH	Valor p
β -globina -Chr.7 73,1 Mb	+	258	7,4	+	P<0,001

β -globina Chr.7 80,1 Mb (OR)	-	254	3,6	-	
β -globina Chr.7 118,3 Mb	-	255	3,5	-	
β -globina Chr.7 127,9 Mb (Uros)	+	259	6,6	+	P<0,001
β -globina Chr.7 130,1 Mb	+	413	9,7	+	P<0,001
β -globina Chr.7 135,0 Mb (OR)	-	261	1,9	-	
β -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	+	261	8	+	P<0,001

Interação	Em 4C	N	% de sobreposição	em Crio- FISH	Valor p
β -globina -Chr.7 73,1 Mb	-	256	3,9	-	
β -globina Chr.7 80,1 Mb (OR)	+	256	12,9	+	P<0,001
β -globina Chr.7 118,3 Mb	-	242	4,1	-	
β -globina Chr.7 127,9 Mb (Uros)	-	263	3	-	
β -globina Chr.7 130,1 Mb	+	256	7	+	P<0,05
β -globina Chr.7 135,0 Mb (OR)		258	6,2	+	P<0,05

Chr.7 80,1 Mb - Chr.7 135,0 Mb		261	5	+	P<0,1
Rad23A - Chr.8 21,8 MB	-	260	3,8	-	
Rad23A - Chr.8 122,3	+	258	8,1	+	P<0,001

Referências

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

5 Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. 2002. Capturing chromosomal conformation. *Science* 295:1306-11.

Drissen R, Palstra. RJ, Gillemans N, Splinter E, Grosveld F, Philipson 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* 15 37:31-40.

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

20 Palstra, R.J., Tolhuis, B, Splinter, E., Nijmeijer, R, Grosveld, F., and 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., Irnanra, A.M.A, Strouboulzs, J., de Laat, W.,
 and Grosveld, F.G. (2004). Multiple interactions between
 regulatory regions are required to stabilize a active chro-
 5 matin hub. *Genes & Dev.* 18: 1495-1509.

Spilianakis CG, Flavell RA. 2004. Long-range in-
 trachromosomal interactions no T helper type 2 cytokine lo-
 cus. *Nat Immunol* 5:1017-27.

Tolhui.s, B., Palstra, R.J., Splinter; E.,
 10 &osveld, F., and de Laat, W. 2002. Looping and interaction
 between hypersensitive sites in the active beta-globin lo-
 cus. *Molecular Cell* 10: 1453-65. Vakoc CR, Letting DL, Ghel-
 dof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker 3,
 Blobel GA. 2005. Proximity among distant regulatory elements
 15 at the betaglobin locus requires GATA 1 and FOG-1. *Mol Cell.*
 17:453-62
 Todas publicações nas especificações supramenciona-
 das estão aqui incorporadas pela referência. Várias modifi-
 cações e variações dos métodos descritos e sistema da inven-
 ção ficarão aparentes aos versados na tecnologia sem fugir
 20 do escopo e espírito da invenção. Embora a invenção seja
 descrita com relação às modalidades específicas preferidas,
 deve-se entender que a invenção reivindicada não se limitará
 indevidamente a tais modalidades específicas. Efetivamente,
 várias modificações das maneiras descritas para realizar a
 25 invenção que são óbvias aos versados em biologia molecular
 ou campos relacionados estarão no escopo das reivindicações
 seguintes.

REIVINDICAÇÕES

1. Método para analisar a frequência de interação de uma sequência de nucleotídeos alvo com uma ou mais sequências de nucleotídeos de interesse (por exemplo, um ou
5 mais loci genômicos), **CARACTERIZADO** pelo fato de que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado;

(b) digerir o DNA reticulado com uma enzima de restrição primária;

10 (c) ligar as sequências de nucleotídeos reticuladas;

(d) reverter a reticulação;

(e) digerir as sequências de nucleotídeos com uma enzima de restrição secundária;

15 (f) ligar uma ou mais sequências de DNA de composição de nucleotídeo conhecida ao(s) sítio(s) de digestão de enzima de restrição secundária disponível(s) que flanqueia(m) uma ou mais sequências de nucleotídeos de interesse;

20 (g) amplificar uma ou mais sequências de nucleotídeos de interesse usando pelo menos dois iniciadores de oligonucleotídeo, em que cada iniciador hibridiza para as sequências de DNA que flanqueiam as sequências de nucleotídeos de interesse;

25 (h) hibridizar a(s) sequência(s) amplificada(s) em um arranjo; e

(i) determinar a frequência de interação entre as sequências de DNA.

2. Método, de acordo com a reivindicação 1, **CARACTERIZADO** pelo fato de que a reação da ligação na etapa (f) resulta na formação de círculos de DNA.

3. Método, de acordo com a reivindicação 1 ou 2, **CARACTERIZADO** pelo fato de que a sequência de nucleotídeos alvo é selecionada do grupo que consiste em um rearranjo genômico, promotor, um intensificador, um silenciador, um isolante, uma região de anexação de matriz, uma região de controle de locus, uma unidade de transcrição, uma origem de replicação, um ponto quente de recombinação, um ponto de ruptura de translocação, um centrômero, um telômero, uma região densa em genes, uma região pobre em genes, um elemento repetitivo e um sítio de integração (viral).

4. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que a sequência de nucleotídeos alvo é uma sequência de nucleotídeos que está associada com uma doença, ou causa a mesma, ou é localizada até 15 Mb ou mais em um padrão de DNA linear de um locus que é associado com uma doença, ou causa a mesma.

5. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que a sequência de nucleotídeos alvo é selecionada do grupo que consiste em: *AML1*, *MLL*, *MYC*, *BCL*, *BCR*, *ABL1*, *IGH*, *LYL1*, *TAL1*, *TAL2*, *LMO2*, *TCR α / δ* , *TCR β* e *HOX* ou outros loci associados com doença descritos em "Catalogue of Unbalanced Chromosome Aberrations in Man" 2ª edição. Albert Schiuzel. Berlin: Walter of Gzrryter, 2001. ISBN 3-11-011607-3.

6. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que a enzima de restrição primária é uma enzima de restrição que reconhece um sítio de reconhecimento de 6-8 bp.

5 7. Método, de acordo com a reivindicação 6, **CARACTERIZADO** pelo fato de que a enzima de restrição primária é selecionada do grupo que consiste em BglII, HindIII, EcoRI, BamHI, SpeI, PstI e NdeI.

10 8. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que a enzima de restrição secundária é uma enzima de restrição que reconhece um sítio de reconhecimento de seqüência de nucleotídeos de 4 ou 5 bp.

15 9. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que o sítio de reconhecimento da enzima de restrição secundária está localizado em mais que cerca de 350 bp do sítio de restrição primária na seqüência de nucleotídeos alvo.

20 10. Método, de acordo com qualquer uma das reivindicações anteriores, **CARACTERIZADO** pelo fato de que a seqüência de nucleotídeos é marcada.

25 11. Método para analisar a freqüência de interação de uma seqüência de nucleotídeos alvo com uma ou mais seqüências de nucleotídeos (por exemplo, um ou mais loci genômicos), **CARACTERIZADO** pelo fato de que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado;

(b) digerir o DNA reticulado com uma enzima de

restrição primária;

(c) ligar as seqüências reticuladas de nucleotídeo;

(d) reverter a reticulação;

5 (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária;

(f) circular as seqüências de nucleotídeo;

(g) amplificar a uma ou mais seqüências de nucleotídeos que são ligadas à seqüência de nucleotídeos
10 alvo;

(h) hibridizar opcionalmente as seqüências amplificadas em um arranjo; e

(i) determinar a freqüência de interação entre as seqüências de DNA.

