

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-523201

(P2004-523201A)

(43) 公表日 平成16年8月5日(2004.8.5)

(51) Int.CI.⁷

F 1

テーマコード(参考)

C12N 15/09

C12N 15/00

Z N A A

2 G O 4 5

C12Q 1/68

C12Q 1/68

A

2 G O 5 4

G01N 21/78

G01N 21/78

C

4 B O 2 4

G01N 33/53

G01N 33/53

M

4 B O 6 3

G01N 33/566

G01N 33/566

審査請求 未請求 予備審査請求 有 (全 116 頁) 最終頁に続く

(21) 出願番号	特願2001-585298 (P2001-585298)	(71) 出願人	502416567 ザ テルドレンズ マーシー ホスピタル アメリカ合衆国 ミズーリ州 64108 カンザス シティ ギルハム ロード 2401
(86) (22) 出願日	平成13年5月15日 (2001.5.15)	(74) 代理人	100080159 弁理士 渡辺 望穂
(85) 翻訳文提出日	平成14年11月18日 (2002.11.18)	(74) 代理人	100090217 弁理士 三和 晴子
(86) 國際出願番号	PCT/US2001/015674	(72) 発明者	ノール ジョーン エイチ. エム. アメリカ合衆国 カンザス州 66213 オーバーランド パーク ジレット ロ ード 12226
(87) 國際公開番号	W02001/088089		
(87) 國際公開日	平成13年11月22日 (2001.11.22)		
(31) 優先権主張番号	09/573,080		
(32) 優先日	平成12年5月16日 (2000.5.16)		
(33) 優先権主張国	米国(US)		
(31) 優先権主張番号	09/854,867		
(32) 優先日	平成13年5月14日 (2001.5.14)		
(33) 優先権主張国	米国(US)		

最終頁に続く

(54) 【発明の名称】シングルコピーノムのハイブリダイゼーションプローブおよびその作製方法

(57) 【要約】

本発明は、前記の問題を解決し、既知の標的核酸のシングルコピーと推測される配列領域とハイブリダイゼーションする、標識されたシングルコピー核酸を含有する核酸（例えば、DNA）ハイブリダイゼーションプローブを提供するものである。概して、本発明のプローブは、標的核酸の配列を、標的がその一部分となっているゲノム中の既知の繰り返し配列と比較することにより、設計されている；標的中のシングルコピー配列（すなわち、特異性がないことにより、シングルコピー配列のハイブリダイゼーションシグナルをふさいでしまう可能性のあるような、繰り返し配列を全く含んでいない配列）と導出できる情報がある。想像されるように、この初期段階では、標的とゲノム上の繰り返し配列の知見が必要であり、これらの情報は、ヒトゲノム計画やそれに関連したバイオインフォマティクスの研究により、飛躍的に利用できるようになってきている。さらに、必要とするシングルコピー配列を検索するために、市販されているコンピューターソフトウェアを利用することができます。

ここでいうプローブとは、標的配列に相補性であるものが最も好ましい；即ち、プローブであるヌクレオチドと標的である配列が、100%相補性があるものである。さらに広く言えば、プローブが、標的である配列と適切にハイブリダイゼーションする限り、相補性が100%より低いものも使用することができる；すなわち、プローブと標的配列の相補鎖である配列とが、少なくとも80%の相同性があったほうがよい。より好ましくは、少なくとも約90%である。

シングルコピーと推測される配列に対応する核酸断片は、PCR増幅、精製したゲノム断

10

20

【特許請求の範囲】**【請求項 1】**

既知の配列である標的核酸中のシングルコピーと導出された配列にハイブリダイゼーションする、標識されたシングルコピーである核酸を含有する核酸ハイブリダイゼーションプローブで、該核酸プローブが少なくとも約 50 ヌクレオチドの長さを有することを特徴とするプローブ。

【請求項 2】

前記プローブが、該標的核酸中の各々がシングルコピーであると導出された配列にハイブリダイゼーションする、複数の異なる標識された核酸を含有し、該核酸プローブのそれぞれが少なくとも約 50 ヌクレオチドの長さを有することを特徴とする請求項 1 に記載のプローブ。10

【請求項 3】

前記核酸プローブが、少なくとも約 100 ヌクレオチドの長さを有することを特徴とする請求項 1 に記載のプローブ。

【請求項 4】

前記核酸プローブが、少なくとも約 2000 ヌクレオチドの長さを有することを特徴とする請求項 3 に記載のプローブ。

【請求項 5】

該標的核酸が DNA, RNA, mRNA からなる群の中から選択されることを特徴とする請求項 1 に記載のプローブ。20

【請求項 6】

該標的核酸が DNA であることを特徴とする請求項 5 に記載のプローブ。

【請求項 7】

該核酸プローブが 1 本鎖であることを特徴とする請求項 1 に記載のプローブ。

【請求項 8】

前記プローブが、標的核酸がその一部となっているゲノムの繰り返し配列にハイブリダイゼーションする、プロッキング核酸配列を必然的に含んでいないことを特徴とする請求項 1 に記載のプローブ。

【請求項 9】

前記核酸プローブが、蛍光色素応答性ラベル、蛍光色素、発色性化学物質、複合タンパク質、抗体、抗原、およびその混合物からなる群から選択される標識ラベルで標識されていることを特徴とする請求項 1 に記載のプローブ。30

【請求項 10】

前記核酸プローブが、蛍光色素応答性ラベルで標識されていることを特徴とする請求項 9 に記載のプローブ。

【請求項 11】

前記プローブと該標識配列と相補的な配列とが、少なくとも約 80 % の相同性を有することを特徴とする請求項 1 に記載のプローブ。

【請求項 12】

前記プローブが、該標的配列に相補性であることを特徴とする請求項 11 に記載のプローブ。40

【請求項 13】

標的核酸配列および該標的配列の少なくとも一部にハイブリダイゼーションする核酸プローブを含有する反応溶液を調製し、前記プローブの該標的核酸配列へのハイブリダイゼーションを引き起こす段階を含むハイブリダイゼーション方法において、既知の配列の標的核酸中のシングルコピーと導出された配列にハイブリダイズする標識されたシングルコピー核酸を、前記プローブとして使用し、核酸プローブが少なくとも約 50 ヌクレオチドの長さを有することを含むことを特徴とするハイブリダイゼーションの改良方法。

【請求項 14】

前記プローブが、該標的核酸の各々がシングルコピーであると推論される配列間隔にハイ50

ブリダイズする、複数の互いに異なる標識された核酸を含有し、該核酸プローブのそれぞれが、少なくとも約50ヌクレオチドの長さを有することを特徴とする請求項13に記載のハイブリダイゼーション方法。

【請求項15】

前記核酸プローブが、少なくとも約100ヌクレオチドの長さを有することを特徴とする請求項13に記載のハイブリダイゼーション方法。

【請求項16】

前記核酸プローブが少なくとも約2000ヌクレオチドの長さを有することを特徴とする請求項15に記載のハイブリダイゼーション方法。

【請求項17】

該標的核酸がDNA, RNA, mRNAからなる群の中から選択されることを特徴とする請求項13に記載のハイブリダイゼーション方法。 10

【請求項18】

該標的核酸がDNAであることを特徴とする請求項17に記載のハイブリダイゼーション方法。

【請求項19】

前記核酸プローブが1本鎖であることを特徴とする請求項13に記載のハイブリダイゼーション方法。

【請求項20】

前記プローブが、標的核酸がその一部となっているゲノムの繰り返し配列にハイブリダイゼーションする、プロッキング核酸配列を必然的に含んでいないことを特徴とする請求項13に記載のハイブリダイゼーション方法。 20

【請求項21】

該核酸プローブが、蛍光色素応答性ラベル、蛍光色素、発色性化学物質、複合タンパク質、抗体、抗原、およびその混合物からなる群から選択される標識ラベルで標識されていることを特徴とする請求項13に記載のハイブリダイゼーション方法。

【請求項22】

前記核酸プローブが、蛍光色素応答性ラベルで標識されていることを特徴とする請求項21に記載のハイブリダイゼーション方法。

【請求項23】

前記ハイブリダイゼーション方法が、in situハイブリダイゼーション、サザンプロッティング、核酸が固定されているその他の方法からなる群から選択されることを特徴とする請求項13に記載のハイブリダイゼーション方法。 30

【請求項24】

前記プローブと該標識配列の一部である配列に、少なくとも約80%の相同性を有することを特徴とする請求項13に記載のハイブリダイゼーション方法。

【請求項25】

前記プローブが、該標的配列に相補性であることを特徴とする請求項24に記載のハイブリダイゼーション方法。

【請求項26】

ゲノムの一部を形成している標的核酸配列に対するハイブリダイゼーションプローブを作製する方法であり、前記方法が、該標的配列の少なくとも1つのシングルコピーである配列のシークエンスを決定し、かつ該シングルコピー配列の少なくとも一部に、ハイブリダイズするハイブリダイゼーションプローブを構築する段階を含んでいることを特徴とするプローブ作製方法。 40

【請求項27】

該標的核酸配列のシークエンスを決定し、該ゲノム中に見られる繰り返し配列を決定し、および前記の少なくとも1つのシングルコピー配列のシークエンスを決定するために、該標的核酸配列と繰り返し配列のシークエンスを比較する段階を含むことを特徴とする請求項26に記載のプローブ作製方法。 50

【請求項 2 8】

前記プローブ作製方法が、該シングルコピーである配列の少なくとも一部を取得し、かつ該シングルコピー配列の一部を精製する段階を含むことを特徴とする請求項 2 6 に記載のプローブ作製方法。

【請求項 2 9】

前記精製方法が、P C R の実行を含むことを特徴とする請求項 2 8 に記載のプローブ作製方法。

【請求項 3 0】

該ハイブリダイゼーションプローブを標識する段階を含んでいることを特徴とする請求項 2 6 に記載のプローブ作製方法。 10

【請求項 3 1】

該標的核酸配列および前記プローブが、D N A であることを特徴とする請求項 2 6 に記載のプローブ作製方法。

【請求項 3 2】

前記ハイブリダイゼーションプローブが、該シングルコピー配列と、少なくとも 80 % の相同性を有することを特徴とする請求項 2 6 に記載のプローブ作製方法。

【請求項 3 3】

前記ハイブリダイゼーションプローブが、シングルコピーである配列と相補性であることを特徴とする請求項 3 2 に記載のプローブ作製方法。

【請求項 3 4】

配列表の配列番号 4 2 9 ~ 4 4 6 からなる群から選択される 1 つの配列と、少なくとも 80 % の相同性を有することを特徴とする D N A 配列。 20

【請求項 3 5】

前記 D N A 配列が、少なくとも 90 % の相同性を有することを特徴とする請求項 3 4 に記載の D N A 配列。

【請求項 3 6】

配列表の配列番号 4 2 9 ~ 4 4 6 および 4 8 0 ~ 6 1 3 からなる群から選ばれることを特徴とする D N A 配列。

【請求項 3 7】

配列番号 4 2 9 で始まり、配列番号 4 2 9 ~ 4 4 6 と 4 8 0 ~ 6 1 3 として同定される配列の近傍の一対を、P C R プライマーとして使用し合成された P C R 増幅配列と、少なくとも 80 % の相同性を有することを特徴とする D N A 配列。 30

【請求項 3 8】

少なくとも 90 % の相同性を有することを特徴とする請求項 3 7 に記載の D N A 配列。

【請求項 3 9】

標識され、シングルコピー核酸と導出されたプローブが、ゲノムと反応し、する段階を含み、プローブのハイブリダイゼーションを引き起こし、かつ、もしプローブがゲノムの互いに異なる 3 つ以上の部位とハイブリダイズするならば、新規な繰り返し配列であると決定できる、ゲノム中の未知の繰り返し配列の存在を同定する方法。

【請求項 4 0】

プローブが 10 ケ所以上の部位とハイブリダイズするならば、新規な繰り返し配列であると決定できる段階を含むことを特徴とする請求項 3 9 に記載の同定方法。 40

【請求項 4 1】

前記核酸が、2 倍化または 3 倍化した配列部分に由来することを特徴とする請求項 1 に記載のプローブ。

【請求項 4 2】

2 倍化または 3 倍化した配列領域にハイブリダイゼーションするシングルコピーである核酸を選択する段階を含むことを特徴とする請求項 1 3 に記載のハイブリダイゼーション方法。

【請求項 4 3】

10

20

30

40

50

前記決定段階が、2倍化または3倍化した配列領域からのシングルコピーである配列を選択する段階を含んでいることを特徴とする請求項26に記載のプローブ作製方法。

【請求項44】

次に示される段階を含む染色体の切断点を同定する方法：

既知配列であり、それぞれ該切断点の両側にハイブリダイゼーションするように設計された、互いに異なり、標識されたシングルコピーである1対の核酸プローブを提供し；該一対のプローブが、該切断点を含む染色体の標的配列と反応して、プローブの標的配列とのハイブリダイゼーションを引き起こし、
該切断部位を決定する手段として、前記のハイブリダイゼーションしたプローブを同定する。

10

【請求項45】

該切断点の片側にハイブリダイズするように設計された、既知の配列に対する複数の互いに異なり、標識されたシングルコピーである核酸プローブを提供し、該切断点のもう一方にハイブリダイズするように設計された、既知の配列に対する複数の互いに異なり、標識されたシングルコピーである核酸プローブを提供することを特徴とする請求項44に記載の染色体切断点の同定方法。

【発明の詳細な説明】

【0001】

技術分野

本発明は、真核生物由来の試料中に含まれる特異的な核酸配列の存在を同定するために、細胞遺伝学および分子遺伝学の分野において有用なシングルコピーハイブリダイゼーションプローブを設計する方法に広く関する；例えば、そのようなプローブは、後天性または先天性の遺伝病の検出手段として、in situハイブリダイゼーションにより、特異的な染色体位置を解析するのに用いることができる。詳しくは、そのようなプローブ、そのプローブを使用したハイブリダイゼーション方法、およびプローブを作製する技術に関する。プローブには、ゲノムの繰り返し配列が必然的に存在せず、従来のプローブでは必要とされていた繰り返し配列を不活化する必要がない。

20

【0002】

背景技術

染色体異常は、後天性または先天性など、様々な遺伝病と関連している。このような染色体異常は、個々の染色体の過剰または欠損（異数性）、染色体の一部分の過剰または欠損（欠失、倍加、重複および標識染色体を含む）、または染色体の再構成の、主に3つのタイプに分類される。後者の範疇には、転座（染色体の一部が、ある染色体上から他の染色体上に移ること）、逆位（染色体の一部分の方向性が逆転すること）、挿入（染色体の一部が、ある染色体から他の染色体の中へ移ること）、および同腕染色体（相同染色体部分に由来する染色体腕）が含まれる。染色体異常は、一連の細胞のみ（モザイク現象）、または全ての細胞に存在しているのかもしれない。先天性または生まれつき保有する様々なタイプの染色体異常は、250人に約1人の頻度で発生するが、そうした異常は、本質的に良性、重篤または致死的なものであるかもしれない。染色体異常は、普遍的な現象であり、白血病やその他の癌といった後天性の病気の診断に役立つことが多い。

30

【0003】

ハイブリダイゼーションプローブは、染色体の解析や染色体異常の診断のために、従来開発された。プローブは、クローン化、增幅されたゲノム配列、またはcDNAを含んでいる。例えば、米国特許5447841、5663319および5756696号明細書には、標的染色体DNAの中の核酸部分に相補性のある、標識された核酸を形成するハイブリダイゼーションプローブについての記述がある。しかし、これらのプローブは、繰り返し配列を含んでおり、それ故、標識されたプローブの中で、繰り返し配列に部分的に相補的なプロッキング核酸と組み合わせて使用しなければならない。つまり、この先行技術に記載されたプローブは、結合させてプローブの中にある繰り返し配列をロックする、プロッキング核酸とあらかじめ反応しておくか、ハイブリダイゼーションを行う反応溶液の

40

50

中にそのようなブロッキング核酸を入れておく必要がある。もし、プローブの中の繰り返し配列が、何らかの形で不活性化されないなら、プローブは、繰り返し配列が存在している標的である染色体DNAの複数の部位と反応し、目的であるシングルコピーである配列と特異的に反応しない。これは、ゲノム全体に広く散在した繰り返し配列では、特に顕著な問題となるが、DNA分子上に塊状または連続している配列でも同様である。相補的なブロッキング核酸を使用して繰り返し配列を不活性化する必要があると、存在しているプローブの感受性も低下してしまう。再現性があつて、簡便にシグナルを検出しようとすると、約40～100kbのDNAプローブを必要とする。

【0004】

前記の先行技術では、シングルコピーである配列を含んでいると推測されるクローン化されたプローブは、放射線標識された全ゲノムDNAには結合しないことにより、同定することができる示されている。この他の研究において、ハイブリダイゼーションは、リコンビナントのDNAが、シングルコピー配列、または繰り返しの非常に少ない配列に結合するように、独立して選択されるクローン群を含むプローブで、まず始めに実施される。この先行技術では、前段階として、候補となっているDNAに対し、サザンハイブリダイゼーションやドットプロットにより、標識されたゲノムDNAの実証的なハイブリダイゼーションを行い、シングルコピー配列であることを確認する。標識された全ゲノムDNAを使って、ハイブリダイゼーションによるシグナルが出ると、候補となっているDNAプローブは、繰り返し配列を含んでおり、シングルコピープローブとして候補から除かれる。さらに、全ゲノムDNAとDNAプローブとの、予備実験的なハイブリダイゼーションでは、コピー数の小さい(<100コピー)またはゲノム中に頻繁には出現しないマルチコピーである繰り返し配列は、検出できない可能性がある。

【0005】

実験手法において、プローブから繰り返し配列を、物理的に除去することが提示されている(Craig et al., Hum. Genet., 100:472-476(1997); Durm et al., Biotech., 24:820-825(1998))。この検出手法では、DNA標的に対するプローブを加える前に、ポリメラーゼ連鎖反応(PCR)で増幅されたゲノムプローブを、過剰の精製した繰り返し配列を有するDNAとプレハイブリダイゼーションする過程を含んでいる。その結果、精製したプローブは、繰り返し配列を除去される。この手法は、本質的には、プローブ中の繰り返し配列のハイブリダイゼーションを不可能にする他の手法と非常に良く似ている、しかし、時間がかかることに加え、米国特許5447841および575696号明細書に記載されているプローブに対して、何ら進歩性がないものである。

[発明が解決しようとする課題]

本発明は、繰り返し配列と相補的なブロッキング核酸と組み合わせて使用する必要がなく、過剰の精製した繰り返し配列を有するDNAとプレハイブリダイゼーションする過程も要しないシングルコピーであるハイブリダイゼーションプローブを提供することを目的とする。

【0006】

[課題を解決するための手段]

(1) 既知の配列である標的核酸中のシングルコピーと導出された配列にハイブリダイゼーションする、標識されたシングルコピーである核酸を含有する核酸ハイブリダイゼーションプローブで、該核酸プローブが少なくとも約50ヌクレオチドの長さを有することを特徴とするプローブを提供する。

【0007】

(2) 前記プローブが、該標的核酸中の各々がシングルコピーであると導出された配列にハイブリダイゼーションする、複数の異なる標識された核酸を含有し、該核酸プローブのそれぞれが少なくとも約50ヌクレオチドの長さを有することを特徴とする(1)に記載のプローブを提供する。

【0008】

10

20

30

40

50

(3) 前記核酸プローブが、少なくとも約100ヌクレオチドの長さを有することを特徴とする(1)に記載のプローブを提供する。

【0009】

(4) 前記核酸プローブが、少なくとも約2000ヌクレオチドの長さを有することを特徴とする(3)に記載のプローブを提供する。

【0010】

(5) 該標的核酸がDNA, RNA, mRNAからなる群の中から選択されることを特徴とする(1)に記載のプローブを提供する。

【0011】

(6) 該標的核酸がDNAであることを特徴とする(5)に記載のプローブを提供する。 10

【0012】

(7) 該核酸プローブが1本鎖であることを特徴とする(1)に記載のプローブを提供する。

【0013】

(8) 前記プローブが、標的核酸がその一部となっているゲノムの繰り返し配列にハイブリダイゼーションする、ブロッキング核酸配列を必然的に含んでいないことを特徴とする(1)に記載のプローブを提供する。

【0014】

(9) 前記核酸プローブが、蛍光色素応答性ラベル、蛍光色素、発色性化学物質、複合タンパク質、抗体、抗原、およびその混合物からなる群から選択される標識ラベルで標識されていることを特徴とする(1)に記載のプローブを提供する。 20

【0015】

(10) 前記核酸プローブが、蛍光色素応答性ラベルで標識されていることを特徴とする(9)に記載のプローブを提供する。

【0016】

(11) 前記プローブと該標識配列と相補的な配列とが、少なくとも約80%の相同性を有することを特徴とする(1)に記載のプローブを提供する。

【0017】

(12) 前記プローブが、該標的配列に相補性であることを特徴とする(11)に記載のプローブを提供する。 30

【0018】

(13) 標的核酸配列および該標的配列の少なくとも一部にハイブリダイゼーションする核酸プローブを含有する反応溶液を調製し、前記プローブの該標的核酸配列へのハイブリダイゼーションを引き起こす段階を含むハイブリダイゼーション方法において、既知の配列の標的核酸中のシングルコピーと導出された配列にハイブリダイズする標識されたシングルコピー核酸を、前記プローブとして使用し、核酸プローブが少なくとも約50ヌクレオチドの長さを有することを含むことを特徴とするハイブリダイゼーションの改良方法を提供する。

【0019】

(14) 前記プローブが、該標的核酸の各々がシングルコピーであると推論される配列間にハイブリダイズする、複数の互いに異なる標識された核酸を含有し、該核酸プローブのそれぞれが、少なくとも約50ヌクレオチドの長さを有することを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。 40

【0020】

(15) 前記核酸プローブが、少なくとも約100ヌクレオチドの長さを有することを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。

【0021】

(16) 前記核酸プローブが少なくとも約2000ヌクレオチドの長さを有することを特徴とする(15)に記載のハイブリダイゼーション方法を提供する。

【0022】

(17) 該標的核酸がDNA, RNA, mRNAからなる群の中から選択されることを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。

【0023】

(18) 該標的核酸がDNAであることを特徴とする(17)に記載のハイブリダイゼーション方法を提供する。

【0024】

(19) 前記核酸プローブが1本鎖であることを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。

【0025】

(20) 前記プローブが、標的核酸がその一部となっているゲノムの繰り返し配列にハイブリダイゼーションする、ブロッキング核酸配列を必然的に含んでいないことを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。 10

【0026】

(21) 該核酸プローブが、蛍光色素応答性ラベル、蛍光色素、発色性化学物質、複合タンパク質、抗体、抗原、およびその混合物からなる群から選択される標識ラベルで標識されていることを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。

【0027】

(22) 前記核酸プローブが、蛍光色素応答性ラベルで標識されていることを特徴とする(21)に記載のハイブリダイゼーション方法を提供する。

【0028】

(23) 前記ハイブリダイゼーション方法が、in situハイブリダイゼーション、サザンプロットティング、核酸が固定されているその他の方法からなる群から選択されることを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。 20

【0029】

(24) 前記プローブと該標識配列の一部である配列に、少なくとも約80%の相同性を有することを特徴とする(13)に記載のハイブリダイゼーション方法を提供する。

【0030】

(25) 前記プローブが、該標的配列に相補性であることを特徴とする請求項24に記載のハイブリダイゼーション方法を提供する。

【0031】

(26) ゲノムの一部を形成している標的核酸配列に対するハイブリダイゼーションプローブを作製する方法であり、前記方法が、該標的配列の少なくとも1つのシングルコピーである配列のシークエンスを決定し、かつ該シングルコピー配列の少なくとも一部に、ハイブリダイズするハイブリダイゼーションプローブを構築する段階を含んでいることを特徴とするプローブ作製方法を提供する。 30

【0032】

(27) 該標的核酸配列のシークエンスを決定し、該ゲノム中に見られる繰り返し配列を決定し、および前記の少なくとも1つのシングルコピー配列のシークエンスを決定するために、該標的核酸配列と繰り返し配列のシークエンスを比較する段階を含むことを特徴とする(26)に記載のプローブ作製方法を提供する。

【0033】

(28) 前記プローブ作製方法が、該シングルコピーである配列の少なくとも一部を取得し、かつ該シングルコピー配列の一部を精製する段階を含むことを特徴とする(26)に記載のプローブ作製方法を提供する。 40

【0034】

(29) 前記精製方法が、PCRの実行を含むことを特徴とする(28)に記載のプローブ作製方法を提供する。

【0035】

(30) 該ハイブリダイゼーションプローブを標識する段階を含んでいることを特徴とする(26)に記載のプローブ作製方法を提供する。 50

【 0 0 3 6 】

(3 1) 該標的核酸配列および前記プローブが、D N A あることを特徴とする(2 6)に記載のプローブ作製方法を提供する。

【 0 0 3 7 】

(3 2) 前記ハイブリダイゼーションプローブが、該シングルコピー配列と、少なくとも 8 0 % の相同性を有することを特徴とする(2 6)に記載のプローブ作製方法を提供する。

【 0 0 3 8 】

(3 3) 前記ハイブリダイゼーションプローブが、シングルコピーである配列と相補性であることを特徴とする(3 2)に記載のプローブ作製方法を提供する。 10

【 0 0 3 9 】

(3 4) 配列表の配列番号 4 2 9 ~ 4 4 6 からなる群から選択される 1 つの配列と、少なくとも 8 0 % の相同性を有することを特徴とする D N A 配列を提供する。

【 0 0 4 0 】

(3 5) 前記 D N A 配列が、少なくとも 9 0 % の相同性を有することを特徴とする(3 4)に記載の D N A 配列を提供する。

【 0 0 4 1 】

(3 6) 配列表の配列番号 4 2 9 ~ 4 4 6 および 4 8 0 ~ 6 1 3 からなる群から選ばれることを特徴とする D N A 配列を提供する。

【 0 0 4 2 】

(3 7) 配列番号 4 2 9 で始まり、配列番号 4 2 9 ~ 4 4 6 と 4 8 0 ~ 6 1 3 として同定される配列の近傍の一対を、 P C R プライマーとして使用し合成された P C R 増幅配列と、少なくとも 8 0 % の相同性を有することを特徴とする D N A 配列を提供する。 20

【 0 0 4 3 】

(3 8) 少なくとも 9 0 % の相同性を有することを特徴とする(3 7)に記載の D N A 配列を提供する。

【 0 0 4 4 】

(3 9) 標識され、シングルコピー核酸と導出されたプローブが、ゲノムと反応し、する段階を含み、プローブのハイブリダイゼーションを引き起こし、かつ、もしプローブがゲノムの互いに異なる 3 つ以上の部位とハイブリダイズするならば、新規な繰り返し配列であると決定できる、ゲノム中の未知の繰り返し配列の存在を同定する方法を提供する。 30

【 0 0 4 5 】

(4 0) プローブが 1 0 ケ所以上の部位とハイブリダイズするならば、新規な繰り返し配列であると決定できる段階を含むことを特徴とする(3 9)に記載の同定方法を提供する。

【 0 0 4 6 】

(4 1) 前記核酸が、 2 倍化または 3 倍化した配列部分に由来することを特徴とする(1)に記載のプローブを提供する。

【 0 0 4 7 】

(4 2) 2 倍化または 3 倍化した配列領域にハイブリダイゼーションするシングルコピーである核酸を選択する段階を含むことを特徴とする(1 3)に記載のハイブリダイゼーション方法を提供する。 40

【 0 0 4 8 】

(4 3) 前記決定段階が、 2 倍化または 3 倍化した配列領域からのシングルコピーである配列を選択する段階を含んでいることを特徴とする(2 6)に記載のプローブ作製方法を提供する。

【 0 0 4 9 】

(4 4) 次に示される段階を含む染色体の切断点を同定する方法を提供する： 既知配列であり、それぞれ該切断点の両側にハイブリダイゼーションするように設計された、互いに異なり、標識されたシングルコピーである 1 対の核酸プローブを提供し； 50

該一対のプローブが、該切断点を含む染色体の標的配列と反応して、プローブの標的配列とのハイブリダイゼーションを引き起こし、該切断部位を決定する手段として、前記のハイブリダイゼーションしたプローブを同定する。

【0050】

(45) 該切断点の片側にハイブリダイズするように設計された、既知の配列に対する複数の互いに異なり、標識されたシングルコピーである核酸プローブを提供し、該切断点のもう一方にハイブリダイズするように設計された、既知の配列に対する複数の互いに異なり、標識されたシングルコピーである核酸プローブを提供することを特徴とする(44)に記載の染色体切断点の同定方法を提供する。 10

【0051】

発明を実施するための最良の形態

本発明は、遺伝病や腫瘍の検出に有効な核酸(例えば、DNA)ハイブリダイゼーションプローブに関する。本発明のプローブは、標的配列へのハイブリダイゼーションを検出できる、標識した核酸断片または標識した核酸断片の集合体を形成している。本発明は、上記のプローブの設計、作製および標識する方法、ならびにそのプローブの使用方法に関する。

【0052】

本発明の標識されたプローブは、繰り返し配列を含んでいるかもしれない、どのような標的核酸にも使用することが可能である。このような標的核酸としては、転写産物と一体化した構成要素として繰り返し配列を有する染色体または精製された核のDNA、hnRNAまたはmRNAが挙げられるが、これらに限定されるものではない。後述の本発明の詳細な説明の中で、DNA標的配列やDNAプローブの使用例が述べられている；しかし、この技術に精通した人であれば、(標的配列及びプローブの性質に従って、認識される技術が異なるだけで)これ以外の核酸にも、同様にして応用することができる解釈するであろう。 20

【0053】

本発明のプローブの重要な特徴としては、目的となる標的DNA領域の少なくとも一部に相補的で、かつ標的領域がその一部であるゲノムの繰り返し配列に相補的な配列が全くないうち、"シングルコピー"または"特有の"DNA配列を有していることである。したがって、シングルコピーまたは特有の配列から作製された本発明のプローブは、対応するゲノム内の一つしかない配列と相補的である。 30

【0054】

ごく最近になって、ヒトゲノムには、互いが対立遺伝子ではない2倍化したデュープリコンや、ゲノム中に3つのコピーが存在するトリプリコン呼ばれる高度に類似した領域があることが明らかとなった。そのような配列を含んだ染色体領域のデュープリコンやトリプリコンは、ヒト以外の靈長類の中には、これらの配列を複数保有していないものもいることや、遺伝子重複により生じた配列のそれぞれの間で、非常に相同性が高いことから、近年の進化で発生したものである。コピー数の少ないデュープリコン(またはトリプリコン)は、ゲノム全体にわたり散在したり、同一の染色体の中で連続して何百、何千も重複している従来の繰り返し配列ファミリーとは区別される；それ故、デュープリコンやトリプリコン由来のプローブは、"シングルコピー"プローブの範疇にあるとみなされ、本発明の目的の範囲内である。 40

【0055】

このようなデュープリコンやトリプリコンは、非常に近年になって進化したため、ゲノム領域全体の配列や構成は(-複雑でほぼシングルコピー部分や、既知の繰り返し配列ファミリーと近縁なものを含む-)、それぞれのデュープリコンやトリプリコンで完全に保存されている。数kbからmbの長さを有するデュープリコンやトリプリコンが、報告されている(国際ゲノム配列機構、Nature, 409: 860-922(2001))。デュープリコン/トリプリコンは、しばしば縦列して並んでおり、ほとんど常に同一の 50

染色体上に存在し、ゲノム中で塊を形成している。プローブ配列を分断している間隔は、デュープリコン／トリプリコン領域の大きさ、相互と関連したデュープリコン（またはトリプリコン）の方向（即ち、順行または逆行）、およびどんな長さにせよ、デュープリコンまたはトリプリコンを分断しているそれ以外の配列の長さに規定される。