15 12. Seqüência de nucleotídeos circularizada, **CARACTERIZADA** pelo fato de que compreende uma primeira e uma segunda seqüência de nucleotídeo, em que cada extremidade da primeira e da segunda seqüência de nucleotídeos são separadas por diferentes sítios de reconhecimento de enzima
20 de restrição, e em que a dita primeira seqüência de nucleotídeos é uma seqüência de nucleotídeos alvo e a dita segunda seqüência de nucleotídeos é obtenível por DNA genômico reticulado.

25 13. Método para preparar uma seqüência de nucleotídeos circularizada, **CARACTERIZADO** pelo fato de que compreende as etapas de:

(a) fornecer uma amostra de DNA reticulado;

(b) digerir o DNA reticulado com uma enzima

restrição primária;

(c) ligar as seqüências de nucleotídeos reticuladas;

(d) reverter a reticulação;

5 (e) digerir as seqüências de nucleotídeos com uma enzima de restrição secundária; e

(f) circular as seqüências de nucleotídeo.

14. Método para analisar a freqüência de interação de uma seqüência de nucleotídeos alvo com uma ou mais
10 seqüências de nucleotídeos (por exemplo, um ou mais loci genômicos), **CARACTERIZADO** pelo fato de que compreende o uso de uma seqüência de nucleotídeos conforme definida na reivindicação 12.

15 15. Arranjo de sondas imobilizadas em um suporte, **CARACTERIZADO** pelo fato de que compreende uma ou mais sondas que hibridizam ou são capazes de hibridizar para uma seqüência de nucleotídeos conforme definida na reivindicação 12.

20 16. Conjunto de sondas, **CARACTERIZADO** pelo fato de que é complementar em seqüência à seqüência de ácido nucléico adjacente a cada um dos sítios de reconhecimento das enzimas de restrição primária de uma enzima de restrição primária no DNA genômico.

25 17. Conjunto de sondas, de acordo com a reivindicação 16, **CARACTERIZADO** pelo fato de que as sondas são complementares em seqüência à seqüência de ácido nucléico adjacente a cada lado de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima

de restrição primária no DNA genômico.

18. Conjunto de sondas, de acordo com a reivindicação 16 ou 17, **CARACTERIZADO** pelo fato de que as ditas sondas são complementares na sequência à sequência de
5 ácido nucléico que está menos que 300 pares de base de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

19. Conjunto de sondas, de acordo com qualquer uma
10 das reivindicações 16-18, **CARACTERIZADO** pelo fato de que as sondas são complementares à sequência que está menos que 300 bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

15 20. Conjunto de sondas, de acordo com qualquer uma das reivindicações 16-19, **CARACTERIZADO** pelo fato de que as sondas são complementares à sequência que está entre 200 e 300-bp de cada um dos sítios de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no
20 DNA genômico.

21. Conjunto de sondas, de acordo com qualquer uma das reivindicações 16-20, **CARACTERIZADO** pelo fato de que as sondas são complementares à sequência que está entre 100 e 200 bp ou 0 a 100 bp de cada um dos sítios de reconhecimento
25 da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

22. Conjunto de sondas, de acordo com qualquer uma das reivindicações 16-21, **CARACTERIZADO** pelo fato de que

duas ou mais sondas são capazes de hibridizar para a sequência adjacente a cada sítio de reconhecimento da enzima de restrição primária de uma enzima de restrição primária no DNA genômico.

5 23. Conjunto de sondas, de acordo com a reivindicação 22, **CARACTERIZADO** pelo fato de que as sondas sobrepõem-se, ou sobrepõem-se parcialmente.

 24. Conjunto de sondas, de acordo com a reivindicação 23, **CARACTERIZADO** pelo fato de que a
10 sobreposição é de menos do que 10 nucleotídeos.

 25. Conjunto de sondas, de acordo com qualquer uma das reivindicações 16-24, **CARACTERIZADO** pelo fato de que a sequência de sondas corresponde a toda ou parte da sequência entre cada um dos sítios de reconhecimento da enzima de
15 restrição primária de uma enzima de restrição primária e cada um dos primeiros sítios de reconhecimento de enzima de restrição secundária vizinhos de uma enzima de restrição secundária.

 26. Conjunto de sondas, de acordo com qualquer uma
20 das reivindicações 16-25, **CARACTERIZADO** pelo fato de que cada sonda está a pelo menos 25 mer.

 27. Conjunto de sondas, de acordo com qualquer uma das reivindicações 16-26, **CARACTERIZADO** pelo fato de que cada uma das sondas está a 25-60 mer.

25 28. Processo para preparar um conjunto de sondas, **CARACTERIZADO** pelo fato de que compreende as etapas de:

 (a) identificar cada um dos sítios de reconhecimento da enzima de restrição primária para uma

enzima de restrição primária no DNA genômico;

(b) projetar sondas que são capazes de hibridizar para a sequência adjacente a cada um dos sítios de reconhecimento da enzima de restrição primária no DNA
5 genômico;

(c) sintetizar as sondas; e

(d) combinar as sondas entre si para formar um conjunto de sondas, ou substancialmente um conjunto de sondas.

10 29. Processo de acordo com a reivindicação 28, **CARACTERIZADO** pelo fato de que as sondas são produtos de amplificação PCR.

30. Conjunto de sondas, ou substancialmente um conjunto de sondas, **CARACTERIZADO** pelo fato de que é obtido
15 ou obtenível pelo processo conforme definido na reivindicação 28 ou 20.

31. Arranjo, **CARACTERIZADO** pelo fato de que compreende o arranjo de sondas de acordo com a reivindicação
15, ou substancialmente o conjunto de sondas conforme
20 definido em qualquer uma das reivindicações 16-27 ou 30.

32. Arranjo, **CARACTERIZADO** pelo fato de que compreende o conjunto de sondas conforme definido em qualquer uma das reivindicações 16-27 ou 30.

33. Arranjo, de acordo com a reivindicação 32,
25 **CARACTERIZADO** pelo fato de que o arranjo compreende cerca de 300.000-400.000 sondas.

34. Arranjo, de acordo com qualquer uma das reivindicações 32-33, **CARACTERIZADO** pelo fato de que o

arranjo compreende cerca de 385.000 ou mais sondas, preferivelmente cerca de 750.000 sondas, mais preferivelmente 6 x 750.000 sondas.

35. Arranjo, de acordo com qualquer uma das
5 reivindicações 31-34, **CARACTERIZADO** pelo fato de que o arranjo compreende ou consiste em uma representação do genoma completo de uma dada espécie em baixa resolução.

36. Arranjo, de acordo com a reivindicação 35, **CARACTERIZADO** pelo fato de que uma de cada 2, 3, 4, 5, 6, 7,
10 8, 9 ou 10 sondas ordenadas em um padrão de cromossomo linear é contida no arranjo.

37. Processo para preparar um arranjo, **CARACTERIZADO** pelo fato de que compreende a etapa de imobilizar em um suporte sólido substancialmente o arranjo
15 de sondas conforme definido na reivindicação 15, ou substancialmente o conjunto de sondas conforme definido em qualquer uma das reivindicações 16-27 ou 30.

38. Processo para preparar um arranjo, **CARACTERIZADO** pelo fato de que compreende a etapa de
20 imobilizar em um suporte sólido o arranjo de sondas conforme definido na reivindicação 16, ou substancialmente o conjunto de sondas conforme definido em qualquer uma das reivindicações 16-27 ou 30.

39. Arranjo, **CARACTERIZADO** pelo fato de que é
25 obtido ou obtenível pelo método conforme definido na reivindicação 37 ou 38.

40. Método para identificar uma ou mais interações de DNA-DNA que são indicativas de um estado da doença

particular, **CARACTERIZADO** pelo fato de que compreende a etapa de realizar as etapas (a)-(i) conforme definidas nas reivindicações 1-11, em que na etapa (a) uma amostra de DNA reticulado é fornecida de uma célula doente e uma não doente, e em que uma diferença entre a frequência de interação entre as seqüências de DNA das células doente e não doente indica que a interação de DNA-DNA é indicativa de um estado de doença particular.