【0056】

本発明の中で、核酸配列と関連して“シングルコピー”という用語は、厳密には独特な配列をさす（即ち、対応するゲノム中の1つおよびたった1つの配列と相補的である）が、デュープリコンやトリプリコンもこれに含まれることにする。後述の通り、好ましい様における“シングルコピー”プローブは、ゲノムの3つまたはそれよりも少ない部位にハイブリダイゼーションするものである。10

【0057】

本発明で言うところの“繰り返し配列”とは、標的DNAがその一部であるゲノムに何度も現れる配列のことであり、それぞれの配列の間で少なくとも約60%の相同性がある、より好ましくは、少なくとも約80%の相同性がある配列のことである。この範囲であれば、繰り返し配列は十分な長さをもち、標的DNAに対する、プローブの望ましいハイブリダイゼーションを阻害するような性質をもつ（すなわち、プローブは繰り返し配列の複数のコピーとハイブリダイゼーションするということである）。概して、繰り返し配列はゲノム中に少なくとも約10回（好ましくは少なくとも約50回、より好ましくは少なくとも約200回）現れ、少なくとも約50ヌクレオチドの長さを有し、好ましくは少なくとも約100ヌクレオチドの長さがあることである。繰り返し配列には多様性（例えば、縦列、散在性、パリンドローム、または繰り返し配列を、標的核酸中のいくつかのコピーやゲノム中の何処かと、共有している状態）があり、染色体のセントロメアの近くにあったり、1つの染色体の全体、あるいは2つ以上または全染色体にわたり分散している。通常、いくつかの例外を除き、繰り返し配列は、生理的に機能するタンパク質を発現しない。20

【0058】

繰り返し配列は、2倍体ゲノムに複数存在するコピー中で発生する。コピー数は、少なくとも約10～数万単位までおよび、その中でも繰り返しDNAのAluファミリーは、後者の数万コピーに及ぶ多様性のあるファミリーの典型例である。繰り返しのあるコピーは、塊状になっていたり、ゲノム全体に渡り散在している。繰り返し配列は、ゲノム中の1つまたは2つ以上の部位で塊状になっているが、例えば、各染色体のセントロメアの近くに生じた繰り返し配列や様々な数を有する縦列配列（VNTRs）が挙げられる（Nakamura et al., Science, 235: 1616 (1987)）；繰り返し配列は、1つの染色体全体にわたり分散していることもあり、例えば、Bardoni et al. (Cytogenet. Cell Genet., 46: 575 (1987))で述べられているように、X染色体上にだけ見られる繰り返し配列もある；例えば、Aluファミリーの繰り返し配列のように、全ての染色体にわたり分散している繰り返し配列もある。30

【0059】

それ程複雑ではない繰り返し配列は、遺伝子の中に見られるが、遺伝子をコードしていない配列にも共通して見られる。そのような繰り返し部分は、縦列単位に位置するモノ、ジ、トリ、テトラまたはペンタヌクレオチドのコア配列からなる。このような繰り返し配列を有する縦列要素の数は、異なる個体のゲノムの同一部位でも多様性がある。こうした繰り返しは、ゲノム配列中のコア配列要素の連続した並びを検索することで、見つけることができる。40

【0060】

本発明で言うところの“配列の相同性”とは、2つまたはそれ以上のポリヌクレオチド配列、すなわち、対照となる配列とその配列と比較した目的の配列との関係に関するものである。配列の相同性は、目的の配列と対照となる配列を、最も配列間の類似の程度が高くなるように至適に調整した後で、両者を比較して決定される、つまり両配列間の一致の程50

度によって決めることができる。そうした並べ替えにあたっては、個々の位置を基にして配列の相同性を確かめることができる、例えば、ある特定の位置でヌクレオチドが同一であれば、配列はその位置で“相同的”である。そして、そのような位置相同性の合計数を、対照配列中のヌクレオチドまたは部分配列の合計数で割り算して、配列相同性(%)を求めることができる。配列相同性は、「Computational Molecular Biology」, Lesk A.N.著, ED., Oxford University Press, New York (1988); 「Biocomputing: Informatics and Genome Project」, Smith D.W.著, ed., Academic Press, New York (1993); 「Computer Analysis of Sequence Data, Part III」, Griffin A.M., Griffin H.G.著, eds., Humana Press, New Jersey (1994); 「Sequence Analysis in Molecular Biology」, von Heinge G.著, Academic Press (1997); 「Sequence Analysis Primer」, Gribskov M. and Devereux J.著, eds., M. Stockton Press, New York (1991); およびCarillo H., Lipman D., SIAM J. Applied Math., 48:1073 (1988) に記載されているものを含め、公知の方法により迅速に計算することができるが、これらの方法に限定されるものではない。配列の相同性を決定する、より好ましい方法としては、検索した配列間で最も高い一致性を与えるような方法である。配列相同性を決める方法としては、特定の配列間の相同性を決定してくれる、上市されたコンピュータープログラムの類である。そのようなプログラムの一例としては、GCGRam package (Devereux et al., Nuc. Ac. Res., 12(1):387 (1984))、BLASTP, BLASTNおよびFASTA (Altschul et al., J. Molec. Biol., 215:403-410 (1990))。BLASTXプログラムは、NCBIやその他(BLAST Manual, Altschul et al., NCBI, NLM, NIH, Bethesda, MD 20894; Altschul et al., J. Molec. Biol., 215:403-410 (1990))から提供されているものも、広く利用することができる。これらのプログラムは、特定の配列と対照配列との配列相同性を、最も高く算出できるように、デフォルトのギャップ荷重を使用して、至適に両配列を並べ替えることができる。例えば、対照配列との“相同性”が、少なくとも95%ある、ヌクレオチドを有するポリヌクレオチドに例示されるように、目的のヌクレオチドは、対照のヌクレオチドの100塩基につき、多くとも5塩基しか違いがない点を除いては、対照の配列と同一である。言い換えれば、対照配列に対して少なくとも95%相同性があるヌクレオチドを有するポリヌクレオチドにおいては、対照配列のヌクレオチドの5%までは欠失しているか、他の塩基に置き換わっている、その他、対照配列の全塩基の5%までは、ヌクレオチドの挿入が起こっている可能性もある。どちらかの配列の逆位は、相同的な試験配列のアンチセンスに対する、対照配列の類似性に基づき、コンピュータープログラムで検出することができる。このような対照配列の多様性は、対照となるヌクレオチド配列の3'または5'末端や、その末端で挟まれた配列、また対照配列のヌクレオチドで独立して、または参照配列内的一群または2つ以上の連続したグループ内に散在しているどんな部分にでも起こる可能性がある。

【0061】

本発明のシングルコピープローブは、少なくとも約50ヌクレオチドの長さがあることが好ましい、より好ましくは、少なくとも約100ヌクレオチドである。この範囲であれば、本発明のプローブは、ササンプロッティングを行うのに十分であるからである。しかし、FISHのような他の分析も行うのであれば、プローブはもう少し長いほうがよい、すなわち、少なくとも500ヌクレオチド、より好ましくは、少なくとも約2000ヌクレオチドの長さである。このようなプローブであれば、欠失、重複、挿入、付加、逆位また

10

20

30

40

50

は転座といった、事実上、どんな種類の染色体再編成を検出するにも使用することができる。

【0062】

本発明に従ってプローブを設計するためには、標的DNA領域の配列が、公知である必要がある。標的領域は、染色体全体である場合も、その部分で再編成が同定された一部分である場合もある。こうした配列の知見があれば、標的領域内にある、シングルコピーまたは特有の配列の境界領域を決定する目的に用いることができる。このような結果は、好ましくは標的領域内の繰り返し配列がある部位から推論することにより得ることができる。通常、コンピューターソフトウェアを利用して、標的領域の配列を、対応するゲノム由来の既知の繰り返し配列と比較する。標的領域内に繰り返し配列が同定されると、その間に介在する配列（すなわち、前後にある繰り返し配列の間に存在する配列）は、シングルコピーであると推測される。10

【0063】

標的配列と繰り返し配列とを比較するための最適な並べ替えは、Smith et al. ., *Adv. Appl. Math.*, 2: 482 (1981) の局所的な相同性検索法、またはNeedleman et al., *J. Mol. Biol.*, 48: 443 (1970) の相同的並べ替え検索法により行うことができる。人間の目による検索手法 (Pearson et al., *Proc. Natl. Acad. Sci.*, 85: 244 (1988); Altshule et al., *J. Molec. Biol.*, 215: 403 - 410 (1990)) から得られた結果は、一般的にSmith et al. ., (1981) やNeedleman et al., (1970) の手法ほど、総括的なものではない。20

【0064】

一度シングルコピー配列の情報が得られると、シングルコピー配列のいくつか（普通は最も長いもの）が、ハイブリダイゼーションプローブを作製するために使用される。この点で、本発明のプローブは、Britten et al., *Methods of Enzymol.*, 29: 363 (1974) で定義されたり、Cantor et al., *Biophysical Chemistry: Part III: The Behavior of Biological Macromolecules*, pp. 1228 - 1230 でも説明されているように、様々な“複雑性”があるのかもしれない。選択したプローブの複雑性は、応用するための（プローブの）設計方法に依存する。概して、標的範囲が大きくなればなるほど、プローブの複雑性も大きくなる。ハイブリダイゼーションの感度が高くなるにつれて、一連の配列を検出するために必要とされる、プローブの複雑性は小さくなる。高感度でかつバックグラウンドが低いほど、より小さくて複雑性の小さいプローブを使用できる。30

【0065】

従来のハイブリダイゼーションの技術を使用して、本発明に従った比較的小さいプローブで、再現性があり、容易に検出できるシグナルを得ることができる。FISHを用い、2 kbオーダーのプローブを使用して、簡便に検出できるシグナルを得た。本発明のプローブは、相同意的なシングルコピー配列であるので、本発明の手法の感度は、先行技術（米国特許5746696号明細書）と比較して、向上している。しかし、それぞれが非繰り返し配列を含んでいる、比較的小さい増幅部分でも、*in situ*ハイブリダイゼーションの正しいシグナルを得るためのプローブとして、組み合わせて使用することも可能である。2ヶ所または3ヶ所存在する標的にハイブリダイゼーションする、組み合わせて使用したシングルコピープローブは、ハイブリダイゼーションのシグナルを強くすることもある。40

【0066】

複数の断片化プローブを使用する応用方法の一例としては、異なる染色体間の転座を検出することにある。プローブの複雑性が増加するにつれて、ゲノムの複数のまとまった領域を、同時に解析することも可能である。1つの染色体に関して、融合または転座のある染50

色体は、1つの標識で検出され、両方の標識を有する変異のない正常な染色体と区別することができる。切断点の一方を標的としたプローブの位置は、切断点のもう一方を標的としたそれと、互いに別々に標識し、検出することができる。

【0067】

本発明では、従来のプローブよりもゲノム密度の高い、シングルコピープローブを作製するが可能である。第21および22番染色体全体がシークエンスされ、シングルコピーである配列が、これらの染色体上に集中する傾向があることが明らかとなった。例えば、第22番染色体では、シングルコピー配列の39%が、たった500～1000bpの配列で分断されている。また、2.3kb以上のシングルコピー配列は、平均して、第21番染色体では29.2kb、第22番染色体では22.3kbの配列で分断されている。10

【0068】

シングルコピープローブを作製するのに必要なゲノム配列の大きさを推定するために、染色体21qおよび22q上の、重複のあり一定の長さをもつ、少なくとも1つのシングルコピー配列を検出できることが明らかとなった。2.0kb以上の長さをもつシングルコピー部分は、これらの染色体の100kbのゲノム配列の大部分（染色体22qでは96%、染色体21qでは88%の領域）に存在する。それ故、シングルコピープローブを、第21および22番染色体の全体に包括的に適用するために、多かれ少なかれ2kb（少なくとも100～150kbに1つ）のシングルコピー配列である方がよい。シングルコピー配列が他の染色体でも同様に分布していると仮定すれば、in situハイブリダイゼーションで、臨床的に非常に重要な染色体の再編成を解析するプローブを開発することが可能である。20

【0069】

一旦、目的の染色体領域中に適切なシングルコピー配列が同定されれば、適切なDNAを増幅してプローブを得るために、PCRを利用するのが好ましい。PCRは、幾何級数的に特異的なDNA断片を増幅する非常に有名な技術で、増幅されるDNA断片の5'末端とその相補鎖の3'末端と相同性のある1対のオリゴヌクレオチドのプライマーを基にして、DNAポリメラーゼの伸張反応の繰り返しを原理とする。

【0070】

PCRの錆型として用いる核酸（例えば、DNA）は、一本鎖、または二本鎖であるが、DNAが一本鎖である時には、二本鎖に変換するのが一般的である。錆型DNAの長さは、50bp位短いこともあるが、普通は少なくとも約100bp、もっと一般的には少なくとも約150bpの長さがあり、10000bpやそれ以上のこともあるが、多くの場合、50000bpを超えることはなく、20000bpを超えることもまれである。DNAは可溶で、一端または両末端が錆型ではないDNAになっていて、DNAがプライマー伸張反応に関与できる条件でのみ、プラスミドやそのようなクローニングベクター中に存在している。標的染色体DNA配列または固定されたDNAに相補的である限り、錆型DNAは、様々な異なる起源に由来することもある。他の試薬と組み合わせた場合、錆型DNAの量は、約1分子から1pmolまでのことが多く、たいていは約50分子から0.1pmolまで、より多くの場合0.01pmolから100fmolまでである。錆型の核酸と接触しているオリゴヌクレオチドプライマーは、アニーリングする条件下では、相補的な錆型DNAとハイブリダイゼーションするのに十分な長さがあるが、伸張反応の条件では、錆型DNAと安定した二重鎖を形成するのには、十分な長さではない。プライマーは、一般的に少なくとも約10ヌクレオチド(nt)の長さがあり、普通は少なくとも15nt、より多くの場合、少なくとも約16ntの長さがあり、30nt前後またはそれ以上の長さがあることもある、その中でも、プライマーの長さは、一般的に18～50ntの長さであり、普通は約20～35ntの長さである。（米国特許5436149号明細書で記載されているように）30～35ntのプライマーおよび正確性の高いDNAポリメラーゼを使用することで、比較的長い増幅産物の収率を上げることができる。3040

【0071】

in situハイブリダイゼーションで得られるシグナルの強さを最大化するために、50

増幅の段階で、それぞれシングルコピー個配列のほぼ全長のDNA断片を作り出す1対のプライマーが好ましい。それ故、連続した、または短い介在配列のある（ソフトウェアは、シングルコピー部分の70%以下の長さで分断されたプライマーペアを排除する）プライマーペアは、一般的に、in situハイブリダイゼーションのプローブを作製することは考えられていない。細胞遺伝学的な調製方法の例外に関して、一般的に、この条件は固定化、クローニングまたは合成核酸である標的にハイブリダイズするプローブには当てはまらない。なぜなら、比較的短いプローブのシグナルの強さは、普通、標的分子の数が増加するに従い、強くなるからである。

【0072】

しかし、PCR反応での収率や反応速度を最適化するために、好ましいプライマーの配列は、この他の条件も満たした方が良い。まず、プライマーの配列は、実質的にプライマー中で相補的であったり、もう1つのプライマーと相補的であってはならない。特に、プライマーの3'末端と同じプライマーまたはもう一つのプライマー内のほかの配列が関与する安定したヘテロな二重鎖（6塩基対以上が関与するもの）の形成につながる潜在的なプライマー配列は排除した方がよい。さらに、一対のプライマーの1つのTmは、もう一方のプライマーとの差が、2以内であったほうがよい。それにより、ほぼ同時に変性したり、アニーリングしたりするのが可能となる。全ての好ましい条件を満たすプライマー配列を同定する技術分野において、有名なコンピューターソフトウェアもある（例えば：http://www-genome.wi.mit.edu/ftp/pub/software/primer.0.5またはhttp://www.oligo.net/Oligo_6_tour.htm等である）。

【0073】

PCR反応溶液は、通常さらに一価のイオン群、二価の陽イオン群および緩衝液を含むことができる。汎用されている一価のイオンとしては、塩化カリウム、酢酸カリウム、酢酸アンモニウム、グルタミン酸カリウム、塩化アンモニウム、硫酸アンモニウム等が挙げられ、溶液中に存在する一価のイオンの量の典型例としては、導電性能を発揮する約500~20000、普通は約1000~10000、より多くの場合は3000~6000μMの範囲の十分な量があればよい。二価の陽イオンとしては、マグネシウムイオン、マンガンイオン、亜鉛イオン等が挙げられる。典型的な陽イオンとしては、マグネシウムイオンである。汎用されるマグネシウムイオンとしては、塩化マグネシウム、酢酸マグネシウム等である。緩衝液中に存在するマグネシウムイオンの濃度としては、0.5~1.0mMの範囲でも可能だが、好ましくは約2~4mM、より好ましくは、約2.25~2.75mM、理想的には2.45mMである。緩衝液中に入っている可能性のある代表的な緩衝液の組成または塩としては、Tris、TRICINE、HEPES、MOPS等が挙げられ、緩衝液の組成物の濃度は、典型的には、約5~150mM、普通は約10~100mM、より多くの場合は約20~50mMの範囲である。そのような好ましい態様の中でも、緩衝液の各成分は、pHが約6.0~9.5の範囲になるのに、十分な量である、非常に好ましくは、7.2でpH7.3である。緩衝液中に存在している可能性のある他の試薬としては、EDTA、EGTA等のようなキレート剤が含まれる。

【0074】

PCR反応溶液中にも存在し、融解温度を下げる成分もある、すなわち、DNAの融解温度を低下させる試薬である。融解温度を下げるのにふさわしい成分としては、2つのヌクレオチドの水素結合での相互作用を、阻害するような成分である。その中でも代表的な塩基対不安定化成分としては、ビタミン、フォルムアミド、尿素、硫化尿素、アセトアミド、メチルウレア、グリシンアミド等を含む、そのような中でも、ビタミンが好ましい。融解温度を下げる成分は、典型的には約20~500mMの範囲の濃度にあり、普通は約50~200mM、より多くの場合は約80~150mMの範囲である。

【0075】

PCRの反応溶液を調製する時には、様々な組成の成分は、都合のよい濃度で組合わせてよい。例えば、緩衝液には、プライマー、DNAポリメラーゼおよび錆型DNAが入っ

10

20

30

40

50

ていてもよいし、またはさまざまな組成の成分のすべてを、同時に投入して、反応溶液にしてもよい。

【0076】

PCRの反応溶液の調製の後に、反応溶液は多様な反応サイクルを適用される。それぞれの反応サイクルは、：（1）変性段階、（2）アニーリング段階、および（3）ポリメラーゼ反応段階を含んでいる。反応サイクルの回数は、行う反応に合わせて、変化させることができるが、普通は少なくとも15サイクル、より多くの場合は少なくとも20サイクルであり、60回前後やそれ以上のこともある。そのような様々な反応回数の中でも、典型的には約20～40回の範囲である。約20回以上、普通約30回以上の方法については、酵素反応によるプライマー伸長に適した条件が維持されるように、反応溶液中にDNAポリメラーゼを追加するのが便利、または好ましいこともある。10

【0077】

変性段階は、反応溶液を上昇した温度まで加熱し、反応溶液中に存在している、すべての二重鎖またはハイブリダイズした核酸が解離するのに十分な時間、その温度で維持する過程を含んでいる。変性するに当たって、反応溶液の温度を、約85～100、普通は約90～98、より多くの場合は約93～96の範囲に上げて、その温度に約3～120秒、普通は約5～30秒間維持する。

【0078】

変性に続いて、PCR反応溶液は、溶液中に存在している錆型DNAとプライマーアニーリングするのに、十分な条件下に置かれる。このような条件になるように、反応溶液の温度は下げられ、至適な効率や特異性が達成されるような温度が、多くの場合選択される、そして一般的には、約50～75、普通は55～70、より多くの場合は60～68の温度範囲である。アニーリングの条件は、約15秒～30分、普通は約30秒～5分の範囲の時間に保持される。20

【0079】

プライマーのDNAとのアニーリングの後、またはその間に、反応溶液は、プライマーは、錆型としてハイブリダイゼーションしたDNAを利用して、5'から3'方向に伸長されるような形態で、プライマー末端のヌクレオチドのポリメラーゼ反応を行うのに、十分な条件にある、すなわち、プライマー伸長反応産物の酵素的な产生に、十分な条件下にある。ポリメラーゼ反応の条件を作り出すのに、反応溶液の温度は、典型的には、約65～75、普通は約67～73の温度まで上げられ、約15秒～20分間、普通は約30秒～5分間の範囲で保持される。30

【0080】

前記変性、アニーリングおよびポリメラーゼ反応のサイクルは、一般にはサーマルサイクラーとして知られている、自動化した機械を使用して行われる。使用することができるサーマルサイクラーとしては、米国特許5612243；5602756；5538871および5475610号明細書に記載されたものが挙げられる。

【0081】

上記の全ての条件に基づいて、一連のプライマーが作製され、正常な個体のゲノムDNAを使用してPCRにより検証される。適切なプライマーに関する知見は、必然的に本発明に対応したPCR産生プローブの範囲を明確にする。このように、配列表の配列番号429で始まり、配列番号429～446および480～613として同定された、後述の一対の配列は、特異的で有用なプローブを調製するのに作製された、それぞれforwardとreverseのPCRプライマーである。それ故、ある有用なプローブは、配列番号429および430を組み合わせることにより調製することができ、その他のプローブは、その配列番号の続きにある一対によって、確定される。包括的に言うと、本発明の特定の好ましいプローブとは、前記に詳述した、連続したプライマー配列の一対によって規定されたプローブに対して、少なくとも80%の相同性があるもので、好ましくは少なくとも約90%の相同性があるものである。40

【0082】

10

20

30

40

50

P C R のほかに、特有の配列に対応する D N A 断片は、deletion mutagenesis、制限酵素反応、核酸合成および D N A ライゲーションを含め、他の様々な方法によって得ることができるが、これらの方法に限定されない。

【 0 0 8 3 】

仮に、ゲノム D N A 断片が、繰り返し配列を含む D N A から増幅反応または精製により得られたら、その断片は、標識付加やハイブリダイゼーションの前に精製しておく必要がある。同一の大きさをもつ D N A 断片の精製は、電気泳動、高速液体クロマトグラフィーを含め、様々な方法により得ることができるが、これらの方法に限定されるものではない。好ましい方法において、増幅された断片は、トリス酢酸バッファーを使用して、シーケン L E アガロースゲルの中で、ゲル電気泳動により分子量に従って分離される (Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual [Cold Spring Harbor Laboratory Press, 1989])、次にエチジウムプロマイドまたはサイバーグリーンのような色素で染色し、紫外線 (300 nm) で可視化し、メスを使ってゲルから切り出す。そして D N A 断片は、Microcon 100 カラム (Millipore, Watertown, MA) を使用し、遠心によりゲル断片から回収される。10

【 0 0 8 4 】

増幅された D N A のフェノール・クロロフォルム抽出は、精製するのに適切な方法ではない。この方法を行ってはみたが、この精製法では、全ての染色体の、その全長にわたり非特異的なハイブリダイゼーションがみられ、それは繰り返し配列のハイブリダイゼーションにより観察される像と一致している (図、省略)。なぜなら、もし、一番最初の鋳型が、ゲノム D N A 配列を含んでいるのであれば、P C R 反応の間、D N A ポリメラーゼは、もう一方のプライマーを通り越して、その後ろにある繰り返し配列に向かって、複製された鎖を伸長していくからである。増幅される産物よりも長い、このような伸長産物は、そのような P C R 反応中、存在している。本発明の方法では、繰り返し配列が増幅される領域の後ろにあるので、伸長産物はそうした配列を含んでしまうと考えられる。P C R 反応溶液のフェノール・クロロフォルム抽出では、このような伸長産物を除去できない。こうしたシーケンスを含んでいる P C R 反応溶液は、標的配列のほかに、繰り返しがノム D N A とハイブリダイゼーションしてしまうかもしれない。それ故、精製したゲノム増幅断片の単離は、(ゲノム D N A から直接に得られたり、P C R により得るにせよ) 本発明いに付随した好ましい態様であり、この分野の技術に熟知した人が容易に想像するものではない。20

【 0 0 8 5 】

精製した断片を、大腸菌、酵母、他の種の中で増殖できるプラスミド、バクテリオファージ、または人工的な染色体クローニングベクターへ挿入しておくと、シングルコピー D N A プローブを、繰り返し調製するのに必要なコスト労力を減らすことも可能である。様々なクローニングベクターは、P C R 産物を含んでいるベクターへの迅速なライゲーションや選択に、最適化されてきた (例えば、米国特許 5 4 8 7 9 9 3 および 5 7 6 6 8 9 1 号明細書参照)。もし、プローブが様々なハイブリダイゼーションに使用されるのであれば、クローン化されたリコンビナントは、同一のゲノムである鋳型 D N A から、何度も P C R で増幅するよりも、多量な産物を生み出すのには、安価である。さらに、クローン化されたプローブ中にゲノムを挿入しておけば、精製の過程で単離する必要がない。何故なら、ベクターと結合した D N A 断片は、繰り返し配列を含んでいる可能性のある、他のどのようなゲノム D N A もない状態で、増殖するからである。最後に、クローン化されたベクターは、無尽蔵なプローブ源を提供するのに等しい、しかし、ネイティブなゲノム D N A を鋳型とすると、培養細胞または他のものから、単離し直さなければならないかもしれない。前記の通り、P C R 増幅で得られたシングルコピー D N A 断片は、ゲル電気泳動により分子量に従って単離したり、子の技術分野で既知のようにカラムで精製される。30

【 0 0 8 6 】

その後で、これらの断片は、蛍光体、酵素との共有結合、またはビオチン、およびアビヂ40

ン、ストレプトアビдин、および特異的な抗体によって認識されるそれ以外のものからなる群から選択される1つのような、非放射性の同定ラベルで標識される。非放射性の同定ラベルには、いくつかのタイプがある。1つは、化学的にプローブと結合したラベルであり、プローブを同定したり、局在化を調べる直接的な手段となる。こうしたタイプの例としては、それに適した波長の光を照射すると、高エネルギー状態に励起され、蛍光を放射するような蛍光色素の類である。プローブは、化学的に合成することができ、またはdATPまたはdUTPのようなヌクレオチドに共有結合し、同定選択ラベルを含有する反応試薬を使用した従来からの方法の中で、ニックトランスレーション(Rigby et al., J. Mol. Biol., Vol 113: 237-251, (1977))またはクレノウ標識(Feinberg et al., Anal. Biochem., Vol 137: 266-267, (1984))のような方法を用いて調製するのが好ましい(しかし、このような方法に限定されるものではない)。DNA断片は、蛍光タグをもつヌクレオチドで直接標識したり、後に詳述するように、DNA断片の中に組込まれた修飾核酸を認識する、蛍光標識された抗体と結合させて、二重標識して間接的にラベルすることができる。ニックトランスレーション(100 μl反応スケール)では、エンドヌクレアーゼを含有していないDNAポリメラーゼI(Roche Molecular Biochemicals, Indianapolis, IN)およびDNaseI(Thermo Fisher Scientific, Waltham, MA)を使用する。それぞれの断片は、DNAポリメラーゼ(20 units/microgram DNA)、DNaseI(10 microgram/100 μl反応溶液)、標識されたヌクレオチド(最終濃度で0.05 mM)およびニックトランスレーション緩衝溶液と一緒にになっている。反応は、15で45分から2時間行い、100~500 bpの範囲で、様々な異なる大きさの標識プローブ断片を得る。

【0087】

一般的に使われている、他のプローブの標識方法や検出方法も、本発明の方法により作製されたシングルコピーDNAプローブに適用できることもある。これらの方は、次のようなものを含んでいる：蛍光色素標識(ハイブリダイゼーションが間違いなく起こったという確認として用いられ、そしてさらに、いくらか距離があるというよりも、ハイブリダイゼーションの部位を正確に検出できる個々の染色分体上の標識を決定する)、適切な反応物と組み合わされた時、識別できるような変化を起こす化学薬品(例えば、アルカリフオスファターゼ、からしのペルオキシダーゼおよびガラクトシダーゼが挙げられ、それぞれ反応して、標的配列の存在とその位置を、色の変化で検出できる。)、および(従来からの方法に従い、プローブがビオチンに適切に結合し、そしてプローブに蛍光色素または酵素を繋げて特異性を付与する、アビдинまたはストレプトアビдинが共有結合した蛍光色素または酵素を添加するビオチン-アビдинシステムのように)特異的に結合している物質を間接的に繋がらせる機構。

【0088】

その他の同定ラベルが、上述のプローブと共に使用される可能性があることは、当然のことである。同定ラベルは、エネルギー転移基、複合タンパク質、抗体・抗原、または放射性同位体のような、蛍光性混合物を含んでいる。

【0089】

染色体へのハイブリダイゼーションや検出は、本発明により作製されたDNAプローブの好ましい使用法である。本発明の方法により作製されたDNAプローブは、(in situハイブリダイゼーションで)細胞中の相補的な核酸や、基質上に固定された核酸と直接的にハイブリダイゼーションすることもある。本発明の好ましい使用法としては、in situハイブリダイゼーションであり、それはこの分野では公知で、米国特許5985549; 5447841; 5756696; 5869237号明細書に記載されている。GalilおよびPardueの以前の研究(Proc. Natl. Acad. Sci., Vol 63: 378-383, 1969)を基礎にして、放射性物質を使用したin situハイブリダイゼーションが、1970年代に確立され(例示として、Ge

Richard et al., Proc. Natl. Acad. Sci., Vol 78 : 3755 - 3759 1981 および Harper et al., Proc. Natl. Acad. Sci., Vol 78 : 4458 - 4460, 1981 参照)、それに続いて、放射性物質を使用しない *in situ* ハイブリダイゼーションの技術が確立された。放射性標識を使わない *in situ* ハイブリダイゼーションの技術は改良され、「Current Protocol in Human Genetics」, Vol. 1, Unit 4.3, (Green-Wiley, New York, 1994) や米国特許 5985549 号明細書の中で、Knoll および Lichter により、染色体ハイブリダイゼーションに用いられる実験手法が、提供されている。これらの技術は、二本鎖の核酸種の形成に依存し、その片方の鎖は標識されたプローブに由来し、もう一方の鎖は検出される標的を含有している。標的分子は、染色体または細胞由来の核酸を含有していてもよい。本発明のプローブを標識して、二本鎖を視覚化するために、非常に多くの方法が開発された。

【0090】

本発明の方法は、繰り返し配列を含む、どんな核酸標的にも使用されることが予定されている。標的核酸配列を含んでいる試料は、細胞核、形態的に普通の細胞（または組織）、染色体、その他の細胞成分、または人工的に合成した核酸から調製することができる。そのような試料は、哺乳動物、好ましくは、生体組織検査や術後、または移植患者のいずれかで、病気または異常に苦しんでいる疑いのある人の体液や組織から得ることができる。

【0091】

一例として、染色体のプレパラートは、次のような方法で作られる：フィトヘマグルチニン刺激した末梢リンパ球を、10% FCS を含んだ RPMI 1640 培養液で、37 で 72 時間培養する。その後に、（終濃度で 100 μg / 10 ml になるように）エチジウムプロマイドを、ハーベストの 1.5 時間に添加する。エチジウムプロマイドを加えてから、40 分後にコルセミド（1 μg / 10 ml、終濃度）を添加して培養する。それから、細胞を遠心してペレットの状態にし、37 で約 20 分間、0.075 M の塩化カリウム濃度の中で培養するのが好ましい。遠心して、再び細胞をペレットの状態にし、公知の細胞遺伝学的な技法を用いて、カルノー固定液（メタノールと酢酸を 3 : 1 の容量比で混合したもの）を、3 回変えることで固定する。[末梢赤球からの染色体の調製法については、「Current Protocol in Human Genetics」, Vol. 1, Unit 4.3, (Green-Wiley, New York, 1994) の Dracopoli et al. の Bang and Donlon の項を参照]。溶解した核または細胞を、染色体の伸長を促すように、湿気のある環境で透明なガラス製のカバースライドや顕微鏡用のスライドガラスの上に滴下する。その後で、カバースライドや顕微鏡用のスライドガラスは、一晩風乾したほうが良く、*in situ* ハイブリダイゼーションに使用されるまで、何年も貯蔵しておくことができる。