41. Método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA, **CARACTERIZADO** pelo fato de que compreende a etapa de realizar as etapas (a)-(i) conforme definidas nas reivindicações 1-11, em que a etapa (a) compreende fornecer uma amostra de DNA reticulado de um sujeito; e em que a etapa (i) compreende comparar a frequência de interação entre as seqüências de DNA com a de um controle não afetado; em que uma diferença entre o valor obtido do controle e o valor obtido do sujeito é indicativa que o sujeito sofre da doença ou síndrome ou é indicativa que o sujeito sofrerá da doença ou síndrome.

42. Método, de acordo com a reivindicação 41, **CARACTERIZADO** pelo fato de que uma transição de frequência de interação baixa para alta é indicativa da localização de um ponto de ruptura equilibrado e/ou desequilibrado.

43. Método, de acordo com a reivindicação 41, **CARACTERIZADO** pelo fato de que um padrão inverso de frequências da interação de DNA-DNA para a amostra do sujeito comparado com o controle é indicativa de um inversão

equilibrada e/ou desequilibrada.

44. Método, de acordo com a reivindicação 41, **CARACTERIZADO** pelo fato de que uma redução na frequência de interação de DNA-DNA para a amostra do sujeito comparada com o controle, em combinação com um aumento na frequência de interação de DNA-DNA para regiões mais distantes, é indicativa de uma deleção equilibrada e/ou desequilibrada.

45. Método, de acordo com a reivindicação 41, **CARACTERIZADO** pelo fato de que um aumento ou uma diminuição na frequência interação de DNA-DNA para a amostra do sujeito comparado com o controle é indicativo de uma duplicação ou inserção equilibrada e/ou desequilibrada.

46. Método, de acordo com qualquer uma das reivindicações 41-45, **CARACTERIZADO** pelo fato de que cariotipagem espectral e/ou FISH é usado antes de realizar o dito método.

47. Método, de acordo com qualquer uma das reivindicações 41-46, **CARACTERIZADO** pelo fato de que a doença é uma doença genética.

48. Método, de acordo com qualquer uma das reivindicações 41-47, **CARACTERIZADO** pelo fato de que a doença é câncer.

49. Método de diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA, **CARACTERIZADO** pelo fato de que compreende a etapa de: realizar as etapas (a)-(i) conforme definidas nas reivindicações 1-11, em que a etapa (a) compreende fornecer uma amostra de DNA reticulado de um

sujeito; e em que o dito método compreende a etapa adicional de: (j) identificar um ou mais loci que foram submetidos a um rearranjo genômico que é associado com uma doença.

50. Método, de acordo com a reivindicação 49,
5 **CARACTERIZADO** pelo fato de que as duas ou mais seqüências amplificadas são diferencialmente marcadas.

51. Método, de acordo com a reivindicação 49,
CARACTERIZADO pelo fato de que as duas ou mais seqüências
amplificadas são identicamente marcadas quando as seqüências
10 residem em diferentes cromossomos.

52. Método, de acordo com a reivindicação 49,
CARACTERIZADO pelo fato de que as duas ou mais seqüências
amplificadas são identicamente marcadas quando as seqüências
residem no mesmo cromossomo a uma distância que é grande o
15 bastante para mínima sobreposição entre sinais de interação
de DNA-DNA.

53. Método de arranjo para identificar um ou mais
agentes que modulam uma interação de DNA-DNA, **CARACTERIZADO**
pelo fato de que compreende as etapas de:

20 (a) colocar em contato uma amostra com um ou mais
agentes; e

(b) realizar as etapas (a) a (i) conforme
definidas nas reivindicações 1-11, em que a etapa (a)
compreende fornecer DNA reticulado da amostra;

25 em que uma diferença entre (i) a freqüência de
interação entre as seqüências de DNA na presença do agente e
(ii) a freqüência de interação entre as seqüências de DNA na
ausência do agente é indicativa de um agente que modula a

interação de DNA-DNA.

54. Método para detectar a localização de um ponto de ruptura equilibrado e/ou desequilibrado (por exemplo, uma translocação), **CARACTERIZADO** pelo fato de compreender as etapas de:

(a) realizar as etapas (a) a (i) conforme definidas nas reivindicações 1-11; e

(b) comparar a frequência de interação entre as seqüências de DNA com a de um controle;

em que uma transição da frequência de interação de DNA-DNA baixa para alta na amostra comparada com o controle é indicativa da localização de um ponto de ruptura.

55. Método para detectar a localização de uma inversão equilibrada e/ou desequilibrada, **CARACTERIZADO** pelo fato de que compreende as etapas de:

(a) realizar as etapas (a) a (i) conforme definidas nas reivindicações 1-11; e

(b) comparar a frequência de interação entre as seqüências de DNA com a de um controle;

em que um padrão inverso de frequências de interação de DNA-DNA para a amostra comparadas com o controle é indicativo de um inversão.

56. Método para detectar a localização de um deleção, **CARACTERIZADO** pelo fato de que compreende as etapas de:

(a) realizar as etapas (a) a (i) conforme definidas nas reivindicações 1-11; e

(b) comparar a frequência de interação entre as

seqüências de DNA com a de um controle;

em que uma redução na frequência de interação de DNA-DNA para a amostra comparada com o controle é indicativa de deleção.

5 57. Método para detectar a localização de um duplicação, **CARACTERIZADO** pelo fato de que compreende as etapas de:

(a) realizar as etapas (a) a (i) conforme definidas nas reivindicações 1-11; e

10 (b) comparar a frequência de interação entre as seqüências de DNA com a de um controle;

em que um aumento ou uma diminuição na frequência de interação de DNA-DNA para a amostra do sujeito comparado com o controle é indicativo de uma duplicação ou inserção.

15 58. Agente, **CARACTERIZADO** pelo fato de que é obtido ou obtenível pelo método de arranjo, conforme definido na reivindicação 53.

20 59. Uso da seqüência de nucleotídeo, conforme definida na reivindicação 12, **CARACTERIZADO** pelo fato de que é para identificar uma ou mais interações de DNA-DNA em uma amostra.

25 60. Uso da seqüência de nucleotídeo, conforme definida na reivindicação 12, **CARACTERIZADO** pelo fato de que é para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

61. Uso do arranjo de sondas, conforme definido na reivindicação 15, ou do conjunto de sondas, conforme

definido em qualquer uma das reivindicações 16-27 ou 30, **CARACTERIZADO** pelo fato de que é para identificar uma ou mais interações de DNA-DNA em uma amostra.

62. Uso do arranjo de sondas, conforme definido na
5 reivindicação 15, ou do conjunto de sondas, conforme definido em qualquer uma das reivindicações 16-27 ou 30, **CARACTERIZADO** pelo fato de que é para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

10 63. Uso do arranjo, conforme definido em qualquer uma das reivindicações 31-36 ou 39, **CARACTERIZADO** pelo fato de que é para identificar uma ou mais interações de DNA-DNA em uma amostra.

15 64. Uso do arranjo, conforme definido em qualquer uma das reivindicações 31-36 ou 39, **CARACTERIZADO** pelo fato de que é para o diagnóstico ou prognóstico de uma doença ou síndrome causada ou associada com uma mudança em uma interação de DNA-DNA.

20 65. Uso, de acordo com qualquer uma das reivindicações 61, 63 e 65, **CARACTERIZADO** pelo fato de que o diagnóstico ou prognóstico é diagnóstico ou prognóstico pré-natal.

25 66. Método, um arranjo de sondas, um conjunto de sondas, um processo, um arranjo, um método de ensaio, um agente ou um uso, **CARACTERIZADO** pelo fato de que é substancialmente da forma aqui descrita e com referência a qualquer dos Exemplos ou Figuras.