【0092】

in situ ハイブリダイゼーションの過程で、染色体を変性する直前に、乾燥して貯蔵してあった染色体のプレパラートは、あらかじめ温めてあった 2 × SSC (組成は Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual [Cold Spring Harbor Laboratory Press, 1989] に載っている) で、37 で 30 分間、前処理をし、その後で、一連のエタノール溶液で（70%，80%，90% および 100% エタノール溶液に、それぞれ 2 分間）、脱水する。

【0093】

標的核酸配列が DNA であるなら、試料中の DNA は、熱処理やアルカリ溶液で変性することができる。[アルカリ変性および Gallo et al., Proc. Natl. Acad. Sci., Vol 63 : 370 - 383 (1969) に関しては、Harper et al., Proc. Natl. Acad. Sci., Vol 78 : 4458 - 4460 (1981) 参照]。DNA 鎮のシェアリング、デグラデーションまた

10

20

30

40

50

は酸化が、最小限で済むように分離され、変性が行われる。

【0094】

現在通用する本発明の好ましい方法では、標識されたシングルコピー プローブを、脱イオン化したホルムアミドに再溶解し、70～75で変性する。鑄型となる染色体は、2×SSC中で70%となるホルムアミドを含む溶液(pH7.0)で変性し、その後、一連のエタノール溶液で(冷えた70%エタノール溶液、室温の80%、90%および100%エタノール溶液で、それぞれ2分間)脱水する。標識したプローブの、その鑄型へのハイブリダイゼーションは、50%ホルムアミド/2×SSC/10%デキストランサルフェイト/BSA[牛血清アルブミン; 1mg/ml終濃度]を含んでいる溶液中で行う。
ハイブリダイゼーションの時間は、使用するプローブの複雑さに依存する。ハイブリダイゼーションの後で、ハイブリダイゼーションしていない過剰のプローブは、洗浄過程で除去される。二本鎖を形成しているものについては、15～30分間の洗浄処理を行う：まず始めに、39～45の50%ホルムアミド/2×SSC溶液、次に、39～45の2×SSC溶液、その後で、室温の1×SSC溶液で15～30分間行う。ハイブリダイゼーションした配列は、関連性のある手段により検出できる。例えば、ジゴキシゲニンドUTPは、ローダミンや蛍光試薬が共有結合した抗体(Roche Molecular Biochemicals, Indianapolis, IN)のように、ジゴキシゲニンに対する抗体によって検出することができるが、このような方法に限定されるものではない。擬陽性に検出された反応物を、様々なSSCやSSC/Trition-X濃度の溶液で洗浄して除去した検出操作の後で、染色体はDAPIのような色素で対比染色され、ハイブリダイゼーションしているプレパラートを、Vectashield(Vecter Laboratories, Burlingame, CA)のような褪色防止溶液で固定する。細胞は、適当なフィルターを使用して蛍光顕微鏡で観察し、電荷結合素子(CCD)で撮影される。

【0095】

本発明で重要なのは、プローブまたは標的DNAは、精製した繰り返しDNAのような非特異的な核酸の競争阻害剤と、前もって反応しておく必要がないということ、また、本発明のプローブは、シングルコピー断片、またはリコンピナントのクローン化プローブが、繰り返し配列を含んでいないということを、実験的に証明する必要がないことである(米国特許5985549; 5447841; 5663319; 5756696号明細書参照)。何故なら、本発明のプローブは、繰り返し配列のないシングルコピーだからである。これにより、ノイズに対するシグナル比が、極めて向上するのである。ノイズに対するシグナル比とは、標識したプローブの非特異的な結合により起こる、バックグラウンドを検出する確率に対し、プローブが標的核酸配列のハイブリダイゼーションのシグナルを正しく検出する確率の比として定義される。

【0096】

本発明のプローブを使用して実施されたハイブリダイゼーション反応自体は、まさしく公知のものである。示唆されたように、ハイブリダイゼーションの2つの典型例は、サザンプロットティングとFISHで、この技術分野に精通した人達には公知である。しかし、本発明のプローブを使用して得られる視覚的な像、指標像と言われるのは、細胞遺伝学的解析、とりわけ分子細胞遺伝学的解析に対して、非常に有効な手段である。このような指標像は、正常および異常な染色体の、顕微鏡および/またはフローサイトメトリックによる同定や、ならびに遺伝子異常の特性決定に有効である。プローブを視覚化する、多様で適用範囲の広い方法が利用でき、異なるプローブ成分の結合像は、例えば、色によって区別することができる。このように、本発明の方法は、1つまたは2つ以上の色(複数色による指標像)および/またはその他の指標方法を用いて視覚化された染色体上に、どんな目的の指標像でも具体的に現すことができる。

【0097】

本発明のプローブを使用して得られる好ましい指標像とは、1つまたは複数の“バンド”を含有しているものであり、ここで言う“バンド”とは、標的DNA配列に結合したプロ

ープを有する、その配列を含有しているゲノム中の参照点を意味し、その結果、形成される二重鎖は、何らかの指標により検出することができる。ハイブリダイゼーションでの洗浄処理や検出条件に依存して、1つのバンドは、単一の染色体領域上で再現性のあるシグナルを与える配列の狭い部分から、単一または複数の染色体上にある複数の領域まで、拡張させることができる。このプローブ由来の指標バンドは、前処理および化学染色により観察されるバンドとは区別される。本発明のプローブによるバンドは、DNA配列との相補性に基づいているが、化学染色により観察されるバンドは、(構造やタンパク質の構成のような)染色体本来の性質に依存する、しかし、そのDNA配列へのハイブリダイゼーションによるバンドではそのようなことはない。さらに、化学染色技術は、メタフェーズの染色体との関連でのみ有用であるが、本発明のプローブによるバンドは、メタフェーズでもインターフェーズのいずれの染色体に対しても有効である。

10

【0098】

後述の実施例では、ゲノム中の標的DNA配列に対してハイブリダイズするように設計された特異的なDNAプローブを、開発、作製、標識および使用するために用いられる好ましい技法を示している。しかし、このような実施例は、写真像により明示され、どの例示も、本発明の全体像の限界としてみなされるものではない。

【0099】

[実施例]

実施例1、HIRA遺伝子プローブの開発

ヒトの第22番染色体上の有名な遺伝子の異常は、染色体のバンド22q11.2にある、HIRA遺伝子の1つの欠損と関連している、すなわち、正常な個体では、HIRA遺伝子は2つ存在しているが、異常がある個体では、1つしか存在していない。この欠損は、DiGeorge and Cardio-Facial-Syndrome (VCFS)のような、半欠損症候群の1つの原因であると考えられている。何故なら、遺伝子産物の量が不十分だと、正常な胎児の発達に支障をきたすからである (Fibison et al., Amer. J. Hum. Genet., 46: 888-95 (1990); Converage et al., Amer. J. Cardiol., 77: 1023-1205 (1996))。ネコ眼症候群を含む他の病気や、融合した第22番染色体に関連した症候群は、この領域由来のゲノム配列が過剰なことに原因がある (Mears et al., Amer. J. Hum. Genet., 55: 134-142 (1994); Kno11 et al., Amer. J. Med. Genet., 55: 221-224 (1995))。このような症候群を持つ個体は、典型的に、第22番染色体の短腕が過剰である。

20

30

30

【0100】

まず最初に、“HIRA”という用語をもとに、国立医学図書館ウェブサイトにあるEntrez Nucleotideソフトウェアを使用して、コンピューターによる検索を実施した。これにより、GenBankの中に、HIRA遺伝子の一連のcDNA配列が同定された。3859bpの長さを有するcDNAの全長 (GenBank Accession No. X81844)を選択した。次に、このcDNA配列を、国立医学図書館のDRAFT配列 (<http://www.ncbi.nlm.nih.gov/Genome/seq/page.cgi?F=HsBlast.html&&ORG=Hs>)を含んでいるゲノム配列と比較した。これは、十分な長さを有するゲノム配列がプローブを作製するのに利用できるかどうかを決定するために行った。このように比較することで、HIRAゲノム配列全体は公知であり、コーディング配列部分は、染色体中で100836bpにわたり広がっていることが明らかとなった。GenBankで利用できる隣接したゲノム配列は、コーディング部分を越えたところまであるので、5'末端の遺伝子プロモーター領域や、3'末端の非翻訳配列やポリA付加シグナル由来の配列を含ませる目的で、コーディング領域よりも、もっと長い部分を選択することも可能である。このようにして、およそ103kbある全ゲノム部分が選択された。第1塩基から103kbにわたる部分は、GenBank Accession No NT_001039の7

40

50

9 8 3 3 4 塩基に対応する。

【0 1 0 1】

次に段階では、選択した 1 0 3 k b ゲノム部分は、対象となるゲノム配列（配列番号の 1 ~ 4 2 8 ）と関連して並べ替えられ、既知の複雑性の高い繰り返し配列ファミリーまたはコンセンサス配列、およびヒトゲノム中に存在していることが知られている、少なくとも 1 7 ヌクレオチドの長さ（モノ、ジ、トリおよびテトラヌクレオチド単位の）の複雑性の小さい繰り返し配列（配列番号 4 4 7 ~ 4 7 9 ）と比較した。この比較には、遺伝子情報研究所のウェブサイト、www.girinst.org に見られる、広く利用されている C E N S O R プログラムを使用した。このプログラムでは、ゲノム部分中にある繰り返し配列の部位や分散状態を決定するために、Smith-Waterman グローバル・アライメント・比較アルゴリズムを利用する。ゲノム配列に関して、Smith-Waterman アライメントを繰り返し行うには、次に示すようなパラメーターを使用した：マージン配列の長さ：5 0 ヌクレオチド、挿入断片を抽出する最小の長さ：1 2 ヌクレオチド、適合断片を組み合わせる最小マージン：3 0 、類似性の限度：2 2 、常に一致している類似性の限度：3 5 、割合限度：2 . 8 、相対的な類似性：2 . 8 、ギャップ・コンスタント D 1 : 2 . 9 5 、ギャップ・コンスタント D 2 : 1 . 9 0 、ミスマッチ・ペナルティー：- 1 . 0 。このような解析結果に基づき、ヒト H I R A 遺伝子のコーディング配列の中、およびその近傍にある繰り返し配列ファミリーの詳細を表す、下記の表を作成した。

【0 1 0 2】

【表 1】

第 1 表

HIRA遺伝子の位置(bp)				HIRA遺伝子の位置に 対応している、配列表に おける位置(bp)	
始点	終点	繰り返しファミリー	配列番号	始点	終点
798411	798434	(AC)	452	1	24
798983	799395	MLT2A1	444	1	434
801257	801348	CHESTER_A	420	132	223
801367	801729	LIME_ORF2	425	757	1089
801746	802032	Alu-Jb	2	2	289
802033	802308	LIME_ORF2	425	1090	1380
802355	802434	L1MB6_5	77	1629	1710
802448	802798	L1M3D_5	66	996	1348
802811	803100	Alu-Y	2	1	290
803104	803189	L1M3D_5	66	907	995
803199	803454	Alu-Jb	2	5	290
803472	803545	Alu-Spqxz	2	2	76
803548	804061	L1MEC_5	345	1860	2392
804079	804365	Alu-Sz	2	6	290
804476	804559	L1P_MA2	348	6242	6321
804625	804885	L1ME_ORF2	425	2287	2568
804936	804997	MLTIE2	106	198	260
805011	805077	MLTIE1	105	420	484

【0 1 0 3】

【表 2】

10

20

30

40

始点	終点	繰り返しファミリー	配列番号	始点	終点
805110	805211	L1PBA_5	359	103	204
805212	805862	L1PBA_5	359	1089	1738
805933	805989	Alu-I	2	234	290
805991	806489	L1PBA_5	359	1749	2247
806510	806624	L1	59	1659	1773
806628	806917	Alu-Sz	2	1	290
806919	807254	L1M2_5	61	2377	2716
807301	808176	L1P_MA2	348	3516	4425
808179	808469	Alu-Sz	2	1	290
808476	808734	L1	59	3268	3525
808735	809426	L1ME_ORF2	425	1411	2105
809429	809860	L1P_MA2	348	5607	6044
809861	809993	Alu-Jb	2	2	134
809996	810282	Alu-Jb	2	2	290
810345	811040	L1	59	4711	5402
811041	811221	L1PB3	358	151	333
811226	811513	Alu-Sx	2	1	287
811515	812032	L1PB1	357	330	863
812096	812394	Alu-Jb	2	1	288
812474	812698	Alu-Jb	2	5	229
812721	812836	Alu-Jo	2	2	117
812862	812901	L1P_MA2	348	4315	4354
812903	813078	L1	59	3028	3222
813079	814102	L1ME_ORF2	425	1113	2166
814323	814410	MER1B	315	242	337
814411	814557	CHARLIE3	7	1	281
814760	814916	L1MB7	78	9	143
815061	815181	Alu-Y	2	1	134
815420	815452	LTR67	279	99	131
816487	816772	Alu-Sx	2	5	290
817180	817270	L1MCC_5	335	1384	1473
817332	817620	Alu-Sg	2	1	290
817634	817909	Alu-Sq	2	1	288
817943	818227	Alu-Sx	2	2	289
818368	818578	HALI	18	1346	1547
818631	818791	LINE2	362	2280	2464
818824	818889	Alu-S	2	223	290
818890	819185	LINE2	362	2465	2749
819328	819450	LINE2	362	1925	2049
819565	819757	LINE2	362	2273	2498
823604	823892	Alu-Jo	2	2	290
826836	827042	Alu-Sxzg	2	84	290
827922	827977	MIR	99	105	160
830762	831371	L1MEC_5	345	1498	2123
831396	831685	Alu-Sx	2	2	290
831687	831774	L1MEC_5	345	2117	2205
831778	832066	Alu-Sx	2	1	290
832155	832288	Alu-FLA	2	5	134
832317	832431	L1MC2	79	666	786
832442	832735	Alu-Sz	2	1	289
832742	832992	L1MC2	79	787	1077
833004	833170	L1ME_ORF2	425	172	340
833177	834590	TIGGER1	148	1	1477
834592	834642	Alu-Jb	2	156	207
834799	834877	Alu-Jb	2	208	290
834907	835194	Alu-Y	2	1	289
835198	835590	TIGGER1	148	1468	1900
835597	835888	Alu-Sx	2	1	290
835946	835979	L1P_MA2	348	4654	4689
836060	836177	MER2	316	229	345
836203	836486	Alu-Sx	2	7	290
836497	836712	MER2	316	1	228
838478	838760	Alu-Sz	2	1	288
838822	839069	Alu-Sx	2	1	288
839086	839373	Alu-Sz	2	1	289
840297	840926	L1MB7	78	269	915
841062	841306	L1MB7	78	7	249
841323	841382	L1ME_ORF2	425	3053	3116

10

20

30

40

【 0 1 0 4 】

【 表 3 】

始点	終点	繰り返しファミリー	配列番号	始点	終点
841408	841697	Alu-Sq	2	1	290
841705	841828	Alu-Jo	2	1	136
841829	842012	L1ME_ORF2	425	2870	3052
842744	842871	MER86	239	51	183
842879	843107	Alu-Spqxz	2	3	230
843109	843271	Alu-Jo	2	9	175
847056	847210	MER104	293	1	179
847256	847351	L1ME4	343	128	224
847413	847551	MIR	99	65	218
847570	847695	L1ME4	343	1	127
847865	848137	Alu-Y	2	1	290
848171	848458	Alu-Sg	2	1	290
848493	848564	L1PA7	355	35	105
848646	848928	Alu-Sc	2	5	290
849186	849435	L1ME_ORF2	425	2527	2796
849450	849745	Alu-Sx	2	5	289
850114	850249	L1P_MA2	348	5447	5610
850250	850761	L1	39	3478	4017
850824	850942	L1ME_ORF2	425	1128	1265
851588	851614	(T)	449 (complement)	1	27
851749	851881	L1ME2	341	357	523
852607	852853	L1MA10	72	664	918
852863	853156	Alu-Sc	2	1	290
853176	853211	L1MA10	72	628	663
853212	853267	L1MA9	75	987	1041
853491	853779	Alu-Sz	2	1	290
859137	859435	Alu-Sx	2	1	290
859436	859456	(A)	449	1	21
859570	859805	L1ME3A	342	215	442
859806	860289	L1ME2	341	375	879
860318	860605	Alu-Y	2	1	290
862194	862481	Alu-Sg	2	1	290
865060	865350	Alu-Sq	2	1	290
867521	867800	Alu-Jb	2	1	288
867836	867876	MIR	99	157	196
869546	869802	LINE2	362	123	413
869923	870118	LINE2	362	1251	1450
870124	870202	Alu-J	2	48	132
870203	870296	LINE2	362	1451	1592
870316	870666	LINE2	362	1708	2097
871000	871075	LINE2	362	2617	2736
871650	871935	Alu-Jo	2	1	290
871936	871960	(GAAAAAA)		4	28
872154	872444	Alu-Sc	2	1	289
874867	874990	L1MB7	78	529	676
878120	878408	Alu-Sx	2	1	290
881003	881054	MLT1G	109	217	268
881130	881266	MLT1G	109	269	480
881293	881346	MLT1G	109	415	469
881762	881891	LINE2B	363	85	229
882448	882740	Alu-Sb0	2	1	290
883566	883716	Alu-Sz	2	1	288
883782	883977	Alu-Sc	2	2	290
883988	884329	L1P_MA2	348	5600	5935
884333	884623	Alu-Sp	2	1	290
884624	885134	L1ME_ORF2	425	2431	2975
885160	885456	Alu-Jb	2	9	290
885460	885742	L1ME_ORF2	425	2949	3252
885744	886031	Alu-Sx	2	1	288
886032	886082	Alu-Sp	2	291	341
886083	886166	L1MB7	78	137	220
886168	886454	Alu-Sc	2	1	290
886535	887059	L1MB7	78	345	901
887169	887460	Alu-Y	2	1	289
887485	887748	L1MD2	337	794	1072
887752	887779	LORII	366	395	422
888253	888318	LINE2	362	2440	2505
888385	888548	LINE2	362	2579	2739

【 0 1 0 5 】

【 表 4 】

始点	終点	繰り返しファミリー	配列番号	始点	終点
888865	888893	LORII	366	394	422
889006	889296	Alu-Jb	2	5	290
889446	889548	Alu-Jo	2	188	290
889549	889677	LIPB3	358	770	897
889842	890133	Alu-Sq	2	1	290
890515	890797	Alu-Sz	2	1	283
890858	890972	LIME2	341	769	885
890986	891024	LORII	366	396	434
891028	891063	LTR66	266	173	207
891126	891536	LINE2	362	1980	2452
891545	891670	LTR16A1	382	9	128
891688	891963	LTR16A	381	146	429
892907	893013	LINE2	362	2636	2747
893851	893924	MLT1L	119	47	119
894528	894849	Alu-Sx	2	1	290
895825	895903	LINE2	362	2592	2664
895912	896083	MER20	317	46	216
897067	897299	MER20	317	2	217
897492	897624	Alu-FLA	2	2	136
897977	898261	Alu-Sc	2	1	290

10

20

30

40

50

【0106】

非繰り返し部分の長さを、これらのデータを元に計算した。例えば、5358bpの非繰り返し配列部分は、その前後にあるAlu-SzおよびAlu-Sx繰り返し要素との境界となる、853779番目と859137番目の塩基の間に存在することが決定された。次に、非繰り返し部分は、それぞれの長さに応じて並べ替えられた。これらの非繰り返し部分のうち、4つ、すなわち、前記で詳述した5358bp配列、3847bp配列(819757から823604に対応する)、3785bp配列(843271から847056に対応する)、および3130bp配列(874990から878120に対応する)が、プローブを作製するために選ばれた。

【0107】

次の段階では、同定した4つのシングルコピー部分を増幅するために、long PCRの手法を用いた。5358bpの増幅産物に対して行った手法について、下記に詳細に記述する。同様の手法で、残りの3つのシングルコピー部分の増幅も行った。

【0108】

最大限長いプローブが、FISHの実験には好ましかった。しかし、これらのプローブを作製するPCR反応を至適化するためには、他の制約も満たさなければならなかった。それにより、非繰り返し配列部分全体よりも、いくらか短い増幅された産物が生じた。Primeコンピュータープログラムは、PCRのためのプローブの選択を至適化するために用いられた(Genetics Computer Group Software, Madison WI)。ロングPCRの至適化したPCRプライマーには、次のような制約があった：長さ30～35のヌクレオチド、GC含量が50～80%；融解温度が65～70；プライマーが、8ヌクレオチドより長いヘアピン構造を持つように、3'末端でセルフアニーリングしない；プライマーが、そのどの位置でも14ヌクレオチドよりも大きいセルフアニーリングをしないこと；プライマーが、標的配列のたった一つの位置でしかアニーリングをしないこと；そしてプライマーの3'末端では8bpより短い長さ、プライマーのどの位置でも14bpより短い長さで、プライマー同士のアニーリングが起きないことである。さらに、ロングPCRを至適化するためには、次のような制約が増幅されるPCR産物にあてはまった：長さが5100～5358ヌクレオチドであること；GC含量が40～60%であること；融解温度が70～95であること；forwardプライマーとreverseプライマーの融解温度の差が、2以下であること。この制約から、517のforwardプライマーと382のreverseプライマーができて、その組み合わせにより928通りのPCR産物が可能である。上述の制約順位を用いたPrimeプログラムは、プライマー対となり得る配列を与えてくれた。後述の第2表の中で示されているように、高順位のプライマーが、合成のために選ばれる。このよう

なプライマーは、商業目的で生産された (Oligos, Etc., Wilsonville)。

【0109】

【表5】

第2表

遺伝子	染色体バンド	GenBank Accession No., 染色体ゲノム配列	最も長いシングルコピー部分の座標	Forward PCR プライマーの座標, 始点 / 終点	Reverse PCR プライマーの座標, 始点 / 終点	PCR プライマーの配列番号 Forward/Reverse	プローブの長さ (bp)
HIRA	22q11.2	NT_001039	853779/859137	853946/853975	859116/859085	429/430	5170
HIRA	22q11.2	NT_001039	819757/823604	819901/819933	823592/823559	431/432	3691
HIRA	22q11.2	NT_001039	843271/847056	843602/843631	846946/846915	433/434	3344
HIRA	22q11.2	NT_001039	874990/878120	875226/875257	878074/878042	435/436	2848

【0110】

このようなプライマーを使用して、5170 bp のプローブを增幅するために、(フェノール抽出で精製された) 1 µg の高分子量ゲノムDNAとオリゴヌクレオチドのプライマ

ーの濃度がそれぞれ 200 μM のプライマーとを用いて、ロング P C R 反応 (50 μl スケール) を行った。下記に示す P C R 条件下で、特異性があり、フィデリティーが高い D N A ポリメラーゼ (LA - T a q , 宝酒造株式会社) を使用した。

S t e p 1 - 9 4 - 5 分

S t e p 2 - 9 8 - 2 0 秒

S t e p 3 - 6 5 - 7 分

S t e p 4 - S t e p 2 に戻って、14 サイクル繰り返す

S t e p 5 - 9 8 - 2 0 秒

S t e p 6 - 6 5 - 7 分 + 1 サイクルにつき 15 秒追加

S t e p 7 - S t e p 5 に戻って、14 サイクル繰り返す

S t e p 8 - 7 2 - 1 0 分

S t e p 9 - 0

S t e p 1 0 - 終了

【 0 1 1 1 】

5170 bp のプローブを増幅するのは、それより短いフラグメントを増幅するのに比べて効率が低いので、最初に行った P C R 反応では、多様なハイブリダイゼーション実験に使用できる、十分な量のプローブを得られなかった。そこで、増幅した反応溶液の 4 μl を、次に示す方法によりもう一度増幅させた：S t e p 1 - 9 4 - 1 . 5 分、その後は最初に行った P C R 反応の S t e p 2 ~ 1 0 を行う。十分な量の 5170 bp のプローブが得られた。再増幅反応を行う代わりに、S t e p 7 を少なくとも 10 サイクル増やすこともできる。

【 0 1 1 2 】

増幅した産物は、ゲル電気泳動により精製し、その後カラムクロマトグラフィーを行った。まずはじめに、増幅した産物は、1 × 变成 T A E バッファー中で、0 . 8 % S e a k e m L E アガロースゲル (F C B i o p r o d u c t s) 上に分離された。それからゲルを、エチジウムプロマイドで染色し、UV 光で視覚化した。適切な長さに相当する D N A 断片を切り出して、U l t r a f r e e - D A スピンカラム (M i l l i p o r e) の中に入れ、5000 g で 10 分間遠心する。D N A は、溶液中に回収され、1 / 10 倍容量の酢酸ナトリウムと 2 . 5 倍容量のエタノールを加えて、-20 度の条件下で沈殿させた。次に、沈殿した D N A を遠心し、70 % の冷エタノールで rinses 、風乾後、滅菌した 20 μl の脱イオン水に溶解した。D N A 断片は、濃度を決定するために、0 . 8 % アガロースゲル (S i g m a) で確認した。

【 0 1 1 3 】

プローブの標識方法、ハイブリダイゼーション法、非特異的に結合したプローブの除去方法およびプローブの検出方法については、K n o l l と L i c h t e r により提示されている、D r a c o p o l i e t a l . , (e d s) 、“ C u r r e n t P r o t o c o l s i n H u m a n G e n e e t i c s V o l u m e 1 ” , U n i t 4 . 3 (G r e e n - W i l l y , N e w Y o r k , 1 9 9 4) 。プローブを標識するために、標識としてジゴキシゲニン - 11 - d U T P を使用して、手短に標準的なニックトランスレーションを行った (R i g b y e t a l . , J . M o l . B i o l . , 1 1 3 : 2 3 7 - 2 5 1 , (1 9 7 7)) 。これにより、一連の重複のある 300 ~ 500 bp の標識された D N A 断片が得られ、その中には、5170 bp のプローブも含まれていた。

【 0 1 1 4 】

それから、標識されたプローブ断片は、1 / 10 容量の酢酸ナトリウム、2 . 5 倍容量のエタノールおよびキャリア D N A を加えることにより (一晩、-20 度) 沈殿させた。翌日、沈殿した D N A を遠心し、凍結乾燥し、125 ng / 20 μl の濃度で脱イオン化した滅菌水中に溶解した。

【 0 1 1 5 】

米国特許 5 4 4 7 8 4 1 , 5 6 6 3 3 1 9 および 5 7 5 6 6 9 6 号明細書に記載されてい

10

20

30

40

50

るタイプのプロッキング核酸有無の条件で、標識したプローブ断片を使用し、変性した正常なヒトのメタフェーズ染色体を用いて、ハイブリダイゼーションの対照実験を実施した。溶解した標識プローブの 20 μl を凍結乾燥させ、10 μl の脱イオン化したフォルムアミドに溶解し、1本鎖の核酸にするために 70 ~ 75 ℃ で 5 分間変性した。対照として、プローブは、C₀t1 DNA (Life Technologies) の 10 μg に対し、125 ng (または 20 μl) の標識プローブを加えて、その混合溶液を凍結乾燥させることで、精製した繰り返しDNAと前反応させた。それから、この混合溶液を、70 ℃、5 分間変性させ、その後、プローブ中の一本鎖繰り返し配列を二重鎖核酸にするために、37 ℃、30 分間の前反応 (すなわちプレアニーリング) を行った。これにより、繰り返し配列と標的の鑄型DNAのような染色体とのハイブリダイゼーションができないなくなる。

10

【0116】

その後、精製した繰り返しDNA (例えば、C₀t1) と混合した、または混合していない変性したプローブは、等量のあらかじめ温めておいたハイブリダイゼーション溶液 (4 × SSC / 2 mg / ml ヌクレアーゼフリーの牛血清アルブミン / 20 % デキストラサンサルフェイト / 30 % 滅菌した脱イオン水) と混ぜて、変性した標的DNA上にかけた。顕微鏡用のスライドに固定された染色体の標的DNAは、50 % ホルムアミド / 2 × SSC 中で、72 ℃ 5 分間変性した。カバースライドを、スライドのプローブハイブリダイゼーション混合溶液上において、蒸発を防ぐためにマニキュア液でシールし、一晩 39 ℃ の湿気のある入れ物の中に静置した。

20

【0117】

ハイブリダイゼーションの後に、非特異的に結合したプローブは、塩濃度および温度といった様々な厳しい条件下で、洗い流された。繰り返し配列がハイブリダイゼーションしないように、前反応を行った標識したプローブおよびそのような前反応を行っていないプローブは、従来からあるFISH法 (Knoll および Lichten, Current Protocols in Human Genetics, Vol 1, Unit 4.3 Green-Wyllie, New York, 1994) を用いて、ローダミンラベルされたジゴキシゲニン-11-dUTPに対する抗体で検出された。染色体DNAは、DAPIで対比染色した。そして、マイクロスライド上の細胞プレパラートを、(Vectashield, Vector Laboratories, Burlingame, CAのような)褪色防止溶液で固定し、適当な蛍光フィルターのセットを用いて蛍光顕微鏡を使用して視覚的に観察した。第1図および第2図は、ハイブリダイゼーションを比較して結果を図示した写真であり、第1図はプロッキング繰り返し配列を使用したハイブリダイゼーションで、一方、第2図は精製した繰り返し配列と前反応を行っていないハイブリダイゼーションの結果である。これらの写真是、2つの正常な染色体 22q11.2 領域上の両方のHIRAアレルへのハイブリダイゼーションを示している。これらの写真を比較して、プロッキング繰り返し配列の存在は、本発明のプローブを使用するにあたり、必須ではないことが示されている。

30

【0118】

第2表に特定されている残りの3つのプローブは、上述のようにPCRで増幅され標識された。これらのプローブは、プローブの有効性を決定するために一連のFISH実験の中で使用された。このように4つのプローブのすべてが、潜在的な繰り返し配列とのプレアニーリング無しに使用され (第6図)、短いほうから数えて3つのプローブを組み合わせて、欠失していることがあらかじめ明らかとなっているDiGeorge / VCF5に罹患した患者から採取した細胞に対して使用された (第12図)。第6図の写真の中で、プローブは、正常な個体の両方の第22番染色体の単一領域とハイブリダイゼーションしている (矢印)。第12図では片方の第22番染色体としかハイブリダイゼーションしていない (矢印)。星印で示されているように、もう一方の第22番染色体は、この領域に欠失があり、プローブとはハイブリダイゼーションしない。

40

【0119】

50

実施例 2

N E C D I N および C D C 2 L 1 遺伝子のプローブの開発

実施例 1 で記載された開発技術は、第 1 番染色体上の既知の遺伝子異常（染色体性の 1 p 3 6 . 3 症候群； S l a v o t i n e k et al . , J . M e d . G e n e t . , 3 6 : 6 5 7 - 6 3 (1 9 9 9) ）および第 1 5 番染色体上の既知の遺伝子異常（ブレーダー - ウィリまたはアンゲルマン症候群）を検出するために、一連のプローブの開発に用いられた。ブレーダー - ウィリまたはアンゲルマン症候群の患者のおおよそ 7 0 % の人達は、N E C D I N 遺伝子を含む核酸配列の片方の欠損を示す（ K n o l l et al . , Amer . J . M e d . G e n e t . , 3 2 : 2 8 5 - 2 9 0 (1 9 8 9) ; N i c h o l l s et al . , Amer . J . M e d . G e n e t . , 3 3 : 6 6 - 7 7 (1 9 8 9) ）。この遺伝子が過剰に存在すると、介在性の複製、過剰なファミリー、2 つの動原体がある第 1 5 番染色体を有する患者は、異常な形質を示す（ C h e n g et al . , Amer . J . H u m . G e n e t . , 5 5 : 7 5 3 - 7 5 9 , (1 9 9 4) ; R e p e t t o et al . , Am . J . M e d . G e n e t . , 7 9 : 8 2 - 8 9 , (1 9 9 8) ）。後述の第 3 表には、シングルコピーと導出された部分、P C R のプライマーを設計した部分、配列番号および合成されてくるプローブの長さが表示されている。

【 0 1 2 0 】

【表 6 】

第3表

遺伝子	染色体バンド	GenBank Accession No., 染色体ゲノム配列	最も長いシングルコピー部分の座標	Forward PCR プライマーの座標、始点 / 終点	Reverse PCR プライマーの座標、始点 / 終点	PCR プライマーの配列番号 Forward/Reverse	プローブの長さ(bp)
CDC2L1 ¹	1p36.3	AL031282	8823/17757	9137/9167	13960/13931	44/443	4823
CDC2L1 ¹	1p36.3	AL031282	8823/17757	13028/13057	17752/17720	445/446	4724
NECDIN	15q11-q13	AC006396	94498/99152	94501/94535	98567/98601	439/440	4166
NECDIN	15q11-q13	AC006396	68031/75948	72122/72156	75666/75637	437/438	3544
NECDIN	15q11-q13	AC006396	76249/79221	76608/76639	78898/78867	441/442	2290

¹2つを合わせると全配列部分をおおうような、DNA プローブ断片を作製するために、2組のプライマー一対を使用した。

【 0 1 2 1 】

P C R 増幅反応は、第3表の C D C 2 L 1 プライマーを使用して行われ、增幅産物は、実施例1で述べられたように、標識、ハイブリダイゼーションを行い、検出された。標識されたプローブは、一連の F I S H で使用され、そのハイブリダイゼーションの結果の写真は、第7図～第10図で示されている。第7図で示されている実験では、最も長い 4 8 2 3 b p のプローブが使用され、鑄型中の潜在的なハイブリダイゼーション繰り返し配列は、精製した繰り返し D N A とプレアニ - リングすることにより、不活化されていた。対照

実験としては、同じプローブが、精製した繰り返しDNAとプレアニ-リングすることなしに使用された(第8図)。このハイブリダイゼーションの結果は、繰り返し配列へのハイブリダイゼーションを防ぐために、精製した繰り返しDNAは必要ではないことを示している。両方の実施例において、ハイブリダイゼーションした1つまたは両方の染色分体を有する染色体は、矢印で示されている。第9図および第10図で示されている実験の中では、精製した繰り返しDNAのプレアニ-リングを行う(第9図)、または行わない条件で(第10図)で、4823および4724bpの長さのプローブが使用された。また、精製した繰り返しDNAの前反応は、本発明のプローブの使用にあたり必要ないことが示された。

【0122】

10

第3図～第5図および第11図で例示されているように、NECDINプローブも、一連のFISHの実験で使用された。これらのプローブは、NECDIN遺伝子の36～62kb末端のDNA配列を検出した。3544bpのプローブ(配列番号437～438)は、MAGEL2遺伝子の3'末端を検出した。第3図の中で、3544bpのプローブは、精製した繰り返し配列を用いてプレアニ-リングをし、正常な個体から採取したメタフェーズの細胞に対して使われた。第4図は、プレアニ-リングをしていない対照実験である。第11図では、片方の第15番染色体上の15q11-q13配列の欠失を含んでいることが知られている、ブレーダー-ウィリ症候群の患者のメタフェーズの細胞に対して、3つのプローブのすべてを組み合わせて使用した。正常な相同染色体は、矢印によって示され、片方の染色分体に対するハイブリダイゼーションを示している。欠失のある染色体の部位は、星印で示されている。それは、プローブのハイブリダイゼーションを示したものではない。

20

【0123】

前述の実施例は、DNA断片を組み合わせて使用しても、DNA断片をそれぞれ別々に使用したときと比較して、同一の結果を示すことを表している。これにより、DNA断片のどれも個々に、または組み合わせて使用しても、ゲノムの他のどの部位にもハイブリダイゼーションしない、それ故、繰り返し配列を必要としていることが確立されている。この事実は、シングルコピーのゲノムプローブを設計・作製する本発明の方法の有効性を、さらに確かなものにしている。

30

【0124】

このような異常を検出する商業的かつ研究中のゲノムプローブを使用するには、今までには、メタフェーズまたはインターフェーズの染色体に対するプローブのアニ-リングの前に、繰り返し配列のハイブリダイゼーションを不可能にしておく必要がある。これにより、実験方法を実施するのに必要な操作段階を増加させ、操作ミスが起こる確率を増加させてしまうが、そのいずれも臨床診断を行う研究室には、受け入れがたいものである。第7図で見られる結果は、このような異常を検出する、この分野で商業的に利用されるゲノムプローブを使用して得られる結果と、同等のものである。それ故、これらのプローブは、このような遺伝病を検出するのに有用である。プローブは、単独でも、ハイブリダイゼーションや検出に必要な他の試薬と組み合わせても、こうした遺伝病を検出するキットとして、臨床試験を行う研究室に提供することもできる。

40

【0125】

本明細書の中で引用された実施例で例示されている以外に、ゲノム配列から作製されたプローブは、先天性、突発性、または後天性の染色体の再構成を検出するのにも利用できる。これらの再構成は、上記で例示した以外に、(腫瘍形成を含め)様々な他の公知の遺伝子異常や疾病に対応している可能性がある。それ故、本発明は、商業的に利用されていなかったり、適切なプローブが作製できないゲノム領域からプローブを、作製するのにも有効である。

【0126】

原則として、本発明の作製方法は、DNA配列が利用でき、ゲノム中の広い意味での繰り返し配列部分が分類化されている、どんなゲノム部分に対するシングルコピーゲノムプロ

50

ープでも、設計、開発および作製に利用することが可能である。そのような同定されている配列は、現在、次に示すような種のゲノムに対し利用できる（`http://www.girinst.org`）：ヒト、マウス、シロイヌナズナ、線虫、ショウジョウバエ、ダニオ。

【0127】

実施例3

この実施例の中で、さらに遺伝子疾患や細胞遺伝学的な異常に特異的な沢山のプローブが、本発明の原理に基づいて作製された。ソフトウェアも開発され、シングルコピープローブを設計する過程を迅速に処理するのに貢献した（`find.pl`, `prim_wkg`, `and prim`, 前記および添付されているCD-R参照）。その後、本発明のプローブの試験を行い、`in situ`ハイブリダイゼーションによりその有効性を確認した。10

【0128】

シングルコピー配列の同定

シングルコピープローブ配列の位置は、連続した長いゲノムDNA配列から直接同定される。この位置は、標的のゲノム配列から繰り返し配列ファミリーの配列を検索するソフトウェアによって決定された。標的配列と前もって同定されている繰り返しファミリーの配列とを比較することにより、標的配列に中にある繰り返し配列の境界を同定し明確にできた。コンピューターソフトウェアの「RepeatMasker」（`http://ftp.genome.washington.edu/RM/RepeatMasker.html` : Smith A. F. A. & Green P., 結果は未公開である）は、連続したゲノム配列中にある繰り返し配列ファミリーの位置、たいてい100kbまでの長さを決定するために使用された。RepeatMaskerは、ゲノム配列と、ヒトゲノム中に複数のコピーが存在する繰り返し配列ファミリーとを、コンパイルすることで比較している（`http://www.girinst.org/~server/repbase.html`）。この繰り返し配列のデータベースには、代表的で共通性のあるヒトの繰り返し配列ファミリーの大部分の配列が登録されている。データベースは、（実施例6で示すように）新たに繰り返し配列ファミリーが発見されても、それを追加登録することにより、増やしていくことができる。20

【0129】

Perl言語（`findi.pl`）は、RepeatMaskerの出力結果から、繰り返し部分の境界と実質的に同等の部分を解明し、パラメーターとして入力した限界（多くの場合は、2kbまで）以上の大きさの、近接したシングルコピー部分を推測し、検索してくれる。このプログラム言語により、RepeatMaskerより出てきた（識別子：`.out`のついた）出力ファイルから、大きさにより並べ替えられたシングルコピー部分の位置と長さが決定される、この出力ファイルには、繰り返しファミリーの部分の位置と長さの情報の一覧表も含まれている。近接したシングルコピー部分の境界は、繰り返し部分の上流側の境界にある1つのヌクレオチドを除去し、前記繰り返し部分の下流側の境界に、1つのヌクレオチドを追加することにより導出された。同一の上流および下流に変化物（1bpの長さ）を有するシングルコピー部分は、近接した繰り返し配列とみなされた。そして、プローブの配列は、ヒトゲノム配列のデータベース（`Altschul et al.`, `J. Mol. Biol.`, 215: 403 - 410 (1990)）と比較されて、ゲノム中のどこかに相同性があるかどうか（例えば、デュープリコン、トリプリコン、またはもっと相同性の低い配列）が、同定された。ゲノム中のどこかに、弱いながらも保存されているプローブ配列は、標的配列とクロスしてハイブリダイゼーションしない。30

【0130】

最も長いシングルコピー部分のPCR增幅反応に対して、オリゴヌクレオチドプライマーが選択された。Unix wrapper言語（`prim_wkg`）は、プライマー（`prim`）を設計するコマンドを含んでいるファイルのスイッチを、何度も変化させ、このようにして、プログラム（`Prime Genetics Computer Group`）4050

p ; Madison , WI)) に入力される後述のパラメーターを変化させることにより、プライマー選択を最適化する：Tm (70 ~ 60) 、G / C 含量 (55 ~ 40 %) 、最小限の間隔配列の長さ (シングルコピー部分の長さの 90 ~ 80 %) 。

【 0131 】

プローブの作製および染色体に対する *in situ* ハイブリダイゼーション DNA 断片は、製造メーカー (Panvera , Madison , WI) により推奨されているよう 10 に、LA - Taq 用いてロング PCR (Cheng et al . , Proc . Natl . Acad . Sci . U . S . A . , 91 : 5695 - 5699 (1994)) で増幅された。 Roche Molecular Biochemicals , Indianapolis , IN ; Stratagene , La Jolla , CA ; および Invitrogen , Carlsbad , CA によって製造されている酵素を含め、ロング PCR 用の他の酵素でも同様の結果が得られることが分かった。増幅産物は、低温融解性アガロースゲル電気泳動、その後に、マイクロコン 100 カラム (Millipore , Bedford MA) を用いたクロマトグラフィーを行って精製した。この操作により、繰り返し配列を含め、混入している PCR の伸長反応産物を除去することができた。

【 0132 】

プローブ断片は、ジゴキシゲニン - dUTP や biotin - dUTP (Roche Molecular Biochemicals , Indianapolis , IN) のような修飾ヌクレオチドを用いて、ニックトランスレーションにより標識された。標識したプローブを変性し、前記条件下 (Knoll and Lichter , Current 20 Protocol in Human Genetics , Vol . 1 , Unit 4 . 3 , Green - Wiley , New York , (1994)) 、プローブを (C₀ t 1 DNA のような) 繰り返し DNA とプレアニ - リングする手法を、一連のハイブリダイゼーションにおいて使用しなかったことを除く) で、顕微鏡のスライドに固定した染色体プレパラートにハイブリダイゼーションさせた。非特異的な結合を除去するために、約 100 kb までの片方の染色体領域由来のプローブを、個々または組み合わせて使用しても、ハイブリダイゼーションした。ハイブリダイゼーション後の洗浄は、2 × SSC で 50 % のホルムアミド中、42 °C で行われ、続いてさらに 39 °C × SSC で 1 回、室温の 1 × SSC で 1 回行われた。ゲノム中に至る所にある類似の関係配列への、プローブのハイブリダイゼーションを排除するために、必要に応じて、洗浄条件を厳しくした。ハイブリダイゼーションしたプローブは、修飾ヌクレオチドに対する (ローダミンやフルオロセインのような) 蛍光色素が結合した抗体で検出された。染色体の同定は、細胞性の DNA を 4' 、 6 - デアミヂノ - 2 - フェニルインドール (DAPI) で対比染色することで行われた。ハイブリダイゼーションした染色体は、ディスプレイが付属し、複数の励起蛍光フィルター機器を備えた蛍光顕微鏡 (Olympus , Melville , NY) で視覚化された。 C₀ t 1 DNA とのプレアニ - リング有無の条件ごとに、個々のプローブまたはそれらの組み合わせについて、少なくとも 20 のメタフェーズ細胞 (50 ~ 100 の核) へのハイブリダイゼーション像を計数した。細胞は、CCD カメラ (Cohu , Inc , San Diego , CA) および CytoVision Chromofluor ソフトウェア (Applied Imaging , Santa Clara , CA) を使用して映像化された。 30 40

【 0133 】

第 4 表には、前記の方法を使用して作製した、既知の疾患や細胞遺伝子に関する異常に特異的な、多くのプローブに関するデータを示している。遺伝子異常の診断は、 1995 ISC N (国際ヒト細胞遺伝学会 (1995) , Mittelman F , 編) に挙げられるように、有名な学会でも利用されている。

【 0134 】

【 表 7 】

【 0 1 3 5 】
【 準 8 】

第4表

疾患	メタフェーズの染色体上のプローブによつて検出される、代表的な細胞遺伝学上の異常	遺伝子または転写産物	配列間隔	GenBank Accession No.	Forward PCR プライマーの座標、始点／終点	Reverse PCR プライマーの座標、始点／終点	PCR プライマー番号 Forward/Reverse
アングルマン症候群 およびブレーダー-ウル 症候群	ish del(15)(q11.2q11.2)(UBE3A-) ish del(15)(q11.2q11.2)(IC/SNRPN-)	UBE3A	IVS 8-IVS 9	AC004600	41085/41119	45354/45525	480/481
重複 15 症候群	ish dup(15)(q11.2q11.3)(UBE3A+); IC/SNRPN++ and ish del(15q11.2q13) (UBE3A++, IC/SNRPN++)	IC/SNRPN	IVS 3' to Exon u1B* IVS 5'-Exon u1B - IVS 3'	AC004737 AC004737	13740/13769 31102/31128	15414/15387 33347/33323	482/483 484/485
カルマン症候群	ish del(X)(p22.3)(KAL1-)	KAL1	-53 kb downstream	AC004602	104433/104465	49470/49441	486/487
ターナー症候群 および遠位 Xp 欠失	ish del(X)(p22.3)(SHOX-) ish del(X)(p22.3)(TBL1-)	SHOX	IVS 1-Exon 2 IVS 2 IVS 3	NT_001151 NT_001151 NT_001151	44615/44646 49637/49669 54537/54387	42042/42012 47051/47473 52251/52217	488/489 490/491 494/493
ダウン症候群	ish (2)(q22.2q22.3)(DSCR4x3) ish t(9;22)(q34;q11.2)(BCR st)	GS2 TBL1 TBL1 DSCR4 ² DSCR4 ² DSCR4 ²	Promoter-IVS 2 IVS 2 3' UTR ~39 kb upstream ~30 kb upstream -20 kb upstream	NT_001457 NT_001159 NT_001159 AP000160 AP000160 AP000160	78970/79000 175379/175409 247264/247293 310077/31041 40725/40754 49973/50006	82994/82960 179665/179633 251290/251257 32999/32965 43078/43045 52409/52376	498/499 500/501 502/503 504/505 506/507 508/509 510/511
慢性骨髓性白血病 および急性リンパ球性白血病	ish t(9;22)(q34;q11.2)(BCR st)	BCR	Proximal to the major breakpoint in CML	U07000	100745/100776	104145/104115	512/513
		BCR	Proximal to the major breakpoint in CML		74699/74728	77903/77874	
		IVS 8		U07000	1114946/1114978	1114571/111426	514/515
	ABL1 ³	Exon 1B-IVS 1B		U07361	27182/27213	29388/29357	516/517
	ABL1 ³	IVS 1B		U07362	91939/222	11035/11004	518/519
	ABL1	IVS 4-IVS 6		U07363	65951/65985	70266/70237	520/521
	ABL1	Exon 11-IVS 11		U07363	78862/78891	83813/83784	522/523

¹ u1B は、部分的に重複している PWS の最短領域から、そして部分的に重複している AS の最短領域から、約 160kb のところにある。

² プローブは、動原体末端染色体の p 脳に見られる配列にも、クロスハイブリダイゼーションする。プローブ配列は、公知の繰り返し配列データベースには、見出されない。

³ 公知の繰り返し配列データベースにも、プローブは、クロスハイブリダイゼーションする。

⁴PM5は、MYH11遺伝子から、約1.3mbテロメア側にあり、この遺伝子はinv(16p)切断点で分断されている。PLA2G10は、PM5から、約200kbテロメア側にある。

疾患	メタフェーズの染色体上のプローブに よって検出される、代表的な細胞遺伝 学上の異常	遺伝子または 転写産物	配列間隔	GenBank Accession No.	Forward PCR プライマーの 座標、始点 / 終点	Reverse PCR プライマーの 座標、始点 / 終点	PCR プライマーの 配列番号 Forward/ Reverse アーバー-
スミス-マグニス 症候群(続き)		<i>FJII</i>	IVS 12-IVS 14	U80184	74247453	87428708	578479
		<i>FJII</i>	IVS 15-Exon 21	U80184	96159647	1173811704	580581
		<i>MFAP4</i>	IVS 2-3' UTR	NT_008760	132621/132654	134663/134634	582583
	ish del(17)(p11.2p11.2)(MFAP4) ish del(17)(p11.2p11.2)(ZNF179;PAP1/ SHMT1)	<i>ZNF179-PAP1.⁵</i>	Between ZNF179 and PAP1. IVS 4	AL033367	98189850	122712241	584585
	ish del(17)(p11.2p11.2)(LGLL/HUGL-)	<i>LGLL</i>	Promoter-Exon 2	AL033367	11941226	5365/5334	586587
		<i>HUGL</i>	Promoter-IVS1				
	ish dup(17)(p11.2p11.2)(PMP22+)	<i>PMP22</i>	(~5 kb upstream)	AC005703	153173/153202	155027/154994	588589
		<i>PMP22</i>	~22 kb downstream	AC005703	215632/215661	217362/217329	590591
		<i>PMP22'</i>	IVS 3	AC005703	1846666/184700	186035/186006	592593
		<i>PMP22</i>	IVS 3	AC005703	176746/176778	179073/179044	594595
		<i>PAFAH1B1</i>	~5 kb downstream	NT_000774	63645/63679	66603/66573	596597
		<i>EIF-3⁶</i>	IVS 24-IVS 27				
		<i>PAFAH1B1</i>	~7-8 kb downstream	NT_000774	68841/68870	71195/71163	598599
		<i>EIF-3</i>	IVS 15-IVS 19				
		<i>PAFAH1B1</i>	~13 kb downstream	NT_000774	75328/75362	78122/78093	600601
		<i>EIF-3</i>	IVS 5-IVS 11				
		<i>JAG1</i>	IVS 5-IVS 8	AL03456.24	152935/153966	151675/157662	6021603
		<i>JAG1</i>	IVS 2-IVS 3	AL03456.24	144875/144904	147028/146995	604605
		<i>JAG1</i>	Exon 1-IVS 2	AL03456.24	133644/135673	139440/139407	606607
	モノソミー-13q32	ish del(13)(q32.3)(ZIC2-)	~5.8 kb downstream	AL355338	111114/111145	116046/116012	608609
	トリソミー-13	ish (13)(q32.3)(ZIC2;3)	~2 kb upstream	AL355338	128595/128627	133039/133006	610611
	ワルフルニー-ン症候群 ハーニッシュ-ン症候群	HD	Exon 67	NT_000102	267614/267643	271120/271091	612613

⁵ プローブは、個々ではなく、他のプローブと組み合わせて用いることにより、ハイブリダイゼーションする。

⁶ プローブは、PAFAH1B1 遺伝子(以前は LUS1 として知られていた)の下流および近傍にある。 EIF-3 と相同意性の発現してくる転写産物は、このあたりの配列座標で見つかる。

【 0 1 3 7 】

実施例 4

この実施例の中では、本発明のプローブを使用して、染色体の切断点を、さらに正確に決定することが行われた。構造的な染色体の再構成は、ある種のガンの場合と同様に、遺伝子疾患や後天性の病気において遺伝性である可能性が高い。そのような再構成は、(逆位

、欠失、重複のような) 単一の染色体内や(例えば、転座のように) 相同染色体間または非相同染色体間の中で起こることもある。本発明のプローブについては、切断が起きている正確な領域を、従来にはない解像度で決定することができる。そのレベルの解像度があると、再構成が起こる段階で散在してしまった遺伝子や核酸配列の検出が可能となり、病因、予後および/または治療において洞察を与えてくれる。遺伝的に関係のある遺伝子疾患において、染色体の切断点を正確に位置付けすることは、欠失または重複の範囲を明らかにし、それ故、遺伝子異常の予後を決定付ける(Cheng et al., Am. J. Hum. Genet., 55: 753-759 (1994))。

【0138】

次の実施例では、シングルコピープローブが、同一の染色体領域の上市されているクローニングプローブより、正確な情報を提供してくれることを、例示している。 10

【0139】

慢性成人骨髓性白血病の多くの場合(CML; 90%) および急性リンパ球性白血病の症例の中には(AAL; 25~30%の成人; 2~10%の小児)(Perkins et al., Cancer Genet. Cytogenet., 96(1): 64-80 (1997); Rubintz et al., J. Pediatr. Hematol. Oncol., 20(1): 1-11 (1998))、染色体9q34と22q11.2の間での相互転座が明らかとなっている(Rowley, Nature, 243: 290-392 (1973))。この転座の結果、異常のある、または派生した第22番染色体は、第22番染色体のBCR(切断点集中領域)プロモーターのところで、第9番染色体上の癌遺伝子ABL1と融合する。癌遺伝子ABL1は、択一的にスプライシングされた、それぞれエキソン1b, 1aと呼ばれる第1エキソンを有し、共通のエキソン2-11と接合した、6kbまたは7kbのmRNAのいずれかを発現する。エキソン1bは、エキソン1aの250kb位の近傍にあり、この非常に長いインtronが転座の絶好の標的となる。CMLにおいて、ABL1遺伝子は、第9番染色体から第22番染色体のBCR遺伝子のプロモーター領域に、BCR-ABL1キメラタンパク質を産生するように転座してくれる(Bernalds, et al., Mol. Cell. Biol., 7: 3231-3236; (1987))。BCR遺伝子は、24のエキソンを有し、160kDのタンパク質をコードしている。BCR遺伝子の切断点は、CMLとALLとで異なる。CMLでは、大部分の切断点は、エキソン12から17の間(すなわち、b1からb5)に相当する、5.8kbのメジャー切断点集中領域(M-BCR)の中で起こる;しかし、多くのALLの場合、BCR遺伝子は、エキソン1と2の間(マイナー切断点集中領域、すなわちm-BCR)で切断される。このように、異なる分子再構成により、異なる大きさのタンパク質が生じ、それぞれ異常を引き起こす。このように遺伝子の再構成を、本発明のプローブにより区別することができる。 20 30

【0140】

DNAレベルでは、2つの色と2つのプローブによる蛍光in situハイブリダイゼーション(FISH)の手法が、メタフェーズおよびインターフェーズの核上のBCR-ABL1転座部位を検出するのに利用できる(Bentz et al., Blood, 83: 1922-1928 (1994); Sinclair et al., Blood, 90: 1395-1402 (1997); Buno et al., Blood, 92: 2315-2321 (1998))。従来のFISHに関して、その技術は、癌遺伝子ABL1の領域(切断点および200kbまでの長さ)を、ある色(例えば、赤色)で検出するプローブと、(切断点の近傍にある)BCR領域を、別の色(例えば、緑色)で検出するプローブとを有するものである。このように、融合した第22番染色体陽性細胞では、赤色と緑色のシグナルが同じ位置にあって、結果として第22番染色体の存在を示すように黄色のハイブリダイゼーションシグナルを与える。しかし、正常な第9番と第22番染色体は、それぞれ独立に、赤色と緑色のシグナルのままになっている。最近では、ABL1遺伝子転座切断点の両側にわたる、もっと大きなDNAプローブが転座部位を検出するために利用されている。この手法では、切断点の近くにあるプローブ部分は、 40 50

異常のある第9番染色体上に残っており、切断点から遠くにある部分は、従来の検出方法と同様に、BCR領域の部位にある。このため、転座した第9番染色体上に余分なシグナル(ES)が観察され、多くの臨床的細胞遺伝学の研究室が使用しているBCR-ABL1 ESの2つの色による転座プローブ(Vysis, Inc., Downers Grove, IL)より劣った手法である(Herens et al., Br. J. Haem., 110(1):214-216(2000); Sinclair et al., Blood, 95(3):738-744(2000))。このESシステムにおいて、ABL1遺伝子プローブカクテルは、ABL1遺伝子よりもはるかに大きいゲノム上の標的を覆う。このカクテルは、ABL1遺伝子の約250kb上流にあるアルギノスクシネット合成酵素の遺伝子から、ABL1遺伝子を通って、その数kb下流まで伸びている。

10

【0141】

BCR、ABL1遺伝子のローカスのシングルコピー部分の境界を同定して、いくつかのシングルコピープローブが作製された。第13および第14図は、それぞれBCR、ABL1遺伝子のローカス内のシングルコピープローブ部分の位置を示している。2kbを越える11の部分が、BCR遺伝子全体にわたり分布し、これまでに、そのうちの5つをプローブとして作製した。さらに、より正確に転座切断点を明確にするために、この遺伝子中の同程度の比較的短いシングルコピー領域(1.5~2.0kb)を、組み合わせて使用することができた。

【0142】

BCR遺伝子(第13図)およびABL1遺伝子(第14図)に対する多様なシングルコピープローブが導出され、オリゴヌクレオチドプライマーセットが導き出された。BCR遺伝子に関して、これらのプローブは、マイナー切断点とメジャーな切断点(第13図および配列番号510~515)を区別する。BCR-ABL1のESプローブを用いた従来のFISH試験では、BCR遺伝子のメジャー(M)な切断点とマイナー(m)な切断点での再構成を区別できない。ALLの患者でよく観察されるマイナーな切断点を検出するm-BCRプローブと、M-BCR(配列番号510~515)を明示するプローブとを組み合わせることで、CMLの大部分の患者で明らかとなっているメジャーな切断点を検出できる。もし、M-BCRのプローブだけが、第22番染色体から、融合した第9番染色体に転座し、m-BCRプローブが、融合した第22番染色体上に残るなら、遺伝子は、メジャーな切断点のところで分断されているのである。

20

【0143】

ABL1遺伝子に関して、プローブのうち2つは、切断点の近傍にあると推測され(配列番号516/517および518/519)、他のプローブは切断点から遠く離れたところにある(配列番号520および531)。CMLの切断点の遠位にある3つのABL1のプローブのハイブリダイゼーションから、プローブは融合した第22番染色体側に移動したことが示され(第15図)、切断点部分にまたがる5つのABL1のプローブを組み合わせた結果から、いくつかのプローブは融合した第9番染色体側に残り、他のプローブは、融合した第22番染色体側に移動している(第16図)ことが示されている。これらの結果とこれらのプローブの位置関係から、切断点部分は第14図の11004bpから65951bpの位置にまたがっているという結論を導くことができる。分断点の領域を、配列番号519と520間のシングルコピー部分由来のハイブリダイゼーションプローブで、より明確することができるということは、この技術分野に熟知した人にはあきらかである。そして、切断点の正確な位置は、正確な分断点領域のゲノム配列から決定することができる。

30

【0144】

上市されているESプローブを使用している最近の研究(Herens et al., Br. J. Haem., 110:214-216(2000))から、およそ10%のCMLの患者では、融合した第9番染色体上で、余分にハイブリダイゼーションしないことが示されている。(プローブの)配列が、ABL1遺伝子の欠失した上流側であるからである。症例の中には、欠失部分が、ASS遺伝子と同じ方向まで及んでいることもあ

40

50

り、そのような欠失は、予後が悪いこと（すなわち、白血病の急性転化）と関連している。約 100 kb から 200 kb までの、比較的短いプローブがハイブリダイゼーションする、約 1 / 3 の CML の患者では、欠失部分の検出率が上昇することから明らかのように、上述の長い FISH プローブは、ASS とABL1 の間の領域に存在する介在性の欠失を検出するのに有効ではない (Sinclear et al., Blood, 95 (3) : 738 - 744 (2000))。患者の中には、元のまたは転座した第 9 番染色体上の ABL1 切断点近傍の配列が欠失している人もいるので、シングルコピー プローブは、この染色体領域中の半接合の範囲を明らかにするために使用されている。欠失切断点と臨床的な症例との相関により、この染色体部分の特定の遺伝子欠損が、初期の急性転化のような臨床的な症状の予後兆候であるかどうかが決まってくる。

10

【0145】

実施例 5

この実施例では、増幅されたハイブリダイゼーションプローブのシグナルが、二倍化したゲノム領域由来のプローブ配列を使用して発生していることが示されている。いくつかのシングルコピー プローブ配列が設計されたが、それは上述の通り、非常に類似したデュープリコンまたはトリプリコンの一部である。

【0146】

急性骨髓性白血病 - タイプ M4 の逆位による切断点部位に近い、染色体 16 p 13 . 1 (配列番号 536 ~ 543) 由来のいくつかのプローブでは、この領域内の配列の、ほとんど完全に 3 倍重複した配列を含んでいる。ゲノムのドラフト配列において、これらのドメインの 2 つは一列に位置し、プローブ配列が 40 kb の配列で分断されている。そして 3 番目の末端部分は、1 . 2 mb で分断されている。3 つの部分の配列は、1 . 5 % しか違っていない。このプローブを用いたハイブリダイゼーションは、2 つの領域に集中しているが、明確に区別できるシグナルではない。一方のハイブリダイゼーションは、連続している始めと 2 番目のパラログに対応しており、もう一方はこの配列の 3 番目のコピーに対応している。

20

【0147】

Smit h - Magenis 症候群において、一般に欠失している染色体 17 p 11 . 2 (配列番号 586 - 587) 由来のプローブは、ほとんど完全なトリプリコンを含んでいる。このプローブは、SHMT1 遺伝子の IVS4 に存在する、ほぼシングルコピーである配列の中の欠失を検出するようにできている。しかし、SMHT1 配列と 99 . 8 % の相同性を示す、およそ 15 kb の配列で分断されている 2 つのパラローガスなサブ配列は、動原体側に約 2 . 7 mb 離れた ZNF127 と PAIP1 遺伝子間のゲノムドラフト中にも検出され、この異常のある患者で、共通して欠失している遺伝子配列でもある。これらの配列の第 17 番染色体の動原体（高度に凝集した領域）までの距離に従い、シングルハイブリダイゼーション部位が観察された。

30

【0148】

ダウン症候群で重要な領域（配列番号 504 / 505 - 508 / 509）由来のプローブはそれぞれ、同じく大きな、80 kb 以上の長さがある、デュープリコン中に結合した。これらの 2 倍重複した配列は、1 . 1 mb 離れており、それぞれ染色体 21 q 22 . 2 バンドの動原体側およびテロメア側に存在している。2 倍重複した配列は、1 . 1 mb の配列で分断されているにも関わらず、それぞれに対するプローブを使用しても、1 つのハイブリダイゼーションシグナルしか、第 21 番染色体のこの領域には検出されなかった。それ故、染色体 17 p 11 . 2 と同じように、第 21 番染色体のこの領域は、同じ位の距離で分断された 2 倍重複した領域を区別できる染色体 16 p 13 . 1 よりも、メタフェーズの染色体中では凝縮しているように思われる。

40

【0149】

そのような領域から作製されたシングルコピー プローブは、当然だが、既知の繰り返し配列要素を欠いている。しかし、それぞれのコピーは、最も厳しいハイブリダイゼーション洗浄条件下でもハイブリダイゼーションしたままなので、プローブは、概してパラローガ

50

スなコピーのすべてにハイブリダイゼーションする。染色体上の多数のきつく塊状になつた部位は、特異的な部分にハイブリダイゼーションするので、このようなハイブリダイゼーション由来のシグナルは、1倍体ゲノムごとに1つ存在するプローブ配列との比較から予想されるよりも、はっきりしている。このように、ゲノム内のデュープリコンまたはトリプリコンは、本発明のプローブで好適な標的の大きさを長くできる。このことは、そのような領域由来の比較的短いプローブは、比較的長いプローブと比較して、同等な強度のシグナルを与えることができることを意味している。2倍重複したゲノムドメイン由来の比較的短いプローブを選択することは、長いシングルコピー部分と予想していたゲノム領域に対するプローブの開発に、特に役に立つ。