Figura 1

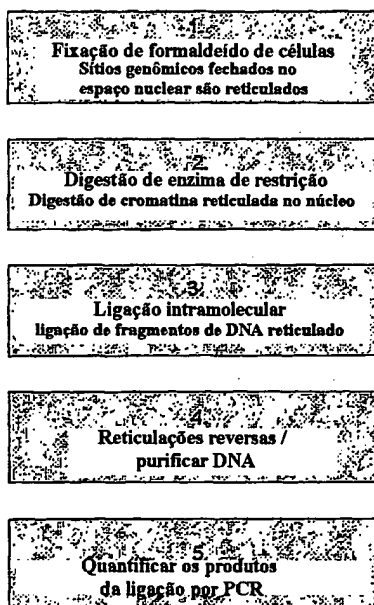


Figura. 2

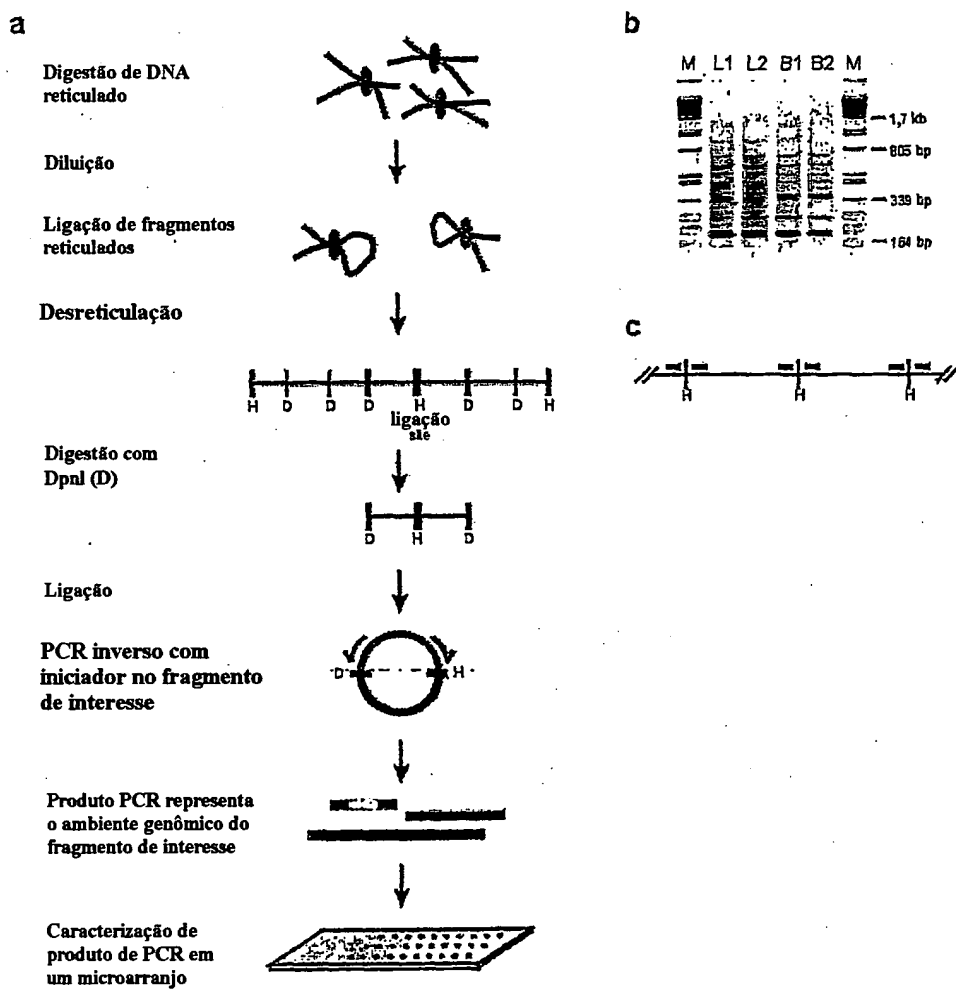


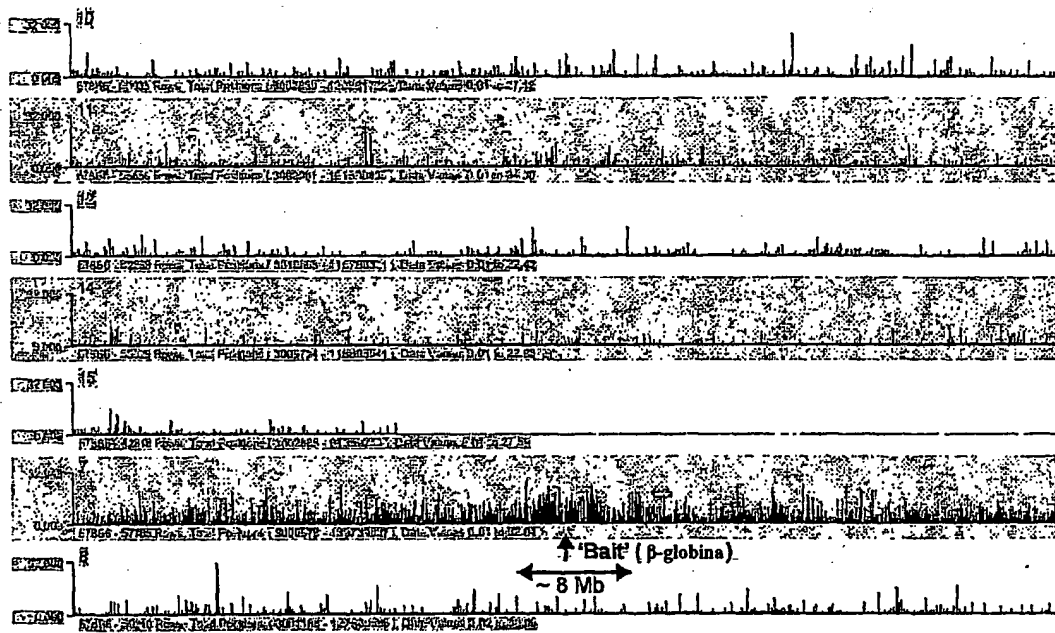
Figura 3

Figura 4

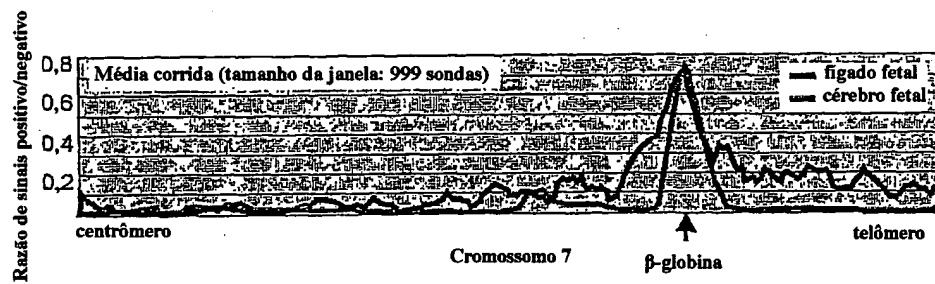
Figura 5

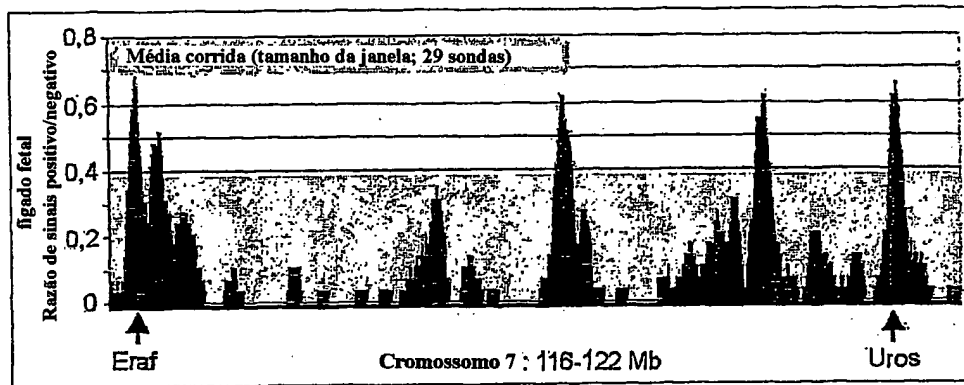
Figura 6

Figura 7

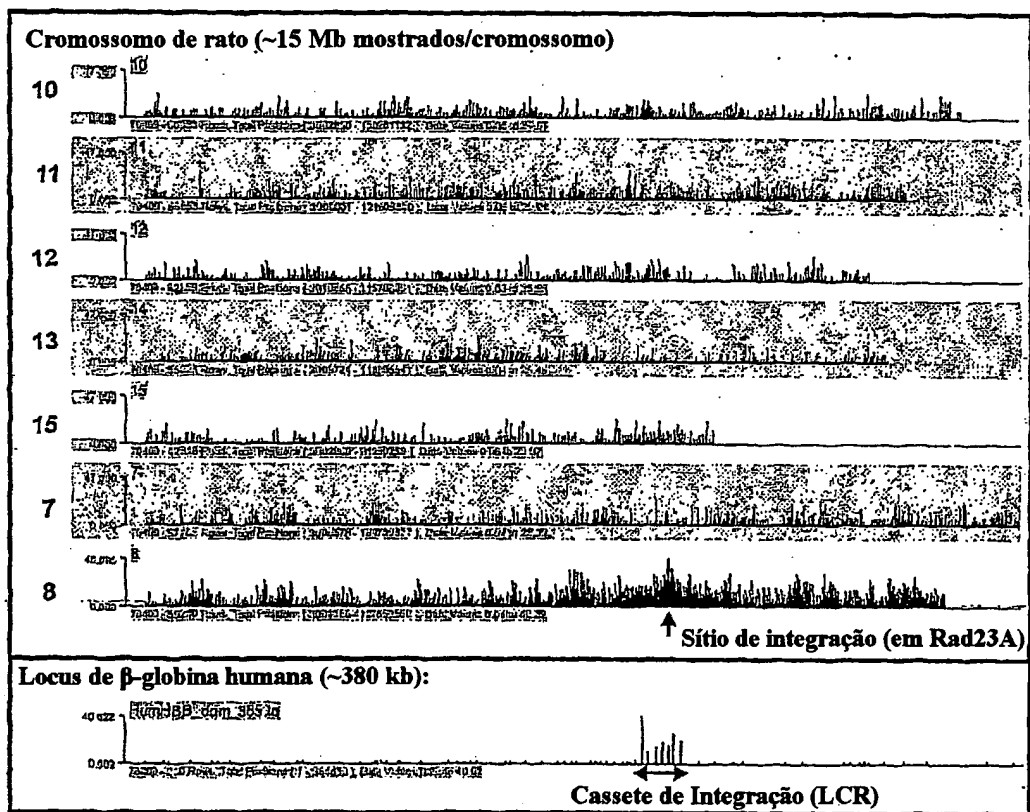


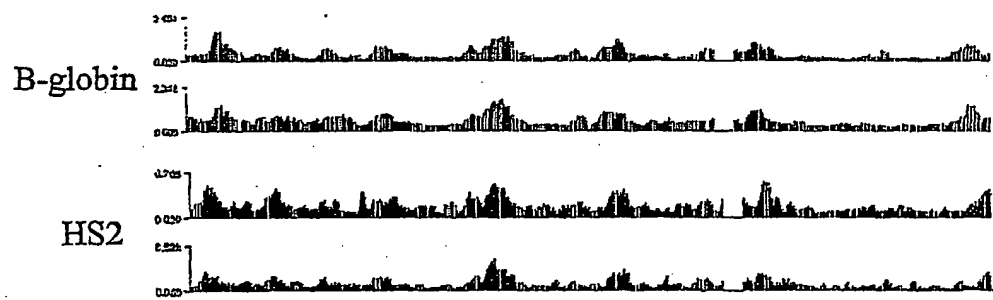
Figura 8

Figura 9

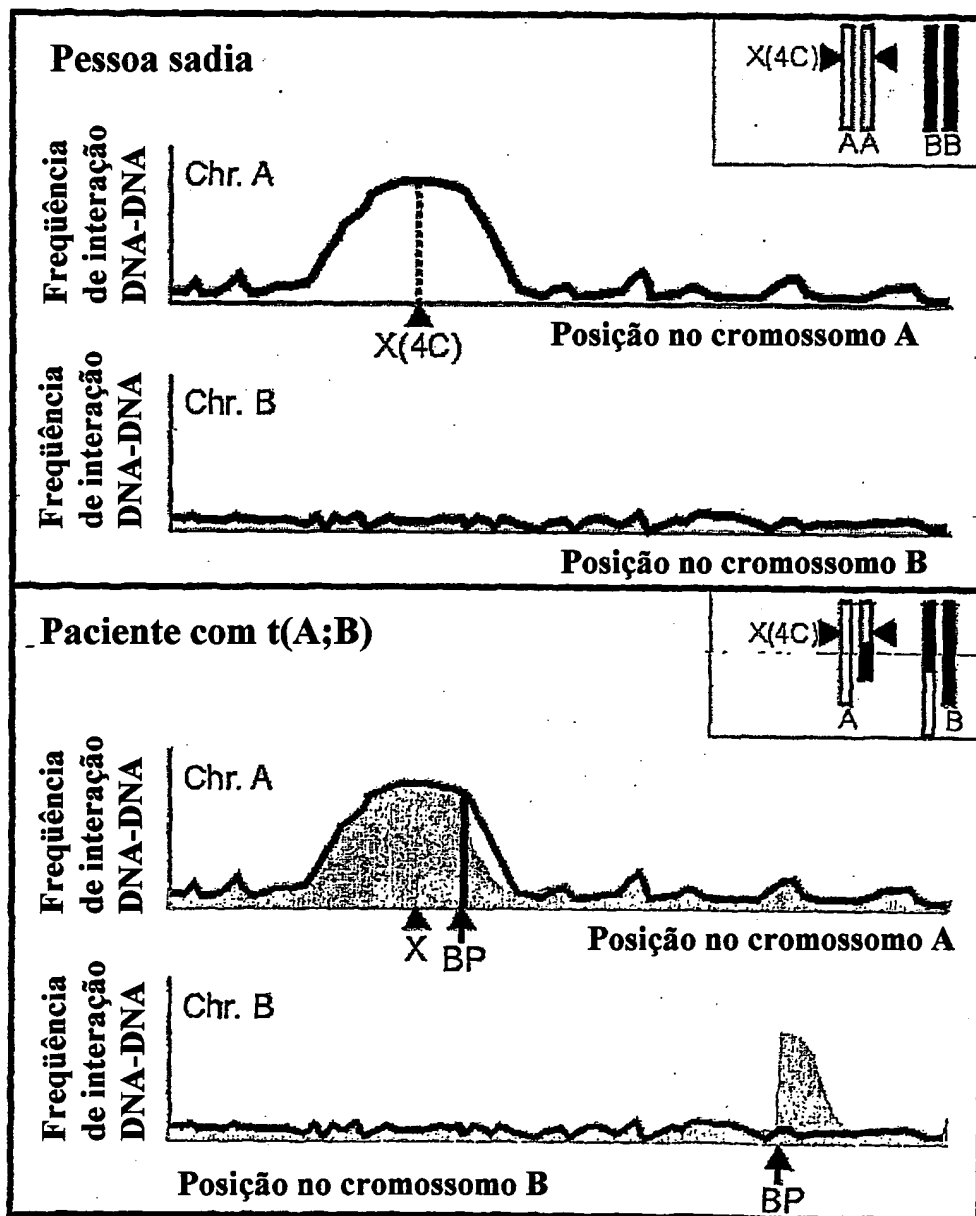


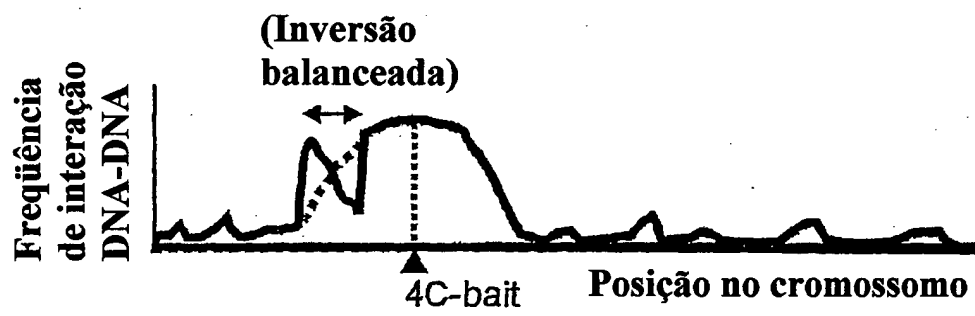
Figura 10

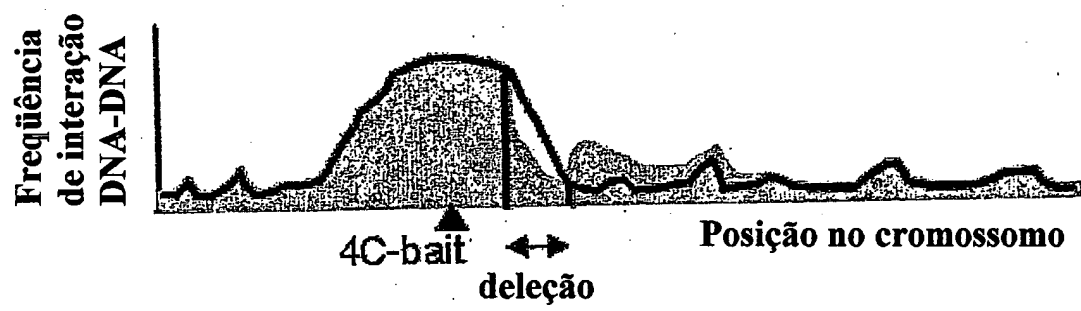
Figura 11

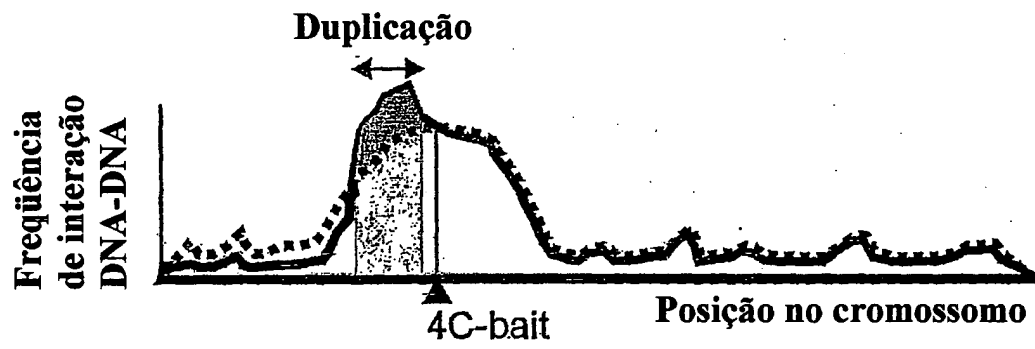
Figura 12

Figura 13

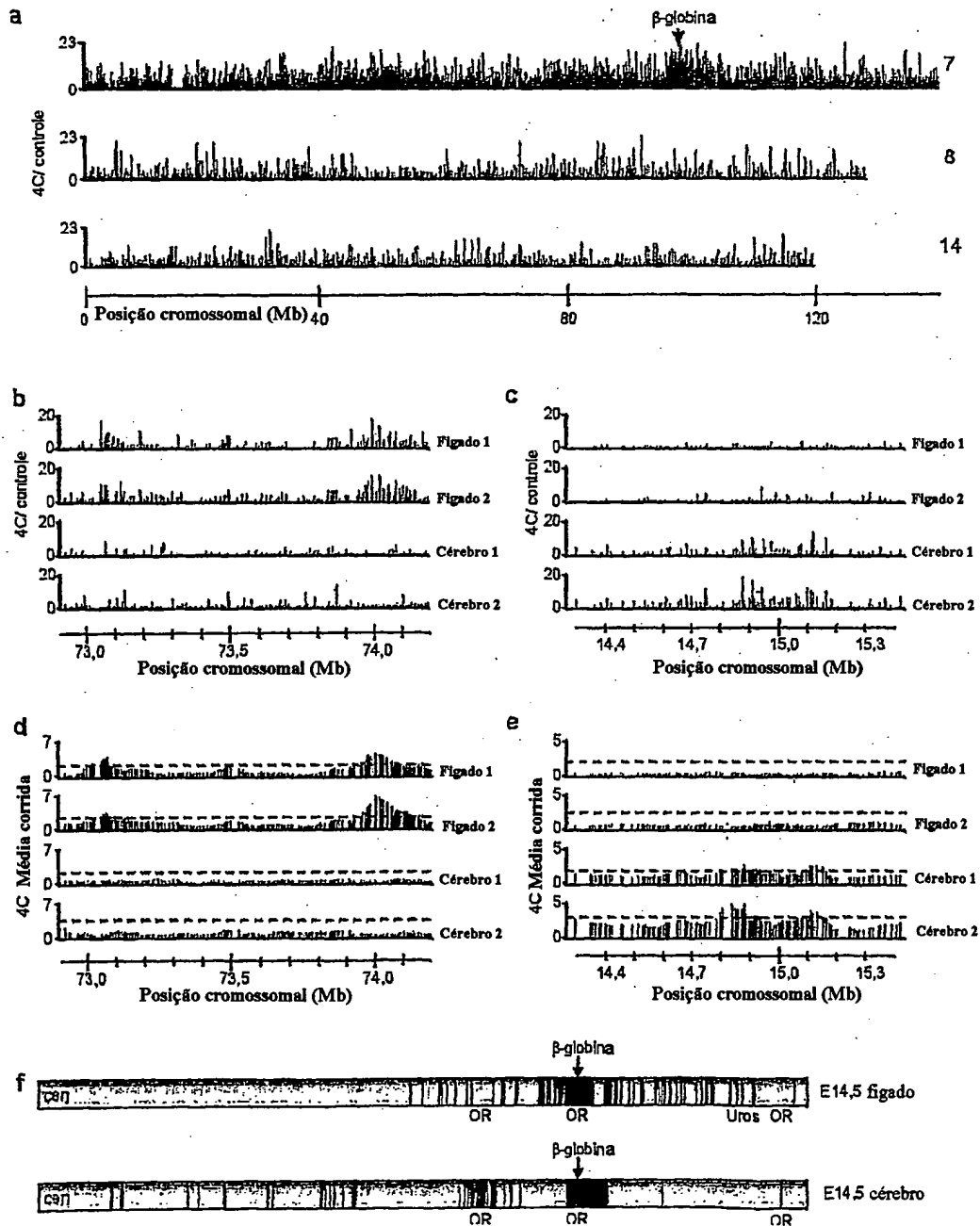


Figura 14

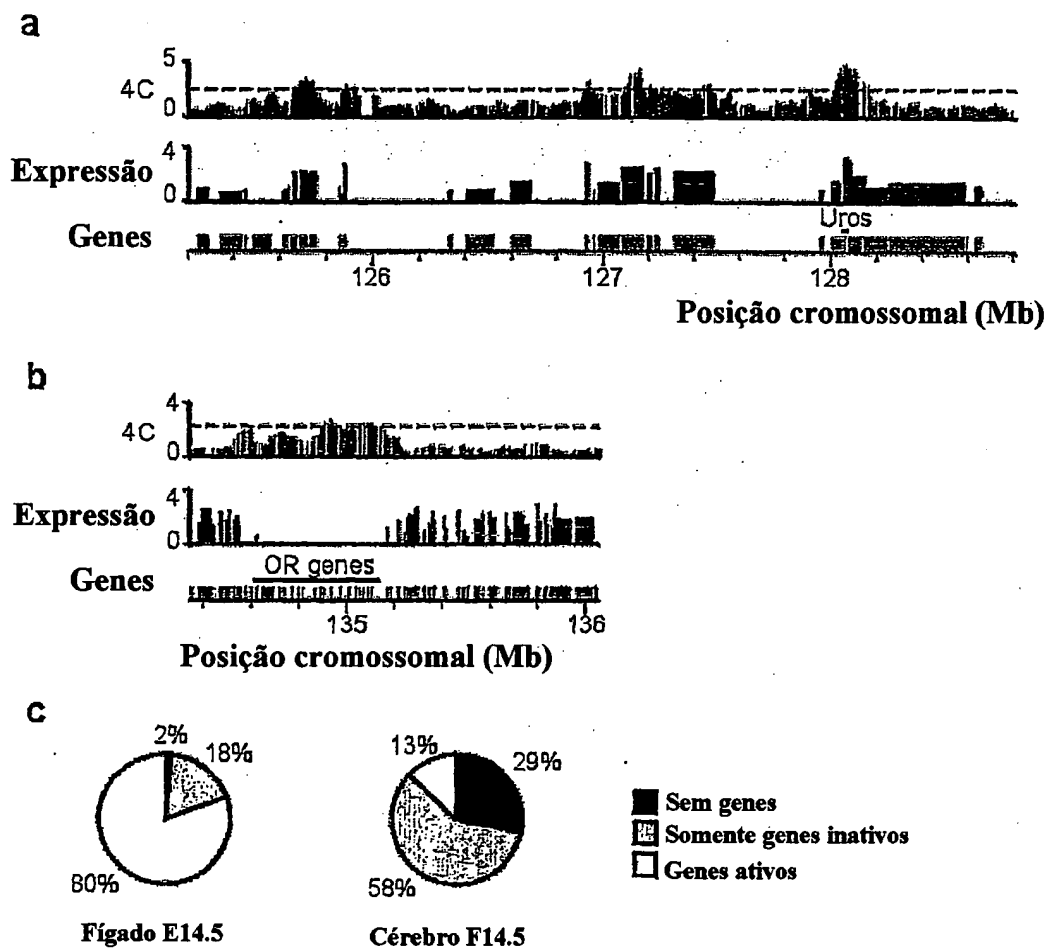


Figura 15

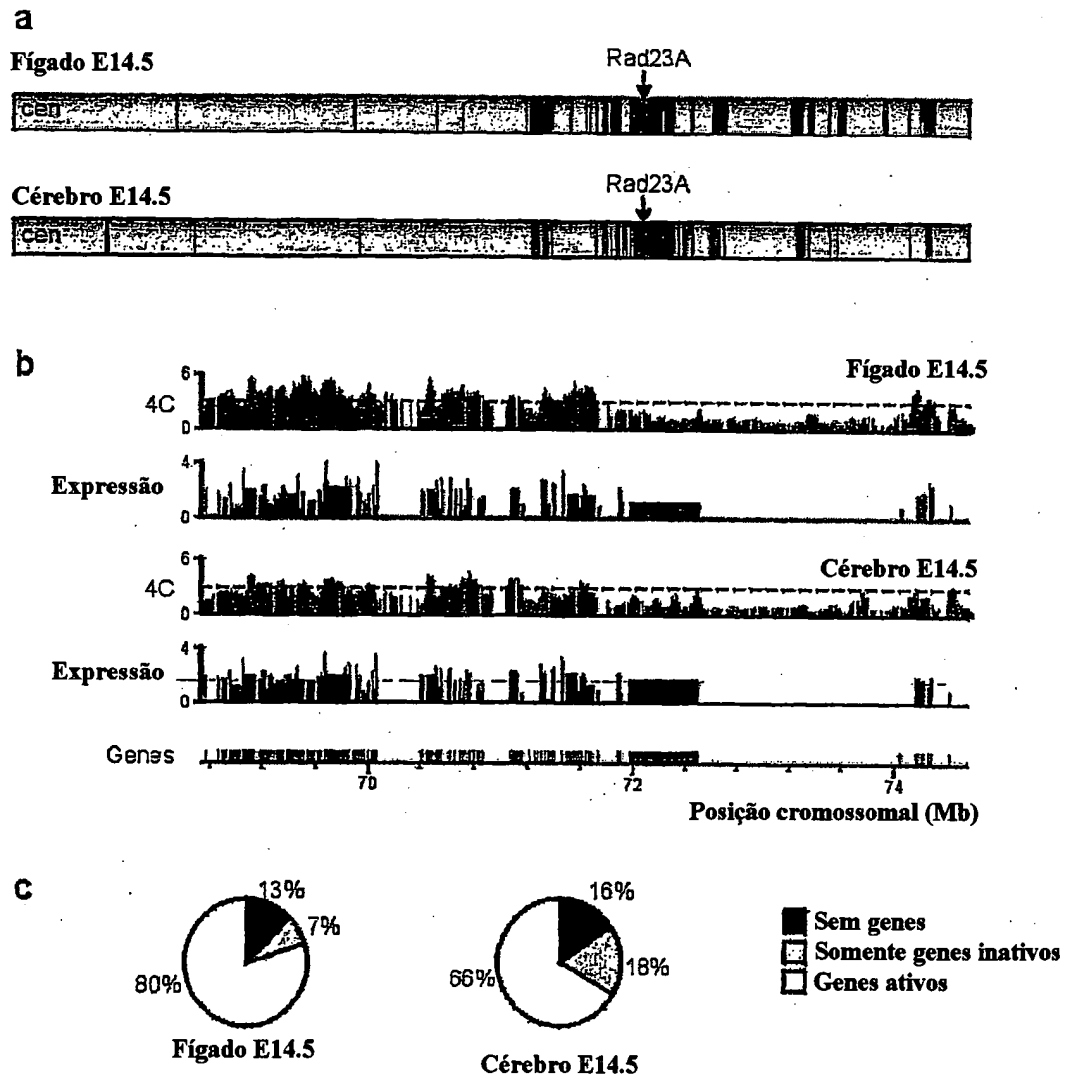


Figura 16

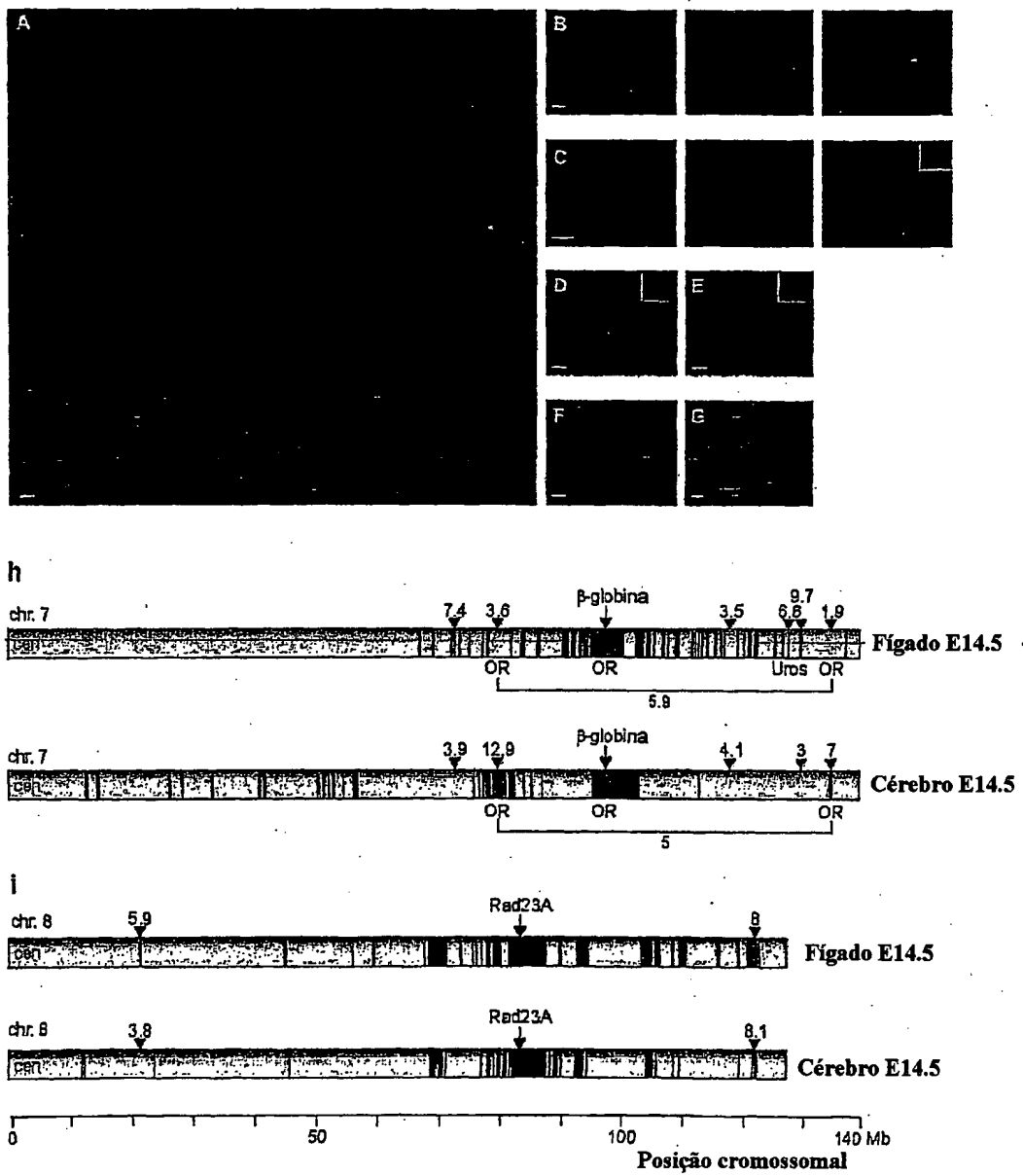


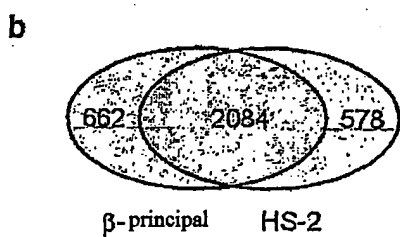
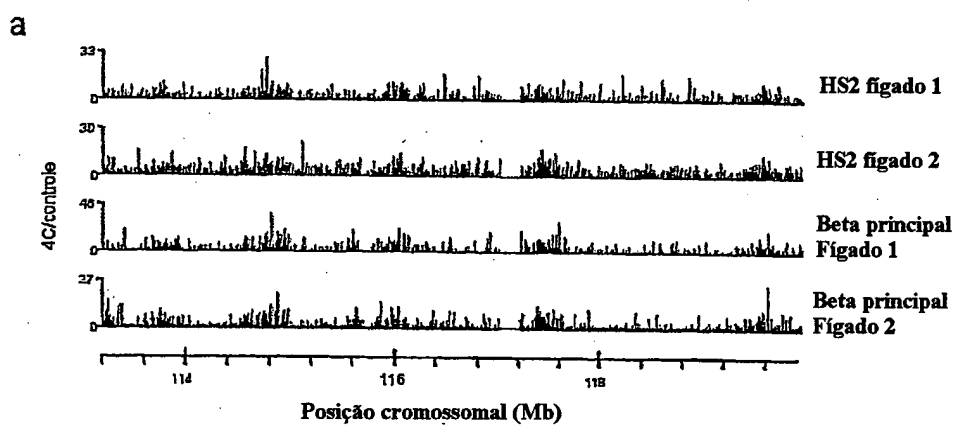