【0150】

10

実施例6

正確なヒトゲノム配列の草案が急速に利用できるようになると、本発明の方法に従い、今まで利用できなかつた数多くの染色体領域に対して、シングルコピープローブの開発が進んだ。現在、最新の総合的なDNA配列のデータベースは、ドラフト配列に存在する繰り返し部分を検出するために利用されているが(<http://www.girinst.org/repbase>)、メタフェーズの染色体に対するシングルコピープローブのハイブリダイゼーションの結果から、これまでに判明していない繰り返し配列を含むプローブもあることが示されている。このことは、草案のゲノム配列中には見つかっていない他の位置と同様にマップされることが知られている、相同染色体バンドへのプローブのハイブリダイゼーションを記述することで明確になった。

20

【0151】

20

ドラフトゲノム配列は不完全であり、ユークロマチンゲノムの90%位までが、配列決定されている(国際ヒトゲノムシークエンス学会、Initial Sequencing and Analysis of the Human Genome, Nature, 409:860-922, (2001))。いくつかの繰り返し配列ファミリー、とりわけ未知の配列の中に存在しているものがあつても、検出されないことが期待された。既知の繰り返し配列の検索にも関わらず、いくつかのユークロマチンシングルコピープローブは、末端動原体染色体(第13, 14, 15, 21および22番染色体)の短腕上に、多数のコピーが優勢的に見つかる、繰り返し配列のホモログを含むように思われる。これらのプローブは、第21番染色体のダウン症候群で重要な領域由来の3つの配列(配列番号504/505, 506/507および508/509で増幅される)と、急性骨髓性白血病-タイプM4で、染色体の逆位の部位に広がっている染色体16p13.1由来の2つの配列(配列番号536/537および538/539で増幅される)を含んでいる。

30

【0152】

30

核小体形成領域と呼ばれる、染色体ドメインは、連続して一列に並んだ何千ものコピーのリボソームRNAを含んでいることが知られている(Sylvester et al., Hum. Genet., 73:193-8 (1986))。ヒトゲノムの配列決定されたゲノム配列には、染色体領域由来の連続配列が欠けている。国際シークエンス会議は、このような配列や、その他の連続した繰り返し配列(例えば、ヘテロクロマチン)を含むクローンを、シークエンス当初の事情を配慮して排除している。何故なら、どのようなクローンからの配列を集めても、不明瞭で、信頼できないと認識しているからである。同じ位置にあるが、リボソームRNA遺伝子とは区別される他の配列も、恐らく染色体の核小体形成領域中に、一列に並んで存在している。このような部分についての配列情報がないので、ゲノム中のどこかにある配列に関係した領域内の配列の分布を、あらかじめ予測することはできない。このようなプローブを用いたシングルコピーFISHは、予測通りにユークロマチン部分への局在化を示しているが、さらに、末端動原体を有する染色体の短腕への、注目すべきハイブリダイゼーションが観察される。ハイブリダイゼーションの前にC₀t1DNAを添加すると、繰り返し配列へのクロスハイブリダイゼーションを除去できた。このような付隨的なシグナルは、末端動原体を有する染色体の短腕上のプロ

40

50

ープと関係して、一列に並んだマルチコピーの配列と一致している。

【0153】

さらに、Repeat Maskerソフトウェアを用いて繰り返し配列を排除したプローブであるにも関わらず、シングルコピーと導出されたプローブの、介在性の繰り返し配列ファミリーに対するハイブリダイゼーションが検出された。Smith-Magenis症候群で共通して欠失している部分にあり、染色体17p11.2にマッピングされる、1つのプローブでは、介在性の繰り返し配列とクロスしてハイブリダイゼーションするものが観察された。染色体9q34の転座切断点の近傍の配列（配列番号518/519および526/527で増幅される）も、これらのプローブをこの領域由来の別のプローブと組み合わせてハイブリダイゼーションさせた結果から、潜在的に介在性の繰り返し配列を含んでいる。これらのプローブ部位にマップされるハイブリダイゼーションシグナルは、クロスしてハイブリダイゼーションしている配列のそれより強くはないが、洗浄処理を厳しい条件にしても、クロスしてハイブリダイゼーションする配列を排除することはできない。この実験事実は、プローブは、非常に多コピーの公知の介在性繰り返し配列というよりは、（Repeat Maskerによって認識されないので、作製されたプローブの中に、偶然に含まれてしまう）今までに判明にしていない繰り返し配列ファミリーを含んでいることがわかる。

【0154】

このようなプローブは全部、これまでに発見されていない繰り返し配列ファミリーの類を含んでいるように思われるが、プローブ自体は、シングルコピーと繰り返し配列の両方で構成されることが多い。染色体DNAに対して、PCRで產生された異なるプローブ配列のセットを何度もハイブリダイゼーションして、これらの配列要素を分離することができる。しかし、それぞれの全プローブ配列は既に判明しているので、さらに新規のシングルコピープローブを作製するために用いられる繰り返し配列のデータベースに、その配列を追加することができる。プローブの配列は長くないので（実際には、介在性繰り返し配列の中には、最も長いシングルコピープローブよりも長いものもある；例えば、L1）；それ故、これらの配列をヒト繰り返し配列ファミリーのデータベースに追加することにより、最小限の計算機のオーバーヘッドで済ませられる。このような以前には未知である繰り返し配列ファミリーを追加することにより、より利用価値のある繰り返し配列データベースを構築でき、シングルコピープローブの設計を向上させることができる。以後、さらに大きな繰り返し配列データベースを用いて設計されたシングルコピープローブは、こうした新規の繰り返し配列を含むことはないであろう。発見的アルゴリズムにより、シングルコピープローブ中のシングルコピー部分の配列の割合を向上させることができる。

【0155】

概して、このように本発明は、ゲノム中で未知であった繰り返し配列ファミリーの存在を決定する手法を提供する。この手法は、標識され、シングルコピーと推測されてる核酸プローブが、ゲノムと反応しプローブのハイブリダイゼーションを引き起こす過程と関連している。もし、プローブが、3つ以上の異なる部位（そして、好ましくは10以上の部位）とハイブリダイゼーションするなら、新規で未知の繰り返し配列が見つかったと判断してよい。

【0156】

上記で引用されたすべての参考文献は、本明細書中に参考文献として、適宜引用されている。さらに、2000年3月27日に出願された公開文書471449号明細書の主な記載事項も、本明細書中の参考文献で引用されている。

【0157】

関連出願

この出願は、2000年5月16日に出願されたシリアル番号09/573080の関連出願である。

【0158】

[配列表]

10

20

30

40

50

本発明に関連して、コンピューターで読み込むことができるA S C I I ファイルの形式で、613の配列を含む配列表が、参考資料として本発明に引用されている、また、37 C F R 1 . 8 2 1 (c)と関連した1枚(1)のオリジナルのコンパクトディスク、37 C F R 1 . 8 2 1 (e)と関連した同じ内容の1つの書面、および37 C F R 1 . 5 2と関連した同じ内容の1つの書面が添付されている。

【0159】

本発明で使用する可能性のあるコンピュータープログラムのソースコードを含むコンピュータープログラムの一覧が、参考資料として本発明に引用されている、また、後記の全3つの電子ファイルを含んでいる1枚(1)のコンパクトディスクおよびそれと同じ内容の書面が添付されている。

10

作成日	ファイルサイズ(バイト)	ファイル名
04/18/01	26 KB	FINDI.PL
04/18/01	19 KB	PRIM.IN
04/18/01	20 KB	PRIM.WKG

【図面の簡単な説明】

第1～12図は、それぞれFISHの結果を、CCDカメラで撮影した写真である。その中で、様々な遺伝子特異的ジゴキシゲニン-dUTP標識プローブは、メタフェーズの細胞にハイブリダイゼーションしており、ローダミンが共有結合したジゴキシゲニンに対する抗体により検出された。また、染色体を、4, 6-ジアミジノ-2-フェニルインドール(DAPI)で、対比染色した。染色体の片方または両方にハイブリダイゼーションしている染色体は、矢印によって示されており、一方星印は、通常であれば予想されるハイブリダイゼーションが見られないことを示している。以下に、図面について詳細に説明する。

20

【図1】第1図は、実施例1で記載されている5170bpのHIRAプローブを用いて、ハイブリダイゼーションを行った結果を示した写真である。このプローブは、精製した繰り返しのあるDNA配列とあらかじめ反応させたものである。

20

【図2】第2図は、第1図の対照実験に対応するハイブリダイゼーションの結果を示した写真である。同一の5170bpのHIRAプローブを使用しているが、精製した繰り返しのあるDNAとの前反応を行っていないものである。

30

【図3】第3図は、精製した繰り返し配列を有するDNAと、あらかじめ反応させた3544bpの15q11-q13プローブを用いて、ハイブリダイゼーションを行った結果を示した写真である。

【図4】第4図は、精製した繰り返し配列を有するDNAと、前もって反応させていない3544bpの15q11-q13プローブを使用した、対照実験の結果を示した写真である。

【図5】第5図は、精製した繰り返し配列を有するDNAと、前もって反応させていない、実施例2で記載された4166bp、3544bpおよび2290bpの15q11-q13プローブを使用して行ったハイブリダイゼーションの結果を示した写真である。

【図6】第6図は、精製した繰り返し配列を有するDNAと、前もって反応させていない、実施例1で記載された5170bp、3691bp、3344bpおよび2848bpのHIRAプローブを使用して行ったハイブリダイゼーションの結果を示した写真である。

40

【図7】第7図は、精製した繰り返し配列を有するDNAと、あらかじめ反応させた、実施例2で記載されている4823bpの1p36.3プローブを用い、メタフェーズのそれぞれ正常な細胞に対してハイブリダイゼーションを行った結果を示した写真である。

【図8】第8図は、精製した繰り返し配列を有するDNAと、前もって反応させていない4823bpの1p36.3プローブを使用した、対照実験であるハイブリダイゼーションの結果を示した写真である。

【図9】第9図は、精製した繰り返し配列を有するDNAと、あらかじめ反応させた、実

50

施例2で記載されている4724bpおよび4823bpの1p36.3プローブを使用して行ったハイブリダイゼーションの結果を示した写真である。シングルコピーハイブリダイゼーションは、相同な第1番染色体の短腕上に観察された。

【図10】第10図は、精製した繰り返し配列を有するDNAと、前もって反応させていない、実施例2で記載されている4724bpおよび4823bpの1p36.3プローブを使用した対照実験であるハイブリダイゼーションの結果を示した写真である。第9図と同様のシングルコピーハイブリダイゼーションであることが示されている。

【図11】第11図は、精製した繰り返し配列を有するDNAと、前もって反応させていない、実施例2で記載された4166bp、3544bpおよび2290bpの15q11-q13プローブを、染色体アレルの15q11-q13の欠失に原因があることが知られているプレーダー・ウィリ症候群に罹患した患者のメタフェーズの細胞に対して使用したハイブリダイゼーションの結果を示した写真である。図の中で、星印は欠失した染色体の位置でハイブリダイゼーションが見られないことを示しており、矢印は染色体に1ヶ所だけハイブリダイゼーションしていることを示している。

【図12】第12図は、精製した繰り返し配列を有するDNAと、前もって反応させていない、実施例1で記載された3691bp、3344bpおよび2848bpのHIRAプローブを、染色体アレルの22q11.2配列の欠失に原因があることが知られているDiGeorge/Velcro-Cardio-Facial症候群(VCFS)に罹患した患者のメタフェーズの細胞に対して使用したハイブリダイゼーションの結果を示した写真である。図の中で、星印は欠失した染色体の位置でハイブリダイゼーションが見られないことを示しており、矢印は正常な対立染色体を示している。

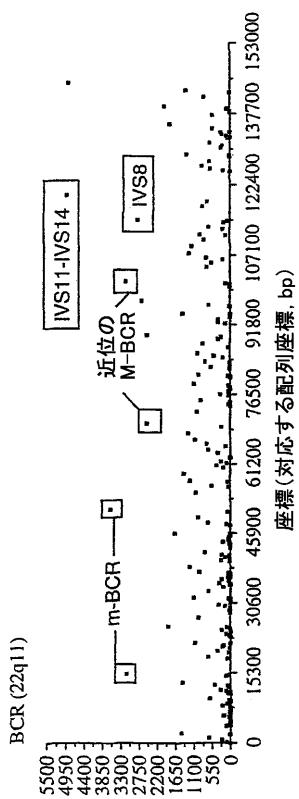
【図13】第13図は、第22番染色体にあるBreakage Cluster Region遺伝子(BCR)プロモーターにみられる、シングルコピープローブ全長に対応する、スキャッタープロットである。この領域の異常は、実施例4で記載されているように、慢性成人骨髓性白血病の症例で共通して見られ、急性リンパ球白血病でも観察される。

【図14】第14図は、第9番染色体にあるABL1遺伝子にみられる、シングルコピープローブ全長に対応する、スキャッタープロットである。この領域の異常は、実施例4で記載されているように、慢性成人骨髓性白血病で共通して見られ、急性リンパ球白血病でも異常が観察される。

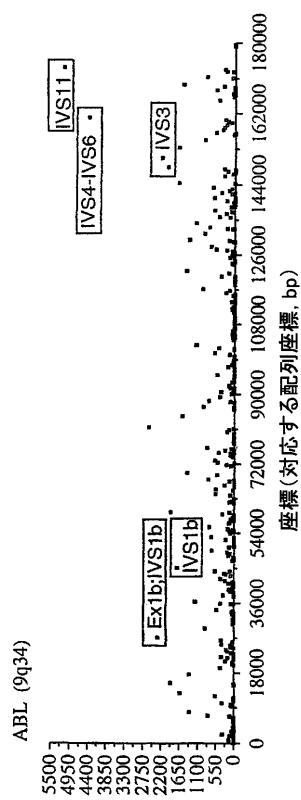
【図15】第15図は、ABL1癌遺伝子(配列番号520~525)由来の染色体である、9q34-特異的ジゴキシゲニン-dUTP標識プローブを使用し、慢性骨髓性白血病(CML)の患者から採取したDAP1染色メタフェーズ細胞にある、ローダミンが共有結合したジゴキシゲニンに対する抗体により検出したFISHの結果を、CCDカメラで撮影した写真である。実施例4で詳述したように、染色体切断点を正確に決定するために、ABL1遺伝子とBCR遺伝子の間の融合領域の下流にあるプローブを使用していることが例示されており、融合した第22番染色体と正常な第9番染色体が示されている。

【図16】第16図は、ABL1癌遺伝子(配列番号516~525)由来の染色体である、9q34-特異的ジゴキシゲニン-dUTP標識プローブを使用し、CMLの患者から採取したDAP1染色メタフェーズ細胞を用い、ローダミン(red)が共有結合したジゴキシゲニンに対する抗体により検出したFISHの結果を、CCDカメラで撮影した写真である。実施例4で詳述したように、染色体領域9q34の切断点を正確に決定するために、BCR遺伝子とABL1遺伝子間の融合部位の、それぞれ由来のプローブを使用していることが例示されており、融合した第22番染色体、融合した第9番染色体および正常な第9番染色体が示されている。

【図13】



【図14】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/88089 A2

(51) International Patent Classification: C12N

(74) Agent: COLLINS, John, M., Hovey, Williams, Timmons

& Collins, Suite 400, 2405 Grand Boulevard, Kansas City,

MO 64108 (US).

(21) International Application Number: PCT/US01/15674

(22) International Filing Date: 15 May 2001 (15.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/573,080 16 May 2000 (16.05.2000) US
09/884,867 14 May 2001 (14.05.2001) US

(71) Applicant (for all designated States except US): CHILDREN'S MERCY HOSPITAL [US/US]; 2401 Gillham Road, Kansas City, MO 64108 (US).

(72) Inventors; and
(75) Inventors/Applicants (for US only): KNOLL, Joan, H., M. [US/US], 12226 Gillette Road, Overland Park, KS 66213 (US). ROGAN, Peter, K. [US/US], 12226 Gillette Road, Overland Park, KS 66213 (US). CAZARRO, Patricia, M. [I]. --.

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JR, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MR, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

{Continued on next page}

(54) Title: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING SAME



WO 01/88089 A2



(57) Abstract: Nucleic acid (e.g., DNA) hybridization probes are described which comprise a labeled, single copy nucleic acid which hybridizes to a deduced single copy sequence interval in target nucleic acid of known sequences. The probes, which are essentially free of repetitive sequences, can be used in hybridization analyses without adding repetitive sequence-blocking nucleic acids. This allows rapid and accurate detection of chromosomal abnormalities. The probes are preferably designed by first determining the sequence of at least one single copy interval in a target nucleic acid sequence, and developing corresponding hybridization probes which hybridize to at least a part of the deduced single copy sequence. In practice, the sequences of the target and of known genomic repetitive sequence representatives are compared in order to deduce locations of the single copy sequence intervals. The single copy probes can be developed by any variety of methods, such as PCR amplification, restriction or exonuclease digestion of purified genomic fragments, or direct synthesis of DNA sequences. This is followed by labeling of the probes and hybridization to a target sequence.

WO 01/88089 A2

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/88089

PCT/US01/15674

-1-

SINGLE COPY GENOMIC HYBRIDIZATION PROBES
AND METHOD OF GENERATING SAME

5

RELATED APPLICATION

This is a continuation-in-part of Serial No. 09/573,080 filed May 16, 2000.

SEQUENCE LISTING

A Sequence Listing containing 613 sequences in the form of a computer readable ASCII file in connection with the present invention is incorporated herein by reference and appended hereto as one (1) original compact disk in accordance with 37 CFR 1.821(c), an identical copy thereof in accordance with 37 CFR 1.821(e), and one (1) identical copy thereof in accordance with 37 CFR 1.52(e).

15

COMPUTER PROGRAM LISTING APPENDIX

A computer program listing appendix containing the source code of a computer program that may be used with the present invention is incorporated herein by reference and appended hereto as one (1) original compact disk, and an identical copy thereof, containing a total of 3 files as follows:

20

	Date of Creation	Size (Bytes)	File Name
	04/18/01	26 KB	FINDI.PL
	04/18/01	19 KB	PRIM.IN
	04/18/01	20 KB	PRIM.WKG

25

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention is broadly concerned with a method for designing single copy hybridization probes useful in the fields of cytogenetics and molecular genetics for determining the presence of specific nucleic acid sequences in a sample of eukaryotic origin, e.g., the probes may be used to analyze specific chromosomal locations by *in situ* hybridization as a detection of acquired or inherited genetic diseases. More particularly, the

WO 01/88089

PCT/US01/15674

-2-

invention pertains to such probes, hybridization methods of use thereof and techniques for developing the probes, where the probes are essentially free of genomic repeat sequences, thereby eliminating the need for disabling of repetitive sequences which is required with conventional probes.

5

Description of the Prior Art

10

Chromosome abnormalities are associated with various genetic disorders, which may be inherited or acquired. These abnormalities are of three general types, extra or missing individual chromosomes (aneuploidy), extra or missing portions of chromosomes (including deletions, duplications, supernumerary and marker chromosomes), or chromosomal rearrangements. The latter category includes translocations (transfer of a piece from one chromosome onto another chromosome), inversions (reversal in polarity of a chromosomal segment), insertions (transfer of a piece from one chromosome into another chromosome) and isochromosomes (chromosome arms derived from identical chromosomal segments). The abnormalities may be present only in a subset of cells (mosaicism), or in all cells. Inherited or constitutional abnormalities of various types occur with a frequency of about one in every 250 human births, with results which may be essentially benign, serious or even lethal. Chromosomal abnormalities are common and often diagnostic in acquired disorders such as leukemia and other cancers.

15

Hybridization probes have been developed in the past for chromosome analysis and diagnosis of abnormalities. The probes comprise cloned or amplified genomic sequences or cDNA. For example, U.S. Patents Nos. 5,447,841, 5,663,319 and 5,756,696 describe hybridization probes in the form of labeled nucleic acids which are complementary to nucleic acid segments within target chromosomal DNA. However, these probes contain repetitive sequences and therefore must be used in conjunction with blocking nucleic acids which are substantially complementary to repetitive sequences in the labeled probes. That is, these prior art probes are either pre-reacted with the blocking nucleic acids so as to bind and block the repetitive sequences therein, or such blocking nucleic acids are present in the hybridization reaction mixture. If the repetitive sequences in the probes are not disabled in some manner, the probes will react with the multiple locations in the target chromosomal DNA where the repetitive sequences reside and will not specifically react with the single

20

25

30

WO 01/88089

PCT/US01/15674

-3-

copy target sequences. This problem is particularly acute with interspersed repeat sequences which are widely scattered throughout the genome, but also is present with tandem repeats clustered or contiguous on the DNA molecule. The requirement for repeat sequence disabilization by using complementary blocking nucleic acids reduces the sensitivity of the existing probes. Reliable, easily detectable signals require DNA probes of from about 40-100 kb.

The prior art also teaches that cloned probes presumed to contain single copy sequences can be identified based on their lack of hybridization to radiolabeled total genomic DNA. In these other studies, hybridization is first performed with probes that contain pools of clones in which each recombinant DNA clone has been individually selected so that it hybridizes to single-copy sequences or very low copy repetitive sequences. A prerequisite step in this prior art is to identify single copy sequences by experimental hybridization of labeled genomic DNA to a candidate DNA probe by Southern or dot-blot hybridization. Positive hybridization with labeled total genomic DNA usually indicates that the candidate DNA probe contains a repetitive sequence and eliminates it from consideration as a single copy probe. Furthermore, an experimental hybridization of a DNA probe with total genomic DNA may fail to reveal the presence of multicopy repetitive sequences that are not abundant (<100 copies) or are infrequent in the genome. Such sequences represent a small fraction of the labeled genomic DNA and the signal they contribute will be below the limits of detection.

It has also been suggested to physically remove repeat sequences from probes by experimental procedures (Craig et al., *Hum. Genet.*, **100**:472-476 (1997); Durm et al., *Biotech.*, **24**:820-825 (1998)). This procedure involves prehybridizing a polymerase chain reaction (PCR)-amplified genomic probe with an excess of purified repetitive sequence DNA prior to applying the probe to the DNA target. The resulting purified probe is depleted of repetitive sequences. This procedure is in principle very similar to other procedures that disable the hybridization of repetitive sequences in probes, but the technique is time-consuming and does not provide any advantages over the probes described in Patents Nos. 5,447,841 and 5,756,696.

30

WO 01/88089

PCT/US01/15674

-4-

SUMMARY OF THE INVENTION

The present invention overcomes the problem outlined above and provides nucleic acid (e.g., DNA) hybridization probes comprising a labeled, single copy nucleic acid which hybridizes with a deduced single copy sequence interval in target nucleic acid of known sequence. Generally speaking, the probes of the invention are designed by comparing the sequence of a target nucleic acid with known repeat sequences in the genome of which the target is a part; with this information it is possible to deduce the single copy sequences within the target (i.e., those sequences which are essentially free of repeat sequences which, due to the lack of specificity, can mask the hybridization signal of the single copy sequences). As can be appreciated, these initial steps require knowledge of the sequences both of the target and genomic repeats, information which is increasingly available owing to the Human Genome Project and related bioinformatic studies. Furthermore, readily available computer software is used to derive the necessary single copy sequences.

The probes hereof are most preferably complementary to the target sequence, i.e., there is a 100% complementary match between the probe nucleotides and the target sequence. More broadly, less than 100% correspondence probes can be used, so long as the probes adequately hybridize to the target sequence, i.e., there should be at least about 80% sequence identity between the probe and a sequence which is a complement to target sequences, more preferably at least about 90% sequence identity.

Nucleic acid fragments corresponding to the deduced single copy sequences can be generated by a variety of methods, such as PCR amplification, restriction or exonuclease digestion of purified genomic fragments, or direct nucleic acid synthesis. The single copy fragments are then purified to remove any potentially contaminating repeat sequences, such as, for example, by electrophoresis or denaturing high pressure liquid chromatography; this is highly desirable because it eliminates spurious hybridization and detection of unrelated genomic sequences.

The probe fragments may then be cloned into a recombinant DNA vector or directly labeled. The probe is preferably labeled by nick translation using a modified or directly labeled nucleotide. The labeled probe is then denatured and hybridized, preferably to fixed chromosomal preparations on microscope slides or alternately to purified nucleic acid immobilized on a filter, slide, DNA chip, or other substrate. The probes can then be

WO 01/88089

PCT/US01/15674

-5-

hybridized to chromosomes according to conventional fluorescence *in situ* hybridization (FISH) methods such as those described in U.S. Patents Nos. 5,985,549 or 5,447,841; alternately, they can be hybridized to immobilized nucleic acids according to the techniques described in Patents Nos. 5,110,920 or 5,273,881. Probe signals may be visualized by any 5 of a variety of methods, such as those employing fluorescent, immunological or enzymatic detection reagents.

Use of the probes of the invention permits more precise chromosomal breakpoint determinations, to a level of resolution heretofore unobtainable by *in situ* hybridization. In 10 such analyses, initial probe sets can be prepared from regions believed to be on opposite sides of the breakpoint. After an initial assay to confirm this, successive additional probes closer to the breakpoint can be designed, using the single copy strategy. In this fashion, the precise region of the breakpoint can be determined.

It has been found that use of putative single copy probes can determine the existence 15 of heretofore unknown repeat sequences in a genome. The heretofore unknown repeated sequence families can then be included in the repetitive sequence database so that these sequences can be used in the design of subsequent single copy probes.

It was also found that probes may contain sequences that are duplicated or triplicated 20 in the genome which can have stronger hybridization due to the increased length of the target sequence. Also, these duplicons or triplicons can be confirmed, as such, using single copy probes which is more difficult with available commercial probes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-12 are respective CCD camera images of FISH experiments wherein 25 various gene-specific digoxigenin-dUTP labeled probes were hybridized on metaphase cells and detected with rhodamine conjugated antibody to digoxigenin and where the chromosomes were counterstained with 4,6-diamidino-2-phenylindole (DAPI). Chromosomes with one or both chromatids hybridized are indicated by arrows whereas a star indicates the absence of normally expected hybridizations. In particular,

Fig. 1 illustrates hybridization results using the 5170 bp HIRA probe described in Example 1, and wherein the probe was reacted with purified repetitive DNA sequences;

WO 01/88089

PCT/US01/15674

-6-

Fig. 2 illustrates a comparative hybridization identical to that depicted in Fig. 1, using the same 5170 bp HIRA probe but without pre-reaction with purified repetitive DNA sequences;

5 Fig. 3 illustrates hybridization results using the 3544 bp 15q11-q13 probe pre-reacted with purified repetitive DNA;

Fig. 4 illustrates results in a comparative experiment using the 3544 bp 15q11-q13 probe without pre-reaction with purified repetitive DNA;

10 Fig. 5 illustrates hybridization results using the 4166 bp, 3544 bp and 2290 bp 15q11-q13 probes described in Example 2, without pre-reaction with purified repetitive DNA sequences;

Fig. 6 illustrates hybridization results using the 5170 bp, 3691 bp, 3344 bp and 2848 bp HIRA probes described in Example 1 without pre-reaction with purified repetitive DNA sequences;

15 Fig. 7 illustrates hybridization results using the 4823 bp 1p36.3 probe described in Example 2 on metaphase cells of a normal individual, with pre-reaction with purified repetitive DNA sequences;

Fig. 8 illustrates a comparative hybridization result using the 4823 bp 1p36.3 probe of Fig. 7 without pre-reaction with purified repetitive DNA sequences;

20 Fig. 9 illustrates hybridization results using the 4724 bp and 4823 bp 1p36.3 probes described in Example 2 with pre-reaction with purified repetitive DNA sequences, and wherein single copy hybridizations were observed on homologous pairs of chromosome 1s;

Fig. 10 illustrates a comparative hybridization result using the 4724 bp and 4823 bp 1p36.3 probes described in Example 2 without pre-reaction with purified repetitive DNA sequences, and depicting the same single copy hybridizations shown in Fig. 9;

25 Fig. 11 illustrates hybridization results using the 4166 bp, 3544 bp and 2290 bp 15q11-q13 probes described in Example 2 without pre-reaction with purified DNA sequences on metaphase cells of a patient affected with Prader-Willi syndrome and known to harbor a deletion of 15q11-q13 sequences for one chromosomal allele, with a star indicating lack of hybridization at the deleted chromosome position and with the arrow indicating hybridization to a single chromosome;

WO 01/88089

PCT/US01/15674

-7-

Fig. 12 illustrates hybridization results using the 3691 bp, 3344 bp and 2848 bp HIRA probes described in Example 1 without pre-reaction with purified DNA sequences on metaphase cells of a patient affected with DiGeorge/Velo-Cardio-Facial Syndrome (VCFS) known to harbor a deletion of 22q11.2 sequences, wherein the star indicating lack of hybridization at the deleted chromosome position and the arrow indicating a normal homolog;

Fig. 13 is a scatterplot of base pair coordinates versus single copy probe lengths found in the Breakage Cluster Region gene (BCR) promoter found on chromosome 22, the disruption of which is common in cases of chronic adult myeloid leukemia and in some cases of acute lymphoblastic leukemia, as described in Example 4;

Fig. 14 is a scatterplot of base pair coordinates versus single copy probe lengths found in the ABL1 gene on chromosome 9, the disruption of which is common in cases of chronic adult myeloid leukemia and in some cases of acute lymphoblastic leukemia, as described in Example 4;

Fig. 15 is a CCD camera image of a FISH experiment using chromosome 9q34-specific, digoxigenin-dUTP labeled probes from the ABL1 oncogene (SEQ ID Nos. 520-525), and detected with rhodamine (red) conjugated antibody to digoxigenin with DAPI stained metaphase cells from a patient with chronic myelogenous leukemia (CML), illustrating the use of probes downstream of the site of fusion between ABL1 gene and BCR gene used to make a precise chromosomal breakpoint determination as explained in Example 4 wherein the derivative chromosome 22 and normal chromosome 9 are indicated; and

Fig. 16 is a CCD camera image of a FISH experiment using chromosome 9q34-specific, digoxigenin-dUTP labeled probes from the ABL1 oncogene (SEQ ID Nos. 516-525) and detected with rhodamine (red) conjugated antibody to digoxigenin, with DAPI stained metaphase cells from a patient with CML, illustrating the use of probes from each side of the site of fusion between BCR and ABL1 genes used to make a precise chromosomal 9q34 breakpoint determination as explained in Example 4, wherein the derivative chromosome 22, derivative chromosome 9 and normal chromosome 9 are indicated.

WO 01/88089

PCT/US01/15674

-8-

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The present invention is concerned with nucleic acid (e.g., DNA) hybridization probes useful for detection of genetic or neoplastic disorders. The probes are in the form of labeled nucleic acid fragments or a collection of labeled nucleic acid fragments whose hybridization to a target sequence can be detected. The invention also pertains to methods of developing, generating and labeling such probes, and to uses thereof.

The labeled probes hereof may be used with any nucleic acid target that may potentially contain repetitive sequences. These target sequences may include, but are not limited to chromosomal or purified nuclear DNA, heteronuclear RNA, or mRNA species that contain repetitive sequences as integral components of the transcript. In the ensuing detailed explanation, the usual case of a DNA target sequence and DNA probes is discussed; however, those skilled in the art will understand that the discussion is equally applicable (with art-recognized differences owing to the nature of the target sequences and probes) to other nucleic acid species.

An important characteristic of the probes of the invention is that they are composed of "single copy" or "unique" DNA sequences which are both complementary to at least a portion of the target DNA region of interest and are essentially free of sequences complementary to repeat sequences within the genome of which the target region is a part. Accordingly, a probe made up of a single copy or unique sequence is complementary to essentially only one sequence in the corresponding genome.

Very recently, it has been discovered that the human genome contains highly similar domains which have been termed duplicons when they are present in two non-allelic copies or triplicons when present in three copies in the genome (Ji et al., *Genome Res.*, 10:597-610 (2000)). Duplication or triplication of chromosomal domains containing such sequences were recent evolutionary events, based on the fact that non-human primates, in some instances, do not contain multiple copies of these sequences, and on the high degree of sequence similarity between different copies of paralogous sequences. These low copy duplicons (or triplicons) are to be distinguished from classic repetitive sequence families, which tend to either be interspersed throughout the genome or to be tandemly reiterated hundreds to thousands of times in the same chromosomal interval; therefore, probes from

WO 01/88089

PCT/US01/15674

-9-

duplicons or triplicons are for purposes of the present invention deemed to be within the ambit of "single copy" probes.

These duplicons or triplicons have evolved so recently that the sequence and organization of an entire genomic domain – which comprises complex, near-single copy segments and adjacent members of known repetitive sequence families – is completely conserved in each duplilon or triplicon segment. Duplilon and triplicon lengths of several kilobases to megabase sizes have been reported (International Genome Sequencing Consortium, *Nature*, 409:860-922 (2001)). The duplicons/triplicons are often tandemly arranged, and are almost always present on the same chromosome, and are therefore clustered in the genome. The length of the interval separating paralogous probe sequences is dictated by the size of the duplicated/triplicated domain, the orientation of the duplicons (or triplicons) relative to each other (i.e. direct or inverted), and the length of unrelated sequence intervals, if any, which separate the duplicons/triplicons.

In the context of the present invention, the term "single copy" with reference to a nucleic acid sequence refers to a sequence which is strictly unique (i.e., which is complementary to one and one only sequence in the corresponding genome) but also covers duplicons and triplicons. Stated otherwise, a "single copy" probe in preferred forms will hybridize to three or less locations in the genome.

As used herein, a "repeat sequence" is a sequence which repeatedly appears in the genome of which the target DNA is a part, with a sequence identity between repeats of at least about 60%, more preferably at least about 80%, and which is of sufficient length or has other qualities which would cause it to interfere with the desired specific hybridization of the probe to the target DNA (i.e., the probe would hybridize with multiple copies of the repeat sequence). Generally speaking, a repeat sequence appears at least about 10 times in the genome (more preferably at least about 50 times, and most preferably at least about 200 times) and has a length of at least about 50 nucleotides, and more preferably at least about 100 nucleotides. Repeat sequences can be of any variety (e.g., tandem, interspersed, palindromic or shared repetitive sequences with some copies in the target region and some elsewhere in the genome), and can appear near the centromeres of chromosomes, distributed over a single chromosome, or throughout some or all chromosomes. Normally, with but few exceptions, repeat sequences do not express physiologically useful proteins.

WO 01/88089

PCT/US01/15674

-10-

Repetitive sequences occur in multiple copies in the haploid genome. The number of copies can range from at least about 10 to hundreds of thousands, wherein the Alu family of repetitive DNA are exemplary of the latter numerous variety. The copies of a repeat may be clustered or interspersed throughout the genome. Repeats may be clustered in one or more locations in the genome, for example, repetitive sequences occurring near the centromeres of each chromosome, and variable number tandem repeats (VNTRs) (Nakamura et al., *Science*, 235:1616 (1987)); or the repeats may be distributed over a single chromosome for example, repeats found only on the X chromosome as described by Bardoni et al. (*Cytogenet. Cell Genet.*, 46:575 (1987)); or the repeats may be distributed over all the chromosomes, for example, the Alu family of repetitive sequences.

Simple repeats of low complexity can be found within genes but are more commonly found in non-coding genomic sequences. Such repeated elements consist of mono-, di-, tri-, tetra-, or penta-nucleotide core sequence elements arrayed in tandem units. Often the number of tandem units comprising these repeated sequences varies at the identical locations among genomes from different individuals. These repetitive elements can be found by searching for consecutive runs of the core sequence elements in genomic sequences.

As used herein, "sequence identity" refers to a relationship between two or more polynucleotide sequences, namely a reference sequence and a given sequence to be compared with the reference sequence. Sequence identity is determined by comparing the given sequence to the reference sequence after the sequences have been optimally aligned to produce the highest degree of sequence similarity, as determined by the match between strings of such sequences. Upon such alignment, sequence identity is ascertained on a position-by-position basis, e.g., the sequences are "identical" at a particular position if at that position, the nucleotides are identical. The total number of such position identities is then divided by the total number of nucleotides or residues in the reference sequence to give % sequence identity. Sequence identity can be readily calculated by known methods, including but not limited to, those described in Computational Molecular Biology, Lesk A. N., ed., Oxford University Press, New York (1988); Biocomputing: Informatics and Genome Projects, Smith D.W., ed., Academic Press, New York (1993); Computer Analysis of Sequence Data, Part I, Griffin A.M., and Griffin H. G., eds., Humana Press, New Jersey (1994); Sequence Analysis in Molecular Biology, von Heinge G., Academic Press (1987);

WO 01/88089

PCT/US01/15674

-11-

Sequence Analysis Primer, Gribskov M. and Devereux J., eds., M. Stockton Press, New York (1991); and Carillo H., and Lipman D., SIAM J. Applied Math., 48:1073 (1988). Preferred methods to determine the sequence identity are designed to give the largest match between the sequences tested. Methods to determine sequence identity are codified in publicly available computer programs which determine sequence identity between given sequences. Examples of such programs include, but are not limited to, the GCG program package (Devereux et al., *Nuc. Ac. Res.*, 12(1):387 (1984)), BLASTP, BLASTN and FASTA (Altschul et al., *J. Molec. Biol.*, 215:403-410 (1990)). The BLASTX program is publicly available from NCBI and other sources (BLAST Manual, Altschul et al., NCBI, NLM, NIH, Bethesda, MD 20894; Altschul et al., *J. Molec. Biol.*, 215:403-410 (1990)). These programs optimally align sequences using default gap weights in order to produce the highest level of sequence identity between the given and reference sequences. As an illustration, by a polynucleotide having a nucleotide sequence having at least, for example, 95% "sequence identity" to a reference nucleotide sequence, it is intended that the nucleotide sequence of the given polynucleotide is identical to the reference sequence except that the given polynucleotide sequence may include up to 5 differences per each 100 nucleotides of the reference nucleotide sequence. In other words, in a polynucleotide having a nucleotide sequence having at least 95% identity relative to the reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. Inversions in either sequence are detected by these computer programs based on the similarity of the reference sequence to the antisense strand of the homologous test sequence. These variants of the reference sequence may occur at the 5' or 3' terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

The single copy probes of the invention preferably should have a length of at least about 50 nucleotides, and more preferably at least about 100 nucleotides. Probes of this length are sufficient for Southern blot analyses. However, if other analyses such as FISH are employed, the probes should be somewhat longer, i.e., at least about 500 nucleotides, and

WO 01/88089

PCT/US01/15674

-12-

more preferably at least about 2000 nucleotides in length. The probes can be used to detect virtually any type of chromosomal rearrangement, such as deletions, duplications, insertions, additions, inversions or translocations.

5 In order to develop probes in accordance with the invention, the sequence of the target DNA region must be known. The target region may be an entire chromosome or only portions thereof where rearrangements have been identified. With this sequence knowledge, the objective is to determine the boundaries of single copy or unique sequences within the target region. This is preferably accomplished by inference from the locations of repetitive sequences within the target region. Normally, the sequence of the target region is compared 10 with known repeat sequences from the corresponding genome, using available computer software. Once the repeat sequences within the target region are identified, the intervening sequences are deduced to be single copy (i.e., the sequences between adjacent repeat sequences).

15 Optimal alignment of the target and repetitive sequences for comparison may be conducted by the local homology algorithm of Smith et al., *Adv. Appl. Math.*, **2**:482 (1981), by the homology alignment algorithm of Needleman et al., *J. Mol. Biol.*, **48**:443 (1970). The results obtained from the heuristic methods (Pearson et al., *Proc. Natl. Acad. Sci.*, **85**:244 (1988); Altschul et al., *J. Molec. Biol.*, **215**:403-410 (1990)) are generally not as comprehensive as the methods of Smith et al. (1981) and Needleman et al. (1970). However, they are 20 faster than these methods.

25 Once the single copy sequence information is obtained, certain of the single copy sequences (normally the longest) are used to design hybridization probes. In this regard, probes may be of varying "complexity" as defined by Britten et al., *Methods of Enzymol.*, **29**:363 (1974) and as further explained by Cantor et al., *Biophysical Chemistry: Part III: The Behavior of Biological Macromolecules*, pp. 1228-1230. The complexity of selected probes is dependent upon the application for which it is designed. In general, the larger the target area, the more complex the probe. The complexity of a probe needed to detect a set 25 of sequences will decrease as hybridization sensitivity increases. At high sensitivity and low background, smaller and less complex probes can be used.

30 With current hybridization techniques, it is possible to obtain reliable, easily detectable signals with relatively small probes in accordance with the invention. A readily

WO 01/88089

PCT/US01/15674

-13-

detectable signal was obtained with a probe on the order of 2 kb in length, using FISH technology. This sensitivity of the present method is improved compared to the prior art (U.S. Patent No. 5,756,696) because the probes of the present invention are homogeneous single copy sequences. However, smaller amplified segments, each comprising non-repetitive sequences, may also be used in combination as probes to achieve adequate signals for *in situ* hybridization. Complex single copy probes that hybridize to duplicated or triplicated targets can also increase hybridization signals.

One application of the use of multiple fragment probes is in the detection of translocations between different chromosomes. Proportionately increasing the complexity of the probe also permits analysis of multiple compact regions of the genome simultaneously. For a single chromosome, the portion of the probe targeted to one side of the breakpoint can be labeled and detected differently from that targeted to the other side of the breakpoint so that the derivative or translocated chromosome is detected by one label and is distinguishable from the intact normal chromosome which has both labels.

The invention makes it possible to produce single copy probes at a higher genomic density than possible using conventional probes. Chromosomes 21 and 22 have been comprehensively sequenced, and it has been determined that adjacent single copy intervals tend to be clustered on these chromosomes. For example, on chromosome 22, 39% of single copy intervals are separated by only 500-1000 bp. Single copy intervals ≥ 2.3 kb are separated, on average, by 29.2 kb on chromosome 21 and by 22.3 kb on chromosome 22.

In order to estimate the size of genomic intervals required to develop single copy probes, the probability of detecting at least one single copy sequence in overlapping, uniform-length genomic intervals on chromosomes 21q and 22q was determined. Single copy segments ≥ 2.0 kb in length are found in the majority of 100 kb genomic intervals of these chromosomes (96% of chromosome 22q and 88% of chromosome 21q). Increasing the size of the genomic sequence to 150 kb results in 99% coverage of chromosome 22q and 96% of chromosome 21q. Therefore, single copy probes should be more or less 2 kb to ensure comprehensive coverage (at least once per 100-150 kb) of chromosomes 21 and 22. Assuming that single copy sequences are similarly distributed on other chromosomes, it should be feasible to develop probes for *in situ* hybridization analysis of most clinically relevant chromosomal rearrangements.

WO 01/88089

PCT/US01/15674

-14-

Once appropriate single copy sequences in the chromosomal region of interest have been identified, PCR is preferably used for amplifying the appropriate DNA to obtain probes. PCR is a well known technique for amplifying specific DNA segments in geometric progression and relies upon repeated cycles of DNA polymerase-catalyzed extension from a pair of oligonucleotide primers with homology to the 5' end and to the complement of the 3' end of the DNA segment to be amplified.

The nucleic acid (e.g., DNA) that serves as the PCR template may be single stranded or double stranded, but when the DNA is single stranded, it will typically be converted to double stranded. The length of the template DNA may be as short as 50 bp, but usually will be at least about 100 bp long, and more usually at least about 150 bp long, and may be as long as 10,000 bp or longer, but will usually not exceed 50,000 bp in length, and more usually will not exceed 20,000 bp in length. The DNA may be free in solution, flanked at one or both ends with non-template DNA, present in a cloning vector such as a plasmid and the like, with the only criteria being that the DNA be available for participation in the primer extension reaction. The template DNA may be derived from a variety of different sources, so long as it is complementary to the target chromosomal or immobilized DNA sequence. The amount of template DNA that is combined with the other reagents will range from about 1 molecule to 1 pmol, usually from about 50 molecules to 0.1 pmol, and more usually from about 0.01 pmol to 100 fmol. The oligonucleotide primers with which the template nucleic acid is contacted will be of sufficient length to provide for hybridization to complementary template DNA under annealing conditions but will be of insufficient length to form stable hybrids with template DNA under polymerization conditions. The primers will generally be at least about 10 nucleotides (nt) in length, usually at least 15 nt in length and more usually at least 16 nt in length and may be as long as 30 nt in length or longer, where the length of the primers will generally range from 18 to 50 nt in length, usually from about 20 to 35 nt in length. The yield of longer amplification products can be enhanced using primers of 30 to 35 nt and high fidelity polymerases (described in U.S. Patent No. 5,436,149).

To maximize the signal intensity obtained during *in situ* hybridization, primer sequence pairs are preferred which, upon amplification, produce a DNA fragment that spans nearly the entire length of each single-copy genomic sequence interval. Hence, contiguous or closely spaced (software excludes pairs that are separated by $\leq 70\%$ of the length of the

WO 01/88089

PCT/US01/15674

-15-

single copy interval) primer pairs are generally excluded from consideration for producing probes for *in situ* hybridization. With the exception of cytogenetic preparations, this criterion is generally not applicable for probes that are hybridized to immobilized cloned or synthetic nucleic acid targets, since signal intensities of shorter probes are usually adequate due to the increased number of target molecules.

However, in order to optimize the yield and kinetics of the PCR reaction, the desired primer sequences are also subject to other criteria. First, a primer sequence should not be substantially self-complementary or complementary to the second primer. In particular, potential primer sequences are excluded which could result in the formation of stable hybrids involving the 3' terminus of the primer and either another sequence in the same or the second primer (defined as ≥ 6 base pairs). Additionally, the T_m of one member of the primer pair should occur within 2°C of its counterpart, which enables them to denature and anneal to the template nearly simultaneously. Software is well known in the art to identify primer sequences that satisfy all of the preferred criteria (see for example: <http://www-genome.wi.mit.edu/ftp/pub/software/primer.0.5/> or http://www.oligo.net/Oligo_6_tour.htm).

The PCR reaction mixture will normally further comprise an aqueous buffer medium which includes a source of monovalent ions, a source of divalent cations and a buffering agent. Any convenient source of monovalent ions, such as KCl, K-acetate, NH₄-acetate, K-glutamate, NH₄ Cl, ammonium sulfate, and the like may be employed, where the amount of monovalent ion source present in the buffer will typically be present in an amount sufficient to provide for a conductivity in a range from about 500 to 20,000, usually from about 1000 to 10,000, and more usually from about 3,000 to 6,000 microohms. The divalent cation may be magnesium, manganese, zinc and the like, where the cation will typically be magnesium. Any convenient source of magnesium cation may be employed, including MgCl₂, Mg-acetate, and the like. The amount of Mg²⁺ present in the buffer may range from 0.5 to 10 mM, but will preferably range from about 2 to 4 mM, more preferably from about 2.25 to 2.75 mM and will ideally be at about 2.45 mM. Representative buffering agents or salts that may be present in the buffer include Tris, Tricine, HEPES, MOPS and the like, where the amount of buffering agent will typically range from about 5 to 150 mM, usually from about 10 to 100 mM, and more usually from about 20 to 50 mM, where in certain preferred

WO 01/88089

PCT/US01/15674

-16-

embodiments the buffering agent will be present in an amount sufficient to provide a pH ranging from about 6.0 to 9.5, where most preferred is pH 7.3 at 72°C. Other agents which may be present in the buffer medium include chelating agents, such as EDTA, EGTA and the like.

5 Also present in the PCR reaction mixtures is a melting point reducing agent, i.e., a reagent that lowers the melting point of DNA. Suitable melting point reducing agents are those agents that interfere with the hydrogen bonding interaction of two nucleotides, where representative base pair destabilization agents include: betaine, formamide, urea, thiourea, acetamide, methylurea, glycinamide, and the like, where betaine is a preferred agent. The melting point reducing agent will typically be present in amounts ranging from about 20 to 500 mM, usually from about 50 to 200 mM and more usually from about 80 to 150 mM.

10 In preparing the PCR reaction mixture, the various constituent components may be combined in any convenient order. For example, the buffer may be combined with primer, polymerase and then template DNA, or all of the various constituent components may be combined at the same time to produce the reaction mixture.

15 Following preparation of the PCR reaction mixture, it is subjected to a plurality of reaction cycles, where each reaction cycle comprises: (1) a denaturation step, (2) an annealing step, and (3) a polymerization step. The number of reaction cycles will vary depending on the application being performed, but will usually be at least 15, more usually at least 20 and may be as high as 60 or higher, where the number of different cycles will typically range from about 20 to 40. For methods where more than about 25, usually more than about 30 cycles are performed, it may be convenient or desirable to introduce additional polymerase into the reaction mixture such that conditions suitable for enzymatic primer extension are maintained.

20 25 The denaturation step comprises heating the reaction mixture to an elevated temperature and maintaining the mixture at the elevated temperature for a period of time sufficient for any double stranded or hybridized nucleic acid present in the reaction mixture to dissociate. For denaturation, the temperature of the reaction mixture will usually be raised to, and maintained at, a temperature ranging from about 85 to 100°C usually from about 90 to 98°C, and more usually from about 93 to 96°C for a period of time ranging from about 3 to 120 seconds, usually from about 5 to 30 seconds.

WO 01/88089

PCT/US01/15674

-17-

Following denaturation, the PCR reaction mixture will be subjected to conditions sufficient for primer annealing to template DNA present in the mixture. The temperature to which the reaction mixture is lowered to achieve these conditions will usually be chosen to provide optimal efficiency and specificity, and will generally range from about 50 to 5

75°C, usually from about 55 to 70°C and more usually from about 60 to 68°C. Annealing conditions will be maintained for a period of time ranging from about 15 seconds to 30 minutes, usually from about 30 seconds to 5 minutes.

Following annealing of primer to template DNA or during annealing of primer to template DNA, the reaction mixture will be subjected to conditions sufficient to provide for polymerization of nucleotides to the primer ends in manner such that the primer is extended in a 5' to 3' direction using the DNA to which it is hybridized as a template, i.e. conditions sufficient for enzymatic production of primer extension product. To achieve polymerization conditions, the temperature of the reaction mixture will typically be raised to or maintained at a temperature ranging from about 65 to 75°C, usually from about 67 to 73°C and maintained for a period of time ranging from about 15 seconds to 20 minutes, usually from 15 about 30 seconds to 5 minutes.

The above cycles of denaturation, annealing and polymerization may be performed using an automated device, typically known as a thermal cycler. Thermal cyclers that may be employed are described in U.S. Patent Nos. 5,612,473; 5,602,756; 5,538,871; and 20 5,475,610.

Based on all the previous criteria, a series of primers were produced and validated by PCR using genomic DNA from normal individuals. Knowledge of suitable primers will necessarily define the corresponding PCR-produced probes in accordance with the invention. Thus, adjacent pairs of sequences identified as SEQ ID Nos. 429-446 and 480-613, 25 beginning with SEQ ID No. 429, are respective forward/reverse PCR primers developed for the production of specific useful probes. Hence, a useful probe may be produced using a combination of SEQ ID Nos. 429 and 430, and additional probes are defined by the succeeding pairs of adjacent SEQ IDs. Broadly speaking, certain preferred probes of the invention should have at least about 80% sequence identity, and more preferably about 90% sequence identity, relative to the probes defined by the above-described adjacent pairs of 30 primer sequences.

WO 01/88089

PCT/US01/15674

-18-

In addition to the PCR, DNA fragments corresponding to unique sequences can also be obtained by a variety of other methods, including but not limited to deletion mutagenesis, restriction digestion, direct synthesis and DNA ligation.

If the genomic fragment is obtained by amplification or purification from DNA containing repetitive sequences, the fragment must then be purified prior to labeling and hybridization. Purification of homogeneously-sized DNA fragments can be accomplished by a variety of methods, including but not limited to electrophoresis and high pressure liquid chromatography. In the preferred method, amplified fragments are separated according to size by gel electrophoreses in Seakem LE Agarose using Tris Acetate buffer (Sambrook, Fritsch & Maniatis, Molecular Cloning: A Laboratory Manual [Cold Spring Harbor Laboratory Press, 1989]), stained with a dye such as ethidium bromide or Syber-Green, visualized with ultraviolet light (300 nm), excised from the gel using a scalpel. Each DNA fragment is then recovered from the gel fragment using a Micro-con 100 (Millipore, Watertown, MA) column by spin centrifugation.

Phenol-chloroform extraction of the amplified DNA is not an adequate method of purification. When this approach was tested, this purification technique resulted in nonspecific hybridization to all chromosomes along their entire length, which is consistent with the pattern produced by hybridization of repetitive sequences (data not shown). This occurs because, during the PCR process, DNA polymerase extends the replicated strand past the position of the second primer into adjacent repetitive sequences if the initial template contains genomic DNA sequences. These extension products which are longer than the amplification product, are present in all such PCR reactions. Since, in the present method, repetitive sequences are adjacent to the segments being amplified, the extension products are likely to contain such sequences. Phenol-chloroform extraction of PCR reactions does not remove such extension products. PCR reaction mixtures containing these sequences may hybridize to repetitive genomic DNA in addition to the target sequence. Hence, isolation of the purified genomic amplification fragment (whether it is obtained directly from genomic DNA or by PCR), is a preferred embodiment of the subject invention and would not be obvious to one skilled in the art.

Insertion of the purified fragments into plasmids, bacteriophages, or artificial chromosome cloning vehicles capable of being propagated in *E. coli*, yeast, or other species

WO 01/88089

PCT/US01/15674

-19-

may be desirable to reduce the cost and labor required for repeated preparation of single copy DNA probes. A variety of cloning vectors have been optimized for rapid ligation and selection for vectors containing PCR products (for example: U.S. Patent Nos. 5,487,993 and 5,766,891). If the probe will be used in multiple hybridizations, then the cloned recombinant form will be less expensive to produce in large quantities than by iterative PCR amplification from the same genomic DNA template. In addition, genomic insert in the cloned probe does not have to be isolated during purification, since the fragment recombined with vector is propagated in the absence of any other genomic DNA that could potentially contain repetitive sequences. Finally, the cloned vehicle provides a potentially inexhaustible source of probe, whereas natural genomic DNA templates may have to be reisolated from cell lines or from other sources. Single copy DNA fragments obtained by PCR amplification as described above are isolated according to size by gel electrophoresis and purified by columns as is well known in the art.

These fragments are then labeled with nonisotopic identifying label such as a fluorophore, an enzymatic conjugate, or one selected from the group consisting of biotin or other moieties recognized by avidin, streptavidin, or specific antibodies. There are several types of non-isotopic identifying labels. One type is a label which is chemically bound to the probe and serves as the means for identification and localization directly. An example of this type would be a fluorochrome moiety which upon application of radiation of proper wavelengths will become excited into a high energy state and emit fluorescent light. The probes can be synthesized chemically or preferably be prepared using the methods of nick-translation (Rigby et al., *J. Mol. Biol.*, 113:237-251, (1977)) or Klenow labeling (Feinberg et al., *Anal. Biochem.*, 137:266-267, (1984)) in the conventional manner using a reactant comprising the identifying label of choice (but not limited to) conjugated to a nucleotide such as dATP or dUTP. The fragments are either directly labeled with a fluorophore-tagged nucleotide or indirectly labeled by binding the labeled duplex to a fluorescently-labeled antibody that recognizes the modified nucleotide that is incorporated into the fragment as described below. Nick-translations (100 μ l reaction) utilize endonuclease-free DNA polymerase I (Roche Molecular Biochemicals, Indianapolis, IN) and DNase I (Worthington Biochemical Corporation, Lakewood, NJ). Each fragment is combined with DNA polymerase (20 units/microgram DNA), DNase I (10 microgram/100 μ l reaction), labeled

WO 01/88089

PCT/US01/15674

-20-

nucleotide (0.05 mm final) and nick translation buffer. The reaction is performed at 15°C for 45 minutes to 2 hours and yields a variety of labeled probe fragments of different nucleotide sizes in the 100 to 500 bp size range.

Other methods for labeling and detecting probes in common use may be applied to the single-copy DNA probes produced by the present method. These include: fluorochrome labels (which resolve labeling on individual chromatids which serves as an affirmation that hybridization occurred unequivocably, and further allows detection precisely at site of hybridization rather than at some distance away), chemical reagents which yields an identifiable change when combined with the proper reactants (for example, alkaline phosphatase, horseradish peroxidase and galactosidase, each of which reacts and provide a detectable color change that identifies the presence and position of the target sequence), and indirect linkage mechanism of specifically binding entities (such as the biotin-avidin system in which the probe is preferably joined to biotin by conventional methods and added to an avidin- or streptavidin-conjugated fluorochrome or enzyme which provides the specificity for attaching the fluorochrome or enzyme to the probe).

It will be recognized that other identifying labels may also be used with the described probes. These include fluorescent compositions such as energy transfer groups, conjugated proteins, antibodies and antigens, or radioactive isotopes.

Chromosomal hybridization and detection are a preferred use of DNA probes generated by the present invention. DNA probes generated by the present method may be hybridized either directly to complementary nucleic acids in cells (*in situ* hybridization) or to nucleic acids immobilized on a substrate. A preferred use of the method is *in situ* hybridization, which is well known in the art, being described in U.S. Patent Nos. 5,985,549; 5,447,841; 5,756,696; 5,869,237. Based on early work of Gall and Pardue (*Proc. Natl. Acad. Sci.*, 63:378-383, 1969), isotopic *in situ* hybridization was established in the 1970s (see Gerhard et al., *Proc. Natl. Acad. Sci.*, 78:3755-3759, 1981 and Harper et al., *Proc. Natl. Acad. Sci.*, 78:4458-4460, 1981 as examples) and subsequently nonisotopic *in situ* hybridization was established. The technique of nonisotopic *in situ* hybridization is reviewed and a protocol is provided for use in chromosomal hybridization by Knoll and Lichter, in *Current Protocols in Human Genetics*, Vol. 1, Unit 4.3 (Green-Wiley, New York, 1994) and in U.S. Patent No. 5,985,549. The technique relies on the formation of duplex

WO 01/88089

PCT/US01/15674

-21-

nucleic acid species, in which one strand is derived from a labeled probe molecule and the second strand comprises the target to be detected. Target molecules may comprise chromosomes or cellular nucleic acids. Numerous methods have been developed to label the probe and visualize the duplex.

5 The method of the present invention is intended to be used with any nucleic acid target containing repetitive sequences. The sample containing the target nucleic acid sequence can be prepared from cellular nuclei, morphologically intact cells (or tissues), chromosomes, other cellular material components, or synthetically produced nucleic acids. The samples may be obtained from the fluids or tissues of a mammal, preferably human, 10 which are suspected of being afflicted with a disease or disorder either from a biopsy or post-mortem, or from plants.

As an example, chromosomal preparations can be made in the following manner: phytohemagglutinin-stimulated peripheral lymphocytes are cultured in RPMI 1640 medium containing 10% fetal calf serum for 72 hours at 37°C. Ethidium bromide (100ug/10ml final) 15 is added 1-1/2 hours prior to harvest. Colcemid (1ug/10ml final) is added during the final 20 min of incubation with ethidium bromide. The cells are then pelleted by centrifugation and incubated preferably in 0.075 M KCl at 37°C for about 20 minutes. Cells are then pelleted again and fixed in 3 changes of Carnoy's fixative (3:1 methanol: acetic acid volumetric ratio) using conventional cytogenetic techniques. [For a review of chromosome preparation from peripheral blood cells, see Bangs and Donlon in Dracopoli et al., eds., 20 *Current Protocols in Human Genetics*, Vol. 1, Unit 4.1 (Green-Wiley, New York, 1994)]. The nuclei or cells in suspension can then be dropped onto clear glass coverslips or microscope slides in a humid environment to promote chromosome spreading. The coverslips or microscope slides can then be preferentially air dried overnight, aged or stored until required for use in *in situ* hybridization.

25 Immediately prior to chromosomal denaturation in the *in situ* hybridization procedure, the dried or stored chromosome preparations can be pretreated in prewarmed 2 x SSC (components are in Sambrook, Fritsch & Maniatis, *Molecular Cloning: A Laboratory Manual* [Cold Spring Harbor Laboratory Press, 1989]) for 30 minutes at 37°C followed by dehydration in an ethanol series (2 minutes each in 70%, 80%, 90% and 100% ethanol).

WO 01/88089

PCT/US01/15674

-22-

When the target nucleic acid sequence is DNA, DNA in the sample can be denatured by heat or alkali. [See Harper et al., *Proc. Natl. Acad. Sci.*, **78**:4458-60 (1981), for alkali denaturation and Gall et al., *Proc. Natl. Acad. Sci.*, **63**:370-383 (1969)]. Denaturation is carried out so that the DNA strands are separated with minimal shearing, degradation or oxidation.

In the preferred current method, the labeled single copy probe is resuspended in deionized formamide and denatured at 70-75°C. The chromosomal template is denatured in a solution containing 70% formamide/2 x SSC, pH 7.0 followed by dehydration in an ethanol series (2 minutes each in cold 70% ethanol and room temperature 80, 90 and 100% ethanol). Hybridization of the labeled probe to the corresponding template is carried out in a solution containing 50% formamide/2 x SSC/10% dextran sulfate/BSA [bovine serum albumin; 1 mg/ml final] for a few hours to overnight. The length of time depending on the complexity of the probe that is utilized. After hybridization, non-hybridizing excess probe is removed by a washing procedure. The duplexes are treated with a series of 15-30 minute washes: first with a solution of 50% formamide/2xSSC at 39-45°C, then 2 x SSC at 39-45°C, followed by a 15-30 minute wash at room temperature in 1 x SSC. The hybridized sequences are detected by relevant means. For example, digoxigenin-dUTP can be but is not limited to detection by an antibody to digoxigenin such as rhodamine or fluorescein conjugated antibody (Roche Molecular Biochemicals, Indianapolis, IN). Following detection, spurious detection reagents are removed by washing in varying SSC and SSC/triton-X concentrations, the chromosomes are counterstained with a dye such as DAPI and the hybridized preparation is mounted in an antifade solution such as Vectashield (Vector Laboratories, Burlingame, CA). The cells are examined by fluorescence microscopy with the appropriate filter sets and imaged with a charge coupled device (CCD).

An important aspect of the present invention is that the probe or target DNA does not require pre-reaction with a non-specific nucleic acid competitor such as purified repetitive DNA or that the probe does not require experimental verification that the single copy fragments or recombinant cloned probes do not contain repetitive sequences (U.S. Patent Nos. 5,985,549; 5,447,841; 5,663,319; 5,756,696) because the probes are single copies without repetitive elements. This results in a significantly improved signal to noise ratio. A signal to noise ratio is defined as a ratio of the probability of the probe detecting a *bona*

WO 01/88089

PCT/US01/15674

-23-

fide signal of hybridization of the target nucleic acid sequence to that of the probability of detecting the background caused by non-specific binding of the labeled probe.

The hybridization reactions carried out using the probes of the invention are themselves essentially conventional. As indicated, two exemplary types of hybridizations are the Southern blot and FISH techniques, well known to those skilled in the art. However, the visual patterns resulting from use of the probes, termed indicator patterns, are extremely useful tools for cytogenetic analyses, especially molecular cytogenetic analyses. These indicator patterns facilitate microscopic and/or flow cytometric identification of normal and abnormal chromosomes and characterization of the genetic abnormalities. Since multiple compatible methods of probe visualization are available, the binding patterns of different components of the probes can be distinguished, for example, by color. Thus, the invention is capable of producing virtually any desired indicator pattern on the chromosomes visualized with one or more colors (a multi-color indicator pattern) and/or other indicator methods.