Figura 17

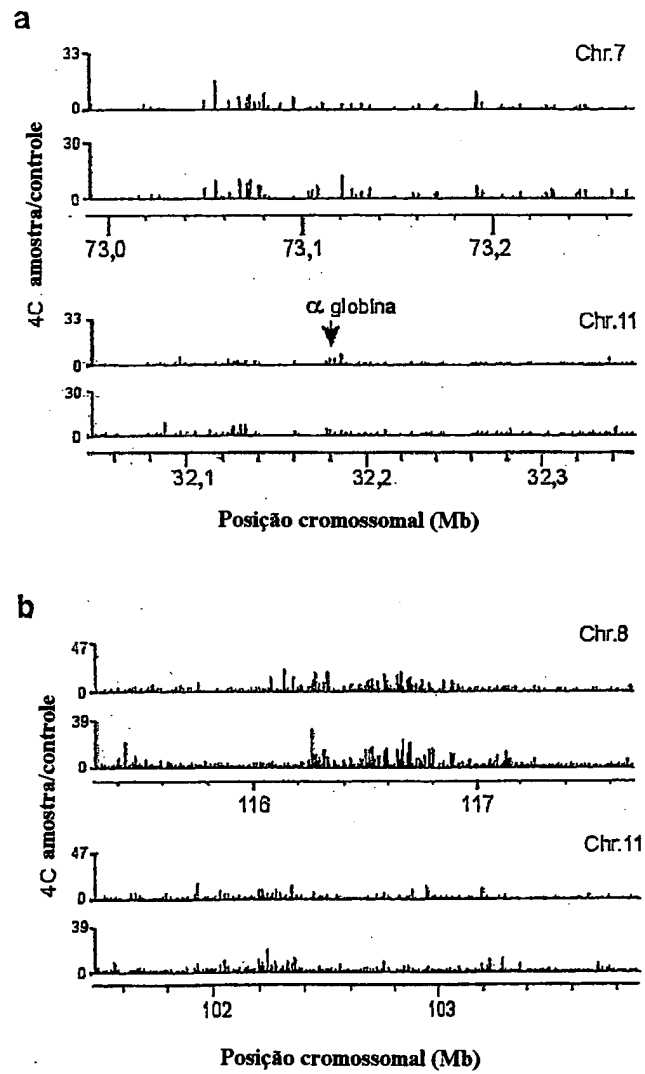
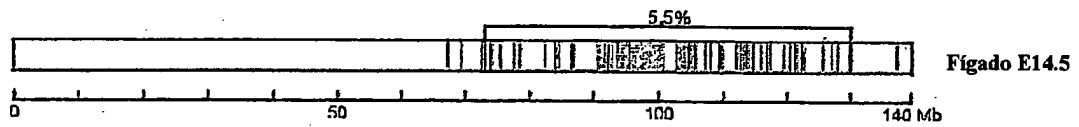
Figura 18

Figura 19

RESUMO

"CAPTURA E CARACTERIZAÇÃO DE CROMATINA CO-LOCALIZADA (4C)"

A presente invenção diz respeito em um aspecto a
5 um método para analisar a frequência de interação de uma
seqüência de nucleotídeos alvo com uma ou mais seqüências de
nucleotídeos de interesse (por exemplo, um ou mais loci
genômicos) que compreende as etapas de: (a) fornecer uma
amostra de DNA reticulado; (b) digerir o DNA reticulado com
10 uma enzima de restrição primária; (c) ligar as seqüências de
nucleotídeos reticuladas; (d) reverter a reticulação; (e)
digerir as seqüências de nucleotídeos com uma enzima de
restrição secundária; (f) ligar uma ou mais seqüências de
DNA de composição de nucleotídeo conhecida ao (s) sítio (s)
15 de digestão de enzima de restrição secundária disponível que
flanqueiam uma ou mais seqüências de nucleotídeos de
interesse; (g) amplificar uma ou mais seqüências de
nucleotídeos de interesse usando pelo menos dois iniciadores
de oligonucleotídeo, em que cada iniciador hibridiza para as
20 seqüências de DNA que flanqueiam as seqüências de
nucleotídeos de interesse; (h) hibridizar as seqüência (s)
amplificada(s) em um arranjo; e (i) determinar a frequência
de interação entre as seqüências de DNA.