Preferred indicator patterns derived from using the probes of the invention comprise one or more "bands," meaning a reference point in a genome comprising a target DNA sequence with a probe bound thereto, and wherein the resulting duplex is detectable by some indicator. Depending on hybridization washes and the detection conditions, a band can extend from the narrow context of a sequence providing a reliable signal to a single chromosome region to multiple regions on single or plural chromosomes. The indicator bands from the probes hereof are to be distinguished from bands produced by pretreatment and chemical staining. The probe-produced bands of the present invention are based upon the complementarity of the DNA sequences, whereas bands produced by chemical staining depend upon natural characteristics of the chromosomes (such as structure or protein composition), but not by hybridization to the DNA sequences thereof. Furthermore, chemical staining techniques are useful only in connection with metaphase chromosomes, whereas the probe-produced bands of the present invention are useful for both metaphase and interphase chromosomes.

The following examples set forth the preferred techniques employed for the development, generation, labeling and use of specific DNA probes designed to hybridize to a target DNA sequence in a genome. It is to be understood, however, that these examples

WO 01/88089

PCT/US01/15674

-24-

are provided by way of illustration and nothing therein should be taken as a limitation upon the overall scope of the invention.

Example I

5 Development of HIRA Gene Probe

A known genetic disorder on human chromosome 22 involves a deletion of one HIRA gene in chromosome band 22q11.2, i.e., in normal individuals; there are two copies of the HIRA gene, whereas in affected individuals, only one copy is present. This deletion is considered to be a cause of haploinsufficiency syndromes such as DiGeorge and Velo-Cardio-Facial Syndromes (VCFS), because insufficient amounts of gene product(s) may disrupt normal embryonic development (Fibison et al., *Amer. J. Hum. Genet.*, **46**:888-95 (1990); Consevage et al., *Amer. J. Cardiol.*, **77**:1023-1205 (1996)). Other syndromes including Cat Eye Syndrome and derivative chromosome 22 syndrome result from an excess of genomic sequences from this region (Mears et al., *Amer. J. Hum. Genet.*, **55**:134-142 (1994); Knoll et al., *Amer. J. Med. Genet.*, **55**:221-224 (1995)). Typically individuals with these syndromes have supernumerary derivative chromosome 22s.

Initially, a computer-based search using the search term "HIRA" was performed using Entrez Nucleotide software at the National Library of Medicine website. This identified a series of cDNA sequences for the HIRA gene in GenBank. The full length cDNA sequence was selected (GenBank Accession No. X81844), having 3859 bp. This cDNA sequence was then compared with the genome sequence which included draft sequences at the National Library of Medicine (<http://www.ncbi.nlm.nih.gov/genome/se-q/page.cgi?F=HsBlast.html&&ORG=Hs>). This was done in order to determine whether genomic sequences of sufficient length were available for probe development. This comparison confirmed that the entire HIRA genomic sequence was known, and that the coding sequence interval spanned a length of 100,836 bp in the chromosome. Since the available contiguous genomic sequence in GenBank exceeded the length of the coding interval, it was possible to select an interval longer than the coding region in order to include sequences from the gene promoter at the 5' end and untranslated sequences and polyadenylation signal at the 3' end. A total genomic interval of approximately 103 kb was thus

WO 01/88089

PCT/US01/15674

-25-

selected. Position 1 of this ~103 kb interval corresponds to position 798,334 in GenBank Accession number NT_001039.

In the next step, the selected 103 kb genomic interval was compared with known high-complexity repeat sequence family members or consensus sequences that are aligned with the test genomic sequences (SEQ ID Nos. 1-428) and all combinations of low-complexity tandem repeat sequences of at least 17 nucleotides in length (mono-, di-, tri-, and tetranucleotide units) known to be present in the human genome (SEQ ID Nos. 447-479). This comparison was done using the publicly available CENSOR program which can be found at the Genetic Information Research Institute website, www.girinst.org. This program utilizes the Smith-Waterman global alignment comparison algorithm to determine the locations and distribution of repeat sequences within the genomic interval. A Smith-Waterman alignment of repetitive with genomic sequences was performed with the following parameters: Length of margin sequence: 50 nt, minimum length to extract insertion: 12 nt, minimum margin to combine matching fragments: 30, similarity threshold: 22, similarity threshold to always keep match: 35, ratio threshold: 2.8, relative similarity threshold: 2.8, gap constant D1: 2.95, gap constant D2: 1.90, and mismatch penalty: -1.0. This analysis generated the following table, which details the coordinates of repetitive sequence family members found in and adjacent to the human HIRA gene coding sequence.

20

Table 1

	Position (bp) in Seq Listing					
	HIRA position (bp)			Corresponding to HIRA Match		
	Begin	End	Repeat Family	SEQ ID NO.	Begin	End
25	798411	798434	(AC)	452	1	24
	798983	799395	MLT2A1	444	1	434
	801257	801348	CHESHIRE_A	420	132	223
	801747	801723	LIMF_ORF2	425	757	1089
	801746	802322	Alu-Y	2	2	299
	802032	802306	LIMF_ORF2	425	1000	1380
30	802355	802434	LIMD6_5	77	1629	1710
	802448	802798	LIMD5_5	66	996	1348
	802811	803100	Alu-Y	2	1	290
	803104	803189	LIMD3_5	66	907	995
	803199	803454	Alu-JD	2	5	290
	803472	803545	Alu-Spnz	2	2	76
35	803548	804061	LIMEC_5	345	1860	2392
	804079	804365	Alu-Sz	2	6	290
	804476	804559	LIP_MA2	348	6242	6321
	804625	804885	LIME_ORF2	425	2287	2568
	804936	804997	MLTIE2	106	198	260
	805011	805077	MLTIE1	105	420	484

WO 01/88089

PCT/US01/15674

-26-

	Begin	End	Repeat Family	SEQ ID NO.	Begin	End
5	805211	805211	L1PB _A _5	359	103	204
	805212	805262	L1PB _A _5	359	1089	1738
	805933	805989	Alu-J	2	234	290
	805950	806489	L1PB _A _5	359	1749	2247
	806210	806242	Alu-Sz	59	1659	1773
	806628	806917	Alu-Sz	2	1	290
	806919	807254	L1M _Z _5	61	2377	2716
	807301	808176	L1P _M A2	348	3516	4425
	808179	808469	Alu-Sz	2	1	290
10	808476	808734	L1	59	3268	3525
	808733	809426	L1MF_ORF2	425	1411	2105
	809429	809860	L1P _M A2	348	5607	6044
	809861	809993	Alu-Jb	2	2	134
	809996	810282	Alu-Jb	2	2	290
	810343	811040	L1	59	4711	5402
	811041	811221	L1PB3	358	151	333
15	811226	811513	Alu-Sx	2	1	287
	811515	812032	L1PB1	357	330	863
	812036	812394	Alu-Jb	2	1	288
	812471	812646	Alu-S	2	5	290
	812721	812836	Alu-Jb	2	2	117
	812862	812901	L1P _M A2	348	4315	4354
20	812903	813078	L1	59	3038	3222
	813079	814102	L1ME _C _ORF2	425	1113	2166
	814323	814410	MER1B	315	242	337
	814411	814557	CHARLIE3	7	1	281
	814780	814916	L1MB7	78	9	143
	815061	815181	Alu-Y	2	1	134
25	815420	815452	L1R67	279	99	131
	816487	816772	Alu-Sx	2	5	290
	817180	817270	L1MC _C _5	335	1384	1473
	817332	817620	Alu-Sg	2	1	290
	817634	818399	Alu-Sq	2	1	288
	817853	818372	Alu-S	2	2	289
30	818368	818578	HAL1	18	1346	1347
	818631	818791	L1NE2	362	2289	2464
	818824	818889	Alu-S	2	223	290
	818890	819185	L1NE2	362	2465	2749
	819328	819450	L1NE2	362	1925	2049
	819565	819757	L1NE2	362	2273	2498
35	823604	823892	Alu-Sx	2	2	290
	826836	827042	Alu-Sx ₂	2	84	290
	827922	827977	MIR	99	105	160
	830762	831371	L1ME _C _5	345	1498	2123
	831396	831683	Alu-S	2	2	290
	831687	831774	L1ME _C _5	345	2117	2205
40	831778	832066	Alu-Sx	2	1	290
	832155	832451	L1MC2	2	5	134
	832442	832735	Alu-Sx	79	666	786
	832742	832992	L1MC2	79	787	1077
	833004	833170	L1ME _C _ORF2	425	172	340
	833177	834590	TIGGER1	148	1	1477
45	834592	834642	Alu-Jb	2	156	207
	834799	834877	Alu-Jb	2	208	290
	834907	835194	Alu-Y	2	1	289
	835198	835590	TIGGER1	148	1468	1900
	835597	835888	Alu-Sx	2	1	290
	835946	835979	L1P _M A2	348	4654	4689
50	836060	836177	MER2	316	229	345
	836168	836486	Alu-Sx	2	7	290
	836497	836712	MER2	316	1	228
	836498	836713	MER2	2	1	288
	838822	839069	Alu-Sx	2	1	288
	839086	839373	Alu-Sz	2	1	289
55	840297	840826	L1MB7	78	269	915
	841062	841306	L1MB7	78	7	249
	841323	841382	L1MF_ORF2	425	3053	3116

WO 01/88089

PCT/US01/15674

-27-

	Begin	End	Repeat Family	SEQ ID NO.	Begin	End
	841408	841697	Alu-Sq	2	1	290
	841705	841828	Alu-Jo	2	1	136
	841829	842012	L1ME_ORF2	425	2870	3052
5	842744	842871	MER86	239	51	183
	843026	843171	Alu-Spxxz	2	3	230
	843109	843271	Alu-Y	2	9	175
	847056	847210	MER104	293	1	179
	847256	847351	L1ME104	343	128	224
10	847413	847551	MIR	99	65	218
	847570	847693	L1ME4	343	1	127
	847863	848137	Alu-Y	2	1	290
	848171	848458	Alu-Sg	2	1	290
15	848493	848564	L1PA7	355	35	105
	848646	848924	Alu-Sx	2	5	290
	849186	849435	L1ME_ORF2	425	2527	2796
	849450	849745	Alu-Sx	2	5	289
	850114	850249	L1P_MA2	348	5447	5610
	850204	850261	L1P_MA2	59	3478	4017
20	850924	850932	L1ME_ORF2	425	1128	1265
	851438	851614	CT	449 (complement)	1	27
	851749	851881	L1ME3A	341	357	523
	852607	852853	L1MA10	72	664	918
	852863	853156	Alu-Sc	2	1	290
25	853176	853211	L1MA10	72	628	663
	853212	853267	L1MA9	75	987	1041
	853491	853779	Alu-Sz	2	1	290
	859137	859435	Alu-Sx	2	1	290
	859436	859456	(A)	449	1	21
30	859570	859805	L1ME3A	342	215	442
	859806	860289	L1ME2	341	375	879
	860318	860505	Alu-Y	2	1	290
	862130	862481	Alu-Sg	2	1	290
	865060	865150	Alu-Sq	2	1	290
35	867521	867800	Alu-Yb	2	1	288
	867836	867876	MIR	99	157	196
	869546	869802	L1NE2	362	123	413
	869923	870118	L1NE2	362	1251	1450
40	870124	870202	Alu-Y	2	48	132
	870203	870296	L1NE2	362	1451	1592
	870316	870666	L1NE2	362	1708	2097
	871000	871075	L1NE2	362	2617	2736
	871650	871935	Alu-Jo	2	1	290
	871936	871960	(GAAAAAA)	4	28	
45	872154	872444	Alu-Sc	2	1	289
	874867	874990	L1MB7	78	529	676
	878120	878408	Alu-Sx	2	1	290
	881003	881054	MLT1G	109	217	268
	881130	881266	MLT1G	109	269	480
50	881293	881346	MLT1G	109	415	469
	881762	881891	L1NE2B	363	85	229
	882140	882740	Alu-Sb0	2	1	290
	883546	883746	Alu-Sg	2	1	288
	883702	883777	Alu-Sx	2	2	290
55	883898	884520	L1P_MA2	348	5600	5926
	884332	884623	Alu-Sp	2	1	290
	884624	885134	L1ME_ORF2	425	2431	2075
	885169	885456	Alu-Yb	2	9	290
	885460	885742	L1ME_ORF2	425	2949	3252
60	885744	886031	Alu-Sx	2	1	288
	886032	886082	Alu-Sp	2	291	341
	886083	886166	L1MB7	78	137	220
	886168	886454	Alu-Sc	2	1	290
	886535	887059	L1MB7	78	345	901
65	887169	887460	Alu-Y	2	1	289
	887483	887748	L1MD2	337	794	1072
	887752	887779	L0R11	366	395	422
	888253	888318	L1NE2	362	2440	2505
	888383	888548	L1NE2	362	2579	2739

WO 01/88089

PCT/US01/15674

-28-

	Begin	End	Repeat Family	SEQ ID NO.	Begin	End
5	88865	888693	LORII	366	394	422
	889006	889296	Alu-Jb	2	5	299
	889249	889248	Alu-Jb	2	188	290
	889249	889477	LPB3	358	770	897
	889421	890133	Alu-Sx	2	1	290
	890515	890797	Alu-Sz	2	1	283
10	890558	890972	LIME2	341	769	885
	890986	891024	LORII	366	396	434
	891028	891063	LTR66	266	173	207
	891126	891536	LINE2	362	1980	2452
	891545	891679	LTR16A1	382	9	128
	891688	891963	LTR16A	381	146	429
15	892907	893013	LINE2	362	2636	2747
	893851	893924	MLTIL	119	47	119
	894528	894849	Alu-Sx	2	1	290
	895825	895903	LINE2	362	2592	2664
	895912	896083	MER20	317	46	216
	897490	897539	MER20	317	2	217
20	897492	897624	Alu-CTTA	2	2	136
	897977	898261	Alu-Sc	2	1	290

The lengths of the non-repetitive intervals were calculated from these data. For example, a non-repetitive interval of 5358 bp was determined between coordinate positions 853779 and 859137 which delineate the boundaries of adjacent Alu-Sz and Alu-Sx repetitive elements. Next, the non-repetitive intervals were sorted based on their respective lengths. Four of these non-repetitive intervals were selected for probe development, namely the above-referenced 5358 bp sequence, a 3847 bp sequence (coordinates 819757 and 823604), a 3785 bp sequence (coordinates 843271 and 847056), and a 3130 bp sequence (coordinates 874990 and 878120).

In the next step, the long PCR technique was used to amplify portions of the four identified single copy intervals. The technique followed for amplification of the 5358 bp interval is described in detail below. Similar techniques were followed for amplification of the remaining three single copy intervals.

Probes of maximal length were desired for FISH experiments. However, in order to optimize the PCR reaction that generated these probes, other constraints had to be met, which resulted in amplification products somewhat shorter than the entire non-repetitive sequence interval. The Prime computer program was employed to optimize the selection of primers for PCR (Genetics Computer Group software package, Madison WI). The PCR primers which were optimized for long PCR were constrained as follows: size of 30-35 nucleotides; GC content of 50-80%; melting temperature of 65-70°C; the primer was not permitted to self-anneal at the 3' end with hairpins of greater than 8 nucleotides; the primer was not permitted to self-anneal at any position with greater than 14; and the primer was

WO 01/88089

PCT/US01/15674

-29-

permitted to anneal only at a single position in the target sequence and primer-primer annealing was limited at the 3' end to less than 8 bp and at any other point less than 14 bp. In addition, certain constraints were applied to the amplified PCR product in order to optimize long PCR: length of 5100-5358 nucleotides; GC content of 40-60%; melting temperature of 70-95°C; difference in forward and reverse primer melting points less than 2°C. This yielded a possible 517 forward primers and 382 reverse primers, and a total of 928 possible products. The Prime program using the foregoing constraints rank ordered potential primer pairs. The top ranked primers were selected for synthesis, as set forth in the following Table 2. These primers were commercially produced (Oligos, Etc., Wilsonville, OR).

-30-

Table 2

Gene	Chromosome Band	GenBank Accession No., Chromosome Genomic Sequence	Coordinates of Longest Single Copy Intervals, Beginning/End	Forward PCR Primer Coordinates, Beginning/End	Reverse PCR Primer Coordinates, Beginning/End	PCR Primer SEQ ID Nos., Forward/Reverse	Probe Length (bp)
HIRA	22q11.2	NT_001039	653779/859137	853946/853975	859116/859085	429/430	5170
HIRA	22q11.2	NT_001039	819757/823604	819901/819933	822592/823559	421/432	3691
HIRA	22q11.2	NT_001039	843271/847056	843602/844631	844546/845615	433/434	3344
HIRA	22q11.2	NT_001039	874990/878120	875226/875257	878074/878042	435/436	2848

5

10

WO 01/88089

PCT/US01/15674

-31-

Using these primers, a long PCR reaction (50 µl) was performed using 1 microgram of high molecular weight genomic DNA (purified by phenol extraction) and 200 µM of each oligonucleotide primer to amplify the 5170 bp probe. Specifically, high fidelity DNA polymerase (LA-Taq, Takara Chemical Co.) was employed using the following thermal cycling protocol:

5 Step 1 - 94°C - 5 minutes
Step 2 - 98°C - 20 seconds
Step 3 - 65°C - 7 minutes
Step 4 - 14 times to Step 2
10 Step 5 - 98°C - 20 seconds
Step 6 - 65°C - 7 minutes + 15 s/cycle
Step 7 - 14 times to Step 5
Step 8 - 72°C - 10 minutes
Step 9 - 0°C
15 Step 10 - END

Because amplification of the 5170 bp probe is less efficient than amplification of shorter fragments, the initial PCR reaction did not yield sufficient quantities of probe for multiple hybridization experiments. Therefore, a 4 µl aliquot of the original DNA amplification reaction was reamplified using the following protocol: Step 1 - 94°C - 1.5 minutes, followed by Steps 2-10 of the original PCR reaction. Sufficient quantities of the 5170 bp probe were obtained. An alternative to reamplification is to increase Step 7 by at least 10 cycles.

The amplified product was then purified by gel electrophoresis followed by column chromatography. First, the amplified product was separated on a 0.8% Seakem 25 LE agarose gel (FC Bioproducts) in 1X modified TAE buffer. The gel was then stained with ethidium bromide and visualized with UV light. The fragment corresponding to the correct interval size was excised in an Ultrafree-DA spin column (Millipore) and centrifuged at 5000g for 10 minutes. The DNA was recovered in solution and precipitated in 1/10 V NaOAc and 2.5 V 95% EtOH (overnight) at -20°C. The 30 precipitated DNA was then centrifuged, rinsed with cold 70% EtOH, air dried and

WO 01/88089

PCT/US01/15674

-32-

resuspended in 20 µl of sterile deionized water. The DNA was checked on a 0.8% agarose gel (Sigma) to determine DNA concentration.

The detailed probe labeling, hybridization, removal of non-specifically bound probe, and probe detection procedures are described by Knoll and Lichter, In: Dracopoli et al., (eds), "Current Protocols in Human Genetics Volume 1", Unit 4.3 (Green-Wiley, New York, 1994). Briefly, in order to label the probe, a standard nick translation reaction was carried out (Rigby et al., *J. Mol. Biol.*, 113:237-251, (1977)) using digoxigenin-11-dUTP as the label. This yielded a series of overlapping 300-500 bp labeled fragments, which together comprised the 5170 bp probe.

10 The labeled probe fragments were then precipitated by adding 1/10 V NaOAc plus 2.5 V 95% EtOH and carrier DNA (overnight, -20°C). On the following day, the precipitated DNA was centrifuged, lyophilized, and resuspended in deionized sterile water at a concentration of 125 ng/20 µl.

15 A comparison set of hybridizations were carried out with normal denatured human metaphase chromosomes, using the labeled probe fragments with and without blocking nucleic acid of the type described in U.S. Patents Nos. 5,447,841, 5,663,319 and 5,756,696. Twenty µl of resuspended labeled probe was then lyophilized and resuspended in 10 µl of deionized formamide and denatured for 5 minutes at 70-75°C to yield single-stranded nucleic acids. For comparison, probes were pre-reacted with 20 purified repetitive DNA by adding 125 ng (or 20 µl) of labeled probe to 10 micrograms of C_{ot} 1 DNA (Life Technologies) and lyophilizing the mixture. This mixture was then denatured for 5 minutes at 70°C followed by pre-reaction (or pre-annealing) for 30 minutes at 37°C to convert the single stranded repetitive sequences in the probe to double stranded nucleic acid. This disables the hybridization between the sequences and the chromosome as target DNA template.

25 Subsequently, the denatured probes with or without purified repetitive DNA (i.e., C_{ot} 1) were mixed with 1 V prewarmed hybridization solution (comprised of 4 x SSC/2 mg/ml nuclease free bovine serum albumin/20% dextran sulfate/30% sterile deionized water) and overlaid onto denatured target DNA. The chromosomal target DNA, fixed to a microscope slide had been denatured at 72°C for 2 minutes in 50% formamide/2 x SSC. A coverslip was placed over the probe hybridization mixture on

WO 01/88089

PCT/US01/15674

-33-

the slide, sealed with nail polish enamel to prevent evaporation and placed in a moist chamber at 39°C overnight.

Following hybridization, non-specifically bound probe was washed off with varying stringencies of salt concentration and temperature. The labeled probes, pre-reacted to disable repetitive sequence hybridization, and the probes without such pre-reaction were detected with rhodamine-labeled antibody to digoxigenin-11-dUTP, using a conventional FISH protocol (Knoll and Lichter, *Current Protocol in Human Genetics*, Vol. 1, Unit 4.3, Green-Wiley, New York, 1994). Chromosomal DNA was counterstained with DAPI. The cell preparations on microscope slides were then mounted in antifade solution (such as Vectashield, Vector Laboratories, Burlingame, CA) and visually examined using a fluorescence microscope with the appropriate fluorochrome filter sets. Figs. 1 and 2 are photographs illustrating the results of the comparative hybridizations, where Fig. 1 is the hybridization with the blocking repetitive sequences, while Fig. 2 is the hybridization without pre-reaction with purified repetitive DNA. These photographs depict hybridization to both HIRA alleles on two normal chromosome 22q11.2 regions. A comparison of the photographs demonstrates that the presence of the blocking repetitive sequences is unnecessary using the probes of the present invention.

The remaining three probes identified in Table 2 were PCR-amplified and labeled as described above. These probes were used in a series of FISH experiments to determine the efficacy of the probes. Thus, all four probes were used together without pre-annealing of potentially repetitive sequences (Fig. 6), and a combination of the three shortest probes were used on cells from a patient affected with DiGeorge/VCFS with a previously confirmed deletion (Fig. 12). In the Fig. 6 photograph, the probe was hybridized to a single region of both chromosome 22s in a normal individual (arrows). In Fig. 12, only one chromosome 22 hybridized (arrow). The other chromosome 22, as indicated by a star, has a deletion of this region and does not hybridize to the probe.

WO 01/88089

PCT/US01/15674

-34-

Example 2

Development of NECDIN and CDC2L1 Gene Probes

The techniques described in Example 1 were used to develop a series of probes for detecting known genetic disorders on chromosome 1 (Monosomy 1p36.3 syndrome; Slavotinek et al., *J. Med. Genet.*, **36**:657-63 (1999)) and on chromosome 15 (Prader-Willi and Angelman Syndromes). Approximately 70% of patients with Prader-Willi or Angelman syndrome exhibit hemizygous deletions of the sequence containing the NECDIN gene (Knoll et al., *Amer. J. Med. Genet.*, **32**:285-290 (1989); Nicholls et al., *Amer. J. Med. Genet.*, **33**:66-77 (1989)). The presence of excess copies of this gene is diagnostic for an abnormal phenotype in patients with interstitial duplication or a supernumerary derivative or dicentric chromosome 15 (Cheng et al., *Amer. J. Hum. Genet.*, **55**:753-759, (1994); Repetto et al., *Am. J. Med. Genet.*, **79**:82-89, (1998)). The following Table 3 sets forth the deduced single copy intervals, PCR primer coordinates, SEQ ID Nos., and the lengths of the resultant probes.

15

-35-

Table 3

Gene	Chromosome Band	GenBank Accession No., Chromosome Genomic Sequence	Coordinates of Longest Single Copy Intervals, Beginning/End	Forward PCR Primer Coordinates, Beginning/End	Reverse PCR Primer Coordinates, Beginning/End	PCR Primer SEQ ID Nos., Forward/Reverse	Probe Length (bp)
CDC2L1 ^a	1p36.3	AL031282	8823/17757	9137/9167	13560/13931	444/443	4823
CDC2L1 ^a	1p36.3	AL031282	8823/17757	13028/13057	17552/17720	445/446	4724
NECDIN	15q11-q13	AC006596	94498391/52	9450/94535	98567/98601	439/440	4166
NECDIN	15q11-q13	AC006596	6803/175948	7212/72156	75666/75637	437/438	3544
NECDIN	15q11-q13	AC006596	76249/79221	76608/76639	78998/78867	441/442	2290

5

15

^aTwo sets of primers were used to generate two DNA probe fragments which, together, spanned the entire interval.

10

WO 01/88089

PCT/US01/15674

-36-

PCR-amplification was performed using the CDC2L1 primers in Table 3, and products were labeled, hybridized and detected as set forth in Example 1. The labeled probes were used in a series of FISH experiments, with images of the hybridizations provided as Figs. 7-10. In the experiment shown in Fig. 7, the longest 4823 bp probe 5 was employed and potential hybridization repetitive sequences was disabled by pre-annealing with purified repetitive DNA. As a comparison, the same probe was used without pre-annealing of purified repetitive DNA (Fig. 8). The hybridizations appear identical demonstrating that the presence of purified repetitive DNA to block repetitive sequence hybridization is unnecessary. In both instances, the chromosomes with one 10 or both chromatids hybridized are indicated by arrows. In the experiments shown in Figs. 9 and 10, the 4823 bp and 4724 bp probes were employed, with (Fig. 9) and without (Fig. 10) pre-annealing of the purified repetitive DNA. Again, pre-reaction of the purified repetitive DNA is shown to be unnecessary using the probes of the invention.

15 The NECDIN probes were also used in a series of FISH experiments, as shown in Figs. 3-5 and 11. These probes detected DNA sequences between 36 and 62 kb distal of the NECDIN gene. The 3544 bp probe (SEQ ID Nos. 437-438) detected the 3' terminus of the MAGEL2 gene. In Fig. 3, the 3544 bp probe was used on metaphase cells from a normal individual, with pre-annealing using purified repetitive sequences; 20 Fig. 4 is a comparison, without pre-annealing. In Fig. 11, all three probes were used in combination, on metaphase cells from a patient affected with Prader-Willi syndrome known to harbor a deletion of 15q11-q13 sequences on one chromosome 15. The normal homolog is indicated by an arrow and shows hybridization to a single chromatid. The location of the deleted chromosome is indicated by a star. It does not show hybridization with the probe.

25 The foregoing examples demonstrate that the mixed combinations of DNA fragments give identical hybridization results, as compared with the fragments when used individually. This establishes that none of the fragments used individually or in combination will hybridize to any other location in the genome and hence, are free of repetitive sequences. This provides an additional confirmation of the validity of the present method for the design and production of single copy genomic probes.

WO 01/88089

PCT/US01/15674

-37-

5 Current use of commercial and research genomic probes to detect these disorders requires that hybridization of repetitive sequences be disabled prior to annealing of the probe to metaphase or interphase chromosomes. This increases the number of steps required to perform the protocol and could potentially increase the chances of procedural errors occurring, any of which would be unacceptable in the clinical diagnostic laboratory. The results present in Fig. 7 are comparable to those obtained using related commercially available genomic probes to detect these abnormalities. Hence, these probes will be useful in the detection of these genetic disorders. The probes themselves or in combination with other solutions necessary for 10 hybridization and detections can be provided to clinical laboratories as kits for detection of these genetic disorders.

15 The probes developed from genomic sequences other than those presented as examples cited herein can also be utilized to detect inherited, sporadic, or acquired chromosomal rearrangements. These rearrangements may correspond to numerous other known genetic abnormalities (including neoplasias) and syndromes besides those examples given above. Hence, the present invention can also be useful for producing probes from genomic regions where no commercial probes are available or the probes are imprecise.

20 In principle, the present method can be utilized to design, develop and produce single-copy genomic probes for any genomic interval where the DNA sequence is available and where a comprehensive set of repetitive sequence elements in the genome has been cataloged. Such catalogs are currently available for genomes for the following organisms (<http://www.girinst.org>): *Homo sapiens*, *Mus musculus*, *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Danio rerio*.

25

Example 3

30 In this example, a number of probes specific to additional genetic disorders and cytogenetic abnormalities were developed using the principles of the invention. Software was also developed and improved to expedite the process of designing single copy probes (findi.pl, prim_wkg, and prim, referred to above and provided on the

WO 01/88089

PCT/US01/15674

-38-

accompanying CD-R). The probes were subsequently tested and their utility confirmed by *in situ* hybridization.

Identification of single copy sequences.

5 The locations of single copy probe sequences are determined directly from long contiguous genomic DNA sequences. The locations were determined by software that aligns the sequences of repetitive sequence family members with the target genomic sequence. Comparison of the target sequence with previously determined sequences of repetitive family members served to identify and delineate the bounds of repetitive elements within the target. The computer program, RepeatMasker (<http://ftp.genome.washington.edu/RM/RepeatMasker.html>; Smit A.F.A. & Green P., unpublished results), was used to determine the locations of repetitive sequence families in contiguous genomic sequences, usually ~100 kb in length. RepeatMasker compares a genomic sequence with a compilation of repetitive sequence families present in multiple copies in the human genome (<http://www.girinst.org/~server/rephbase.html>). This repeat sequence database contains representative and consensus sequences for the majority of human repetitive sequence families. The database can be expanded by addition of newly discovered repetitive sequence families (as shown in Example 6).

20 A Perl script (findi.pl) parsed the coordinates of the boundaries of the repetitive segments from RepeatMasker output, and then deduced and sorted the adjacent single copy intervals by size greater than a parametrized threshold (~2 kb, in most instances). This script determines the locations and lengths of single copy intervals sorted by size from the output file (with the suffix: .out) produced by RepeatMasker, which contains a table of locations and lengths of repeat family elements. The boundaries of adjacent 25 single copy intervals were deduced by subtracting one nucleotide position from the upstream boundary of a repetitive element and adding one nucleotide position to the downstream boundary of the previous element. Single copy intervals with identical upstream and downstream coordinates (1 bp in length) were considered to be adjacent repetitive sequences. Probe sequences were then compared with the human genome sequence database (Altschul et al., *J. Mol. Biol.*, **215**:403-410 (1990)) to determine if there was similarity to sequences elsewhere in the genome (such as duplicons or

WO 01/88089

PCT/US01/15674

-39-

triplicons or other less well conserved intervals). Probe sequences that are weakly conserved elsewhere in the genome do not cross-hybridize to those targets.

Oligonucleotide primers were selected for PCR amplification of the longest single copy intervals. A Unix wrapper script (prim_wkg) iteratively modifies the switches in the file containing the command to design primers (prim), thus optimizing primer selection by changing the following parameters for input to the program, Prime (Genetics Computer Group; Madison, WI): T_m (from 70-60°C), G/C composition (from 55-40%), and minimum interval length (from 90%-80% of the length of the single copy interval).

10

Probe generation and chromosomal in situ hybridization.

DNA fragments were amplified by long PCR (Cheng et al., *Proc. Natl. Acad. Sci. U.S.A.*, **91**:5695-5699 (1994)) with LA-Taq as recommended by the manufacturer (Panvera, Madison, WI). Other enzymes for long PCR have been demonstrated to produce comparable results, including those manufactured by Roche Molecular Biochemicals, Indianapolis, IN; Stratagene, LaJolla, CA; and Invitrogen, Carlsbad, CA. The amplicons were purified by low-melt temperature agarose gel electrophoresis, followed by chromatography with Micro-con 100 columns (Millipore, Bedford MA), which removed contaminating extension products containing repetitive sequences.

20

Probe fragments were labeled by nick translation using modified nucleotides such as digoxigenin-dUTP or biotin-dUTP (Roche Molecular Biochemicals, Indianapolis, IN). Labeled probes were denatured and hybridized to fixed chromosomal preparations on microscope slides using previously described conditions (Knoll and Lichter, *Current Protocols in Human Genetics*, Vol. 1, Unit 4.3, Green-Wiley, New York (1994)), with the exception that preannealing of the probe(s) with repetitive DNA (such as C_nI DNA) was not utilized in a parallel set of hybridizations. Probes from a single chromosome region of ~100kb were hybridized individually or in combination to remove non-specific binding. Post-hybridization washes were performed at 42°C in 50% formamide in 2xSSC, followed by an additional wash at 39°C 2xSSC and one in 1xSSC at room temperature. Wash stringency was increased, if necessary, to remove hybridization of probes to related sequences elsewhere in the genome. Hybridized probes were detected

25

30

WO 01/88089

PCT/US01/15674

-40-

with a fluorochrome (such as rhodamine or fluorescein) tagged antibody to the modified nucleotide. Chromosome identification was performed by counterstaining the cellular DNA with 4', 6-diamidino-2-phenylindole (DAPI). Hybridized chromosomes were viewed with an epifluorescence microscope (Olympus, Melville, NY) equipped with a motorized multi-excitation fluorochrome filter wheel. Hybridization patterns on at least 20 metaphases (and 50-100 nuclei) were scored for each probe or combination of probes, with and without preannealing to C_ot1 DNA. Cells were imaged using a CCD camera (Cohu, Inc, San Diego, CA) and CytoVision ChromoFluor software (Applied Imaging, Santa Clara, CA).

Table 4 sets forth the data generated using the foregoing procedure, with respect to a number of probes specific to known disorders and cytogenetic abnormalities. The abnormality designation makes use of standardized nomenclature as set forth in ISCN 1995, An International System for Human Cytogenetic Nomenclature (1995), Mittelman F, ed.

15

Table 4

Disorder	Representative Cytogenetic Abnormality Detected by Probes on Metaphase Chromosomes	Gene or Transcript ^a	Interval	GenBank Accession No.	Forward PCR Coordinate, Beginning/End	Reverse PCR Coordinate, Beginning/End	Sequence ID
Angelman Syndrome and Prader-Willi Syndrome	ish del(15q11.2)(11p15.5)(DDEA-)	<i>LBE34</i>	WVS 8-IVS 9	AC004600	4108541119	45354545325	480481
Duplication 15 Syndromes	ish dup(15q11.2)(11p15.5)(CNSRNP+) and ish del(15q1.2)(11p1.2)(DDEA+)	<i>I2CSRN</i>	WVS 5 to Exon 1B*	AC004757	13740716179	15140715387	482483
Kallmann Syndrome	ish del(X)(q22.3)(KAL1+)	<i>I2CSRN</i>	WVS 5'-Exon 1B-IVS 3'	AC004757	3110231128	33347733323	484483
Tuber Syndrome and Dibial Xp Deletions	ish del(X)(q22.3)(GSB2-)	<i>KAL1</i>	IVS 5'-Exon 1B upstream	AC004757	4779247121	4870494411	486487
Down Syndrome	ish del(X)(q22.3)(GSB2-)	<i>KAL1</i>	IVS 10	AC006822	1044337104165	107097107072	488489
Chronic Myelogenous Leukemia and Acute Lymphoblastic Leukemias	ish (9;22)(q34;q11.2)(BCR st)	<i>SHOX</i>	IVS 1-Exon 2	NT_001151	4161544466	47305497473	492493
	ish (9;22)(q34;q11.2)(BCR st)	<i>SHOX</i>	IVS 2	NT_001151	49637748969	5225152217	494495
	ish (9;22)(q34;q11.2)(BCR st)	<i>SHOX</i>	IVS 3	NT_001151	51337545387	56821156791	496497
	ish del(X)(q22.3)(TBL1+)	<i>TBL1</i>	Promoter-IVS 2	NT_001159	78977910457	8299482560	508490
	ish (21)(q22.3)(DSCR4+)	<i>DSCR4^b</i>	3' UTR	NT_001159	175376154509	17666517963	502503
	ish (21)(q22.3)(DSCR4+)	<i>DSCR4^b</i>	-30kb upstream	AP001160	217602751275	217602751275	504504
	ish (21)(q22.3)(DSCR4+)	<i>DSCR4^b</i>	-30kb upstream	AP001160	31087310149	32099339265	506507
	ish (21)(q22.3)(DSCR4+)	<i>DSCR4^b</i>	-30kb upstream	AP001160	4072540754	43078439145	508509
	ish (9;22)(q34;q11.2)(BCR st)	<i>BCR</i>	Proximal to the major breakpoint in the BCR	U07000	4997350006	5240735006	510511
	ish (9;22)(q34;q11.2)(BCR st)	<i>BCR</i>	Proximal to the major breakpoint in the BCR	U07000	100745100776	104145104115	510511
	ish (9;22)(q34;q11.2)(BCR st)	<i>BCR</i>	IVS 8	U07000	74689747128	7790377874	512513
	ish (9;22)(q34;q11.2)(BCR st)	<i>ABL1</i>	Exon 1-IVS 1B	U07061	114945114978	117457117426	514515
	ish (9;22)(q34;q11.2)(BCR st)	<i>ABL1</i>	IVS 1B	U07062	2171227713	23938226357	516517
	ish (9;22)(q34;q11.2)(BCR st)	<i>ABL1</i>	IVS 4-IVS 6	U07062	91959222	1035110084	518519
	ish (9;22)(q34;q11.2)(BCR st)	<i>ABL1</i>	Exon 11-IVS 11	U07363	20266720737	20266720737	520521
					7886278891	8381303784	522523

u1B is ~16 kb upstream from the PWS shortest region of overlap and ~35 kb upstream from the AS shortest region of overlap. Probe sequence is not found in public repetitive sequence database. Probe also shows hybridization to a sequence found on the p arm of centromeric chromosomes. Probe sequence is not found in the public repetitive sequence database.

WO 01/88089

PCT/US01/15674

-42-

Disorder	Representative Cytogenetic Abnormality	Gene or Transcript	Interval	GenBank Accession No.	Forward PCR Primer Coordinates, Beginning/End	Reverse PCR Primer Coordinates, Beginning/End	Sequence ID
Detected by Probes on Metaphase Chromosomes							
Chronic Myelogenous Leukemia, Acute Lymphoblastic Leukemia	ish del(19)(q11.2)/ABL(mv)	<i>ABL</i>	IVS 3	U07563	5380752656	5807758946	534925
Williams Syndrome	ish del(7)(q11.23)(q11.23)(MLNK1+)	<i>MLNK1</i>	IVS 3	U07563	5357052604	5784855455	536227
Acute Myelogenous Leukemia-Type M4	ish inv(16)(q13q22)(PMS5 ip) ish inv(16)(q13q22)(PLA2G10 mv, PMS5 ip)	<i>PMS5</i>	IVS 4+IVS 6	U07563	5358245588	5784857171	536229
Rubinstein-Taybi Syndrome	ish del(16)(p13.3)(CREBBP)	<i>CREBBP</i>	IVS 2	U003938	5994759976	6633365367	530931
Acute Lymphocytic Leukemia	ish del(16)(q13.3-q22)(AML1 si) ish (12;21)(q13.3;q22)(TEL/ETV6 inv)	<i>AML1</i>	>20 kb downstream	NT_000691	5319661993	5301534889	534635
Cri-du-Chat Syndrome	ish del(5)(q15.3)(CNND2+)	<i>CNND2</i>	>60 kb downstream	NT_000691	6428464233	2798827958	536237
Langer-Giedion Syndrome	ish del(8)(q23.3q24)(TRPS1+)	<i>TRPS1</i>	IVS 3	NT_000691	6428464233	6768257652	538439
Schild-Magenis Syndrome	ish del(17)(p11.2p11.2)(ADGRG2B+)	<i>ADGRG2B</i>	IVS 12-Exon 15	U003930	6827168300	7198671957	540941
	ish del(17)(p11.2p11.2)(ELL1+)	<i>ELL1</i>	IVS 3	NT_000691	7195771966	7548175452	542543
		<i>PMS5</i>	Exon 15-IVS 20				
			<100 kb upstream & <300 kb downstream				
			IVS 6	NT_025903	313703313812	315675315645	5449245
			IVS 18	NT_000691	583358862	6134763138	546247
			<i>AML1</i>	AP000537	9871298741	102093102872	548249
			IVS 4	NT_000691	9545695480	9728327260	550931
			IVS 3	NT_000691	7254972564	7438574561	552533
			IVS 2	3216508245	4009160062	554955	
			IVS 17	106655169685	171596171945	556357	
			IVS 14	NT_000149	199168199202	203307203343	558359
			IVS 3	NT_000147	2390573935	2771075763	560561
			IVS 3	NT_000147	3075739790	3324153309	562563
			IVS 3	AF119117	1471614748	1778717753	564563
			IVS 3	2520182329	3138941860	366567	
			IVS 1	NT_002386	267751267760	270758270724	566569
			IVS 1	NT_002386	271232271271	274437274404	570571
			Promoter-IVS 1	NT_000770	5644156412	5652458491	572573
			IVS 1	NT_000770	7744277475	79222779189	574575
			IVS 12	U00164	6091941627	730017757	576977

⁴ PMS5 is ~3 mb telomeric of MYTH1 genes, which is disrupted at the inv(16p) breakpoint. PLA2G10 is ~200 kb telomeric of PMS5.

WO 01/88089

PCT/US01/15674

-43-

Disorder	Representative Cyto genetic Abnormality Detected by Protein on Metaphase Chromosomes	Gene or Transcript	Interval	GenBank Accession No.	Forward PCR Primer Beginning/End	Reverse PCR Primer Beginning/End	Sequence ID
Santini-Magnus Syndrome (cont'd)							
ish del(17)(q11.2p11.2) (X)(A4p4)	<i>PLXJ</i>	IVS 12-16S 14	1380184	72617455	97428708	57867579	508561
ish del(17)(q11.2p11.2) (X)(B9pA1p1V)	<i>PLXH4</i>	IVS 15-23 UTR	1069183	96186614	1758116764	5849365	5136215264
ish del(17)(q11.2p11.2) (X)(B9pA1p1V)		Between ZNF779 and KRT11	AL0355367	98149950	1227121241	5849365	
SHMT1	<i>LIG3</i>	Primer: Exon 2	AL0355467	119417226	556556534	5866567	
Primer: IVS1	<i>HUGO</i>	Primer: Exon 2	AC005703	131317152302	159227151994	5885659	
ish del(17)(q11.2p11.2) (X)(B9pA22+1)	<i>BAG22</i>	-5'3' homologous	AL0055013	215623215661	21756232171929	5095591	
	<i>BAG22P</i>	IVS 3	AL0055013	184566187510	186335187506	5025931	
	<i>BAG22P</i>	-5'3' downstream	AL0055013	13267117578	13267117578	5065595	
ish del(17)(q11.3;3p) (PAFAH1B1/EFB-E-3)	<i>PAFAH1B1</i>	IVS 34-IVS 27	NT_000774	65851565673	66683366573	5065597	
Milne-Dickie Syndrome	<i>EFB-3</i>	-7'3' upstream	NT_000774	688416168870	7118571163	5095599	
	<i>PAFAH1B1</i>	IVS 15'-IVS 19	NT_000774	753328775362	78122778093	6001601	
	<i>PAFAH1B1</i>	-13'6' downstream	NT_000774	753328775362	78122778093	6001601	
Alagille Syndrome	<i>EIF-3</i>	IVS 5-IVS 11	AL03546524	153935153966	157675157642	602603	
	<i>JAG1</i>	IVS 2-IVS 3	AL03546524	144875149494	157028516995	6046605	
	<i>JAG1</i>	Exon 1-IVS 2	AL03546524	139449116995	139449116995	6066007	
		-5'8'1' downstream	AL1335338	111141111145	116461160107	6086609	
		-2'8' upstream	AL1335338	128535128627	133315130367	6106611	
		Exon 67	NT_000102	2670614267643	2711202711091	612613	
Monosomy 13q22	ish del(13)(q22.3) (ZIC2-2)						
Trisomy 13	ish (13)(q22.3) (ZIC2-2)						
	ish del(13)(q22.3) (ZIC2-2)						
Wolf-Hirschhorn Syndrome	ish (1p16.2) (G4D)						

Probe was hybridized in combination with offer probes and not individually.

WO 01/88089

PCT/US01/15674

-44-

Example 4

In this example, a more precise chromosomal breakpoint determination was made using the probes of the invention. Structural chromosome rearrangements can be inherited in genetic disease or acquired as in the case of certain cancers. They can occur within a single chromosome (such as an inversion, deletion or duplication) or between homologous or non-homologous chromosomes (i.e. translocations). With the probes of the invention, the precise region of breakage can be determined at a previously unprecedented level of resolution. Such resolution permits detection of genes or sequences that are disrupted in the formation of the rearrangement and may provide insight into etiology, prognosis and/or treatment. In inherited contiguous gene syndromes, precise localization of chromosome breakpoints can define the extent of deletion or duplication and hence the prognosis of the disorder (Cheng et al., Am. J. Hum. Genet., 55:753-759 (1994)).

The following example illustrates how the single copy probes hereof provide more precise information than commercially available cloned probes for the same chromosomal region.

In most cases of chronic adult myeloid leukemia (CML; 90%) and in some cases of acute lymphoblastic leukemia (ALL; 25-30% adults; 2-10% children) (Perkins et al., *Cancer Genet. Cytogenet.*, 96(1): 64-80 (1997); Rubintz et al., *J. Pediatr. Hematol. Oncol.*, 20 (1):1-11 (1998)), a reciprocal translocation between chromosome 9q34 and 22q11.2 is evident (Rowley, *Nature*, 243:290-392 (1973)). The abnormal or derivative chromosome 22 that results from this translocation fuses the ABL1 oncogene on chromosome 9 to the BCR (breakage cluster region) promoter on chromosome 22. The ABL1 oncogene is expressed as either a 6 or a 7 kb mRNA transcript with alternatively spliced first exons, exons 1b and 1a respectively, spliced to the common exons 2-11. Exon 1b is ~250 kb proximal of exon 1a and this very long intron is a primary target for translocations. In CML, the ABL1 gene is translocated from chromosome 9 to the promoter of the BCR gene on chromosome 22 to produce a chimeric BCR-ABL1 protein (Bernards, et al., *Mol. Cell. Biol.*, 7:3231-3236; (1987). The BCR gene contains 24 exons and encodes a 160kD protein. The BCR breakpoints differ in CML and ALL. In CML, most breakpoints occur within the 5.8 kb major breakpoint cluster region (M-

WO 01/88089

PCT/US01/15674

-45-

BCR) which corresponds to exons 12 through 17 (or b1 through b5); whereas in most ALL, the BCR gene breaks between exons 1 and 2 (minor or m-BCR). Thus different molecular rearrangements resulting in differently-sized proteins occur in each disorder. These rearrangements can be distinguished with the probes of the invention.

At the DNA level, dual color-dual probe fluorescence in situ hybridization (FISH) strategies are available to detect the BCR-ABL1 translocation on metaphase chromosomes and interphase nuclei (Bentz et al., *Blood*, **83**:1922-1928 (1994); Sinclair et al., *Blood*, **90**:1395-1402 (1997); Buno et al., *Blood*, **92**:2315-2321 (1998)). With conventional FISH, the initial strategy was to have the probe for the ABL1 oncogene region (distal of the breakpoint and ~200kb in size) detected in one color (i.e., red) and the BCR region (proximal of the breakpoint) in another color (i.e., green). Thus, in derivative chromosome 22 positive cells, one red and one green signal co-localized to give a yellow hybridization signal indicating the presence of a derivative 22 chromosome while the normal chromosome 9 and chromosome 22 remained as independent red and green signals. More recently, larger cloned DNA probes that span both sides of the ABL1 translocation breakpoint have been utilized for detecting translocations. In this strategy, the part of the probe that is proximal to the breakpoint remains on the abnormal chromosome 9 and the part distal to the breakpoint co-localizes with BCR as in the previous strategy. This results in an extra signal (ES) on the translocated chromosome 9 and is the strategy behind the BCR-ABL1 ES dual-color translocation probe (Vysis, Inc., Downers Grove, IL) that many clinical cytogenetics laboratories use (Herens et al., *Br. J. Haem.*, **110**(1):214-216 (2000); Sinclair et al., *Blood*, **95**(3):738-744 (2000)). In the ES system, the ABL1 probe cocktail spans a genomic target significantly larger than the ABL1 gene. This cocktail extends from arginosuccinate synthetase gene (ASS), which is ~250 kb upstream from ABL1, through the ABL1 gene and several kb downstream.

Several single copy probes were designed for ABL1 and BCR by identifying the boundaries of single copy intervals at these loci. Figs. 13 and 14 indicate the locations of potential single copy probes within the BCR and ABL1 loci, respectively. Eleven intervals exceeding 2 kb length are distributed throughout the BCR gene, 5 of which have been currently designed as probes. A similar number of additional shorter single

WO 01/88089

PCT/US01/15674

-46-

copy regions in this gene (between 1.5 and 2.0 kb) could be combined to more precisely delineate translocation breakpoints. In the ABL1 gene, 10 intervals > 2kb are found, six of which have been designed as probes.

Multiple single copy probes for both BCR (Fig. 13) and ABL1 (Fig. 14) have been deduced and oligonucleotide primer sets derived. For BCR, these probes discriminate between the minor (Fig. 13) and major breakpoints (Fig. 13 and SEQ ID Nos. 510-515). Conventional FISH testing with BCR-ABL1 ES probe does not distinguish between rearrangements at the major (M) and minor (m) breakpoints in the BCR gene. A mixture of m-BCR probes detects the minor breakpoint often seen in patients with ALL, and those designated M-BCR (SEQ ID Nos. 510-515) detect the major breakpoint evident in most patients with CML. If all of the probes translocate to the chromosome 9, then the gene is interrupted at the minor breakpoint. If only the M-BCR probes translocate from chromosome 22 to the derivative chromosome 9 and the m-BCR probes remain on the derivative chromosome 22, then the gene is interrupted at the major breakpoint.

For ABL1, two of the probes are predicted to be proximal of the breakpoint (SEQ ID Nos. 516/517 and 518/519), and the others are distal to the breakpoint (SEQ ID Nos. 520 through 531). Hybridization of three ABL1 probes distal to the breakpoint in CML shows that the probes have moved to the derivative chromosome 22 (Fig. 15), whereas a combination of five ABL1 probes that span the breakage interval demonstrates that some probes remain on the derivative chromosome 9 while others move to the derivative chromosome 22 (Fig. 16). Based on these results and the positions of these probes, it can be deduced that the breakpoint interval spans positions 11004 bp through 65951 bp of Fig. 14. It will be evident to one skilled in the art that the region of breakage can be more precisely refined by hybridizing probes from the single copy intervals between Seq. ID Nos. 519 and 520. The exact location at the breakpoint can be then determined from the genomic sequence of the refined breakage region.

Recent studies (Herens et al, *Br. J. Haem.*, 110:214-216 (2000)) utilizing the commercially available ES probe have demonstrated that ~10% of CML patients do not have an extra hybridization signal on the derivative chromosome 9 because sequences

WO 01/88089

PCT/US01/15674

-47-

are deleted upstream of the ABL1 gene. In some instances, the deletions extend as far as the ASS gene and such a deletion is associated with poor prognosis (e.g., blast crisis). These large FISH probes are not useful for detecting interstitial deletions in the region between ASS and ABL1, as evidenced by an increased deletion detection rate of up to 1/3 of CML patients when shorter probes of ~100-200 kb are hybridized (Sinclair et al., *Blood*, 95(3):738-744(2000)). Since some patients harbor deletions of sequences proximal to the ABL1 breakpoint on the derivative or translocated chromosome 9, single copy probes are being used to delineate the extent of hemizygosity in this chromosomal region. Correlation of deletion breakpoints with clinical outcomes will determine if the loss of specific genes in this chromosomal interval is prognostic for clinical findings such as early blast crisis.

Example 5

This example demonstrates that increased hybridization signals can be generated using probe sequences from duplicated genomic domains. Several single copy probe sequences were designed, which were a part of highly similar duplon or triplicon domains as previously described.

Several probes from chromosome 16p13.1 (SEQ ID Nos. 536-543), close to the inversion breakpoint in Acute Myelogenous Leukemia - Type M4, contain sequences that are a near perfect triplication of sequences in this region. In the genome draft sequence, two of these domains are tandemly arranged, separating the probe sequences by 40 kb, and a third telomeric interval is separated by 1.2 mb. The sequences of the three intervals differ by only ~1.5%. Hybridization with this probe demonstrated two clustered, but clearly separable signals. One hybridization corresponds to the combined first and second paralogs and the other to the third copy of this sequence.

A probe from the chromosome 17p11.2 interval that is commonly deleted in Smith-Magenis Syndrome (SEQ ID Nos. 586-587) contained a near-perfect triplication. The probe was intended to detect a deletion within a near single-copy sequence in IVS4 of the SHMT1 gene. However, two paralogous subsequences separated by ~15 kb, exhibiting 99.8% identity with the SMHT1 sequence, were also detected in the genome draft between the ZNF127 and PAIP1 genes ~2.7 mb centromeric but also within the

WO 01/88089

PCT/US01/15674

-48-

genetic interval commonly deleted in patients with this disorder. Due to the proximity of these sequences to the chromosome 17 centromere (which is a highly condensed region), a single hybridization locus was observed.

Two probes from the Down syndrome critical region (SEQ ID Nos. 504/505-508/509) were each embedded in the same large duplilon, which was ≥ 80 kb in length. These duplicated sequences are separated by 1.1 mb, and reside, respectively at the centromeric and telomeric ends of chromosome band 21q22.2. Despite the fact that the duplicated sequences were separated by 1.1 mb, a single hybridization signal was detected in this region of chromosome 21 using each probe. Therefore, this region of chromosome 21, like chromosome 17p11.2, seems more condensed in metaphase chromosomes than sequences in 16p13.1, in which duplicated regions separated by similar distances were distinguishable.

Single copy probes developed from such regions, of course, lack known repetitive sequence elements. However the probes generally hybridize to all of the paralogous copies, since each of the copies remain hybridized even under the most stringent hybridization wash conditions. Because multiple, tightly clustered sites on the chromosome are hybridized in a specific interval, the hybridization signal produced from these hybridizations is brighter than that expected from a comparable probe sequence which was represented once per haploid genome. Thus, these genomic duplicons or triplicons increase the effective target size of the probe. This implies that shorter probes from such regions can produce hybridization signals comparable in intensity to those generated by longer probes. Selection of shorter probes from duplicated genomic domains will be particularly useful for development of probes for genomic regions where long single copy intervals are underrepresented.

Example 6

The increasing availability of accurate draft human genome sequences has facilitated development of single copy probes in accordance with the invention for many previously inaccessible chromosomal regions. Although the most current comprehensive up-to-date sequence databases have been used to detect repetitive elements (<http://www.girinst.org/rebase>) present in these draft sequences,

WO 01/88089

PCT/US01/15674

-49-

hybridization of single copy probes to metaphase chromosomes has revealed that several probes contain previously unrecognized repetitive sequences. This was determined by documenting hybridization of a probe to the homologous chromosomal band where it is known to be mapped as well as other locations not found in the draft genome sequence.

The draft genome sequence is incomplete, with ~90% of the euchromatic genome having been sequenced (International Human Genome Sequencing Consortium, Initial Sequencing and Analysis of the Human Genome, *Nature*, 409:860-922, (2001). It was anticipated that some repetitive sequence families, especially those present among the missing sequences, would have not been detected. Despite screening for known repetitive sequences, several euchromatic single copy probes appear to contain homologs of repetitive sequence families that are predominantly found in multiple copies on the short arms of acrocentric chromosomes (chromosomes 13, 14, 15, 21, and 22). These probes included three sequences derived from the Down Syndrome critical region on chromosome 21 (amplified with SEQ ID Nos. 504/505, 506/507 and 508/509), two sequences from chromosome 16p13.1 that straddle the site of chromosomal inversion in Acute Myelogenous Leukemia, Type M4 (amplified with SEQ ID Nos. 536/537 and 538/539).

Such chromosomal domains, termed nucleolar organizer regions, are known to contain thousands of copies of the ribosomal RNA cistrons arranged in long tandem arrays (Sylvester et al., *Hum. Genet.*; 73:193-8 (1986)). The human genome draft genome sequence is devoid of contiguous sequences from these chromosomal regions. The International Sequencing Consortium eliminated clones containing these and other tandemly repeated sequences (e.g., heterochromatin) from consideration early in the sequencing effort, since it was recognized that any assembly of sequences from such clones would be ambiguous, and thus unreliable. Other sequences distinct from, but co-localizing with, ribosomal RNA genes would most likely also be tandemly arranged in chromosomal nucleolar organizer regions. Because of the lack of sequence information from these intervals, the distribution of sequences within them that are related to sequences elsewhere in the genome has not been previously appreciated. While single copy FISH with these probes demonstrated localization to the expected euchromatic

WO 01/88089

PCT/US01/15674

-50-

intervals, additional significant hybridization to the short arms of several acrocentric chromosomes was seen. Addition of C₁DNA prior to hybridization removed cross-hybridization to these repetitive sequences. These additional signals are consistent with tandemly organized multiple copies of sequences related to the probe on short arms of acrocentric chromosomes.

In addition, hybridization of presumed single copy probes to interspersed repetitive sequence families was detected, despite the fact that these probes were filtered for repetitive sequences using RepeatMasker software. One probe, mapping to chromosome 17p11.2, within the interval commonly deleted in Smith-Magenis syndrome (amplified with SEQ ID No. 586/587), was found to cross-hybridize with interspersed repeats. Sequences close to the translocation breakpoint at chromosome 9q34 (amplified with SEQ ID Nos. 518/519 and 526/527) also potentially contain interspersed repetitive sequences, based on hybridization of combinations of these probes with another probe from this region. The hybridization signals at the mapped locations for these probes were not stronger than those observed for the cross-hybridizing sequences, nor were the cross-hybridizing sequences removed by increasing the stringency during the washing procedures. This suggests that the probes contain previously unrecognized repetitive sequence families, rather than highly divergent copies of known interspersed repeats (whose failure to be recognized by RepeatMasker software would have led to their inadvertent inclusion in the designed probe).

Although all of these probes appear to contain members of previously unrecognized repetitive sequence families, the probes themselves are likely to be composed of both single copy and repetitive sequences. It is feasible to separate these sequence components by iterative hybridization of different PCR-generated subsets of each probe sequence to chromosomal DNA. However, since each entire probe sequence is known, the sequence can be added to the repeat sequence database used to develop new, additional single copy probes. The probe sequences are not lengthy (some interspersed repeat families are, in fact, longer, e.g., L1, than the longest single copy probe); therefore, minimal additional computational overhead is incurred by addition of these sequences to the database of human repetitive sequence families. The addition of these previously unknown repetitive sequence family members results in a more

WO 01/88089

PCT/US01/15674

-51-

comprehensive repetitive sequence database, that in turn improves the design of single copy probes. Single copy probes subsequently designed using the larger repeat sequence database, will not contain these novel repetitive sequences. This heuristic algorithm improves the purity of single copy sequences in single copy probes.

Generally speaking, the invention thus provides a method of determining the existence of previously unknown repeat sequence families in a genome. This method involves reacting a labeled, putative single copy nucleic acid probe with the genome, and causing the probe to hybridize. If the probe hybridizes at more than three different locations (and preferably at more than ten different locations), then it is likely that a new, previously unknown repeat sequence has been found.

All references cited above are expressly incorporated by reference herein. In addition, the subject matter of Disclosure Document #471449 filed March 27, 2000, is also incorporated by reference herein.

WO 01/88089

PCT/US01/15674

-52-

We claim:

1. A nucleic acid hybridization probe comprising a labeled, single copy nucleic acid which will hybridize to a deduced single copy sequence interval in target nucleic acid of known sequence, said nucleic acid probe having a length of at least about 50 nucleotides.

2. The probe of claim 1, said probe including a plurality of different, labeled nucleic acids each of which will hybridize to respective deduced single copy sequence intervals in said target nucleic acid, each of said nucleic acid probes having a length of at least about 50 nucleotides.

3. The probe of claim 1, said nucleic acid probe having a length of at least 100 nucleotides.

4. The probe of claim 3, said nucleic acid probe having a length of at least about 2000 nucleotides.

5. The probe of claim 1, said target nucleic acid being selected from the group consisting of DNA, RNA and mRNA.

6. The probe of claim 5, said target nucleic acid being DNA.

7. The probe of claim 1, said nucleic acid probe being single stranded.

8. The probe of claim 1, said probe being essentially free of blocking nucleic acid sequences which will hybridize to repeat sequences within the genome of which said target nucleic acid is a part.

WO 01/88089

PCT/US01/15674

-53-

9. The probe of claim 1, said nucleic acid probe being labeled with a label selected from the group consisting of fluorochrome-responsive labels, fluorochromes, colorimetric chemical, conjugated proteins, antibodies, antigens, and mixtures thereof.

10. The probe of claim 9, said nucleic acid probe being labeled with a fluorochrome-responsive label.

11. The probe of claim 1, there being at least about 80% sequence identity between said probe and a sequence which is a complement to said target sequence.

12. The probe of claim 11, said probe being complementary to said target sequence.

13. In a hybridization method including the steps of preparing a reaction mixture comprising a target nucleic acid sequence and a nucleic acid probe which hybridizes to at least a portion of said target nucleic acid sequence, and causing said probe to hybridize to said target nucleic acid sequence, the improvement which comprises using as said probe a labeled, single copy nucleic acid which hybridizes to a deduced single copy sequence interval in target nucleic acid of known sequence, said nucleic acid probe having a length of at least about 50 nucleotides.

14. The method of claim 13, said probe including a plurality of different, labeled nucleic acids each of which hybridizes to respective deduced single copy sequence intervals in said target nucleic acid, each of said nucleic acid probes having a length of at least about 50 nucleotides.

15. The method of claim 13, said nucleic acid probe having a length of at least 100 nucleotides.

WO 01/88089

PCT/US01/15674

-54-

16. The method of claim 15, said nucleic acid probe having a length of at least about 2000 nucleotides.

17. The method of claim 13, said target nucleic acid being selected from the group consisting of DNA, RNA and mRNA.

18. The method of claim 17, said target nucleic acid being DNA.

19. The method of claim 13, said nucleic acid probe being single stranded.

20. The method of claim 13, said probe being essentially free of blocking nucleic acid sequences which hybridizes repeat sequences within the genome of which said target nucleic acid is a part.

21. The method of claim 13, said nucleic acid probe being labeled with a label selected from the group consisting of fluorochrome-responsive labels, fluorochromes, colorimetric chemical, conjugated proteins, antibodies, antigens, and mixtures thereof.

22. The method of claim 21, said nucleic acid probe being labeled with a fluorochrome-responsive label.

23. The method of claim 13, said hybridization method selected from the group consisting of *in situ* hybridization, Southern blot, and other methods in which nucleic acid is immobilized.

24. The method of claim 13, there being at least about 80% sequence identity between said probe and a sequence which is a complement to said target sequence.

WO 01/88089

PCT/US01/15674

-55-

25. The method of claim 24, said probe being complementary to said target sequence.

26. A method of developing a hybridization probe for a target nucleic acid sequence forming a part of a genome, said method comprising the steps of:
determining the sequence of at least one single copy sequence in said target nucleic acid sequence; and
developing a hybridization probe which hybridizes to at least a part of said single copy sequence.

27. The method of claim 26, including the steps of:
determining the sequence of said target nucleic acid sequence;
determining the repeat sequences found in said genome; and
comparing said sequence of said target nucleic acid sequence and said repeat sequences in order to determine said sequence of said at least one single copy sequence.

28. The method of claim 26, said probe developing step comprising the steps of obtaining at least a part of said single copy sequence, and purifying said part of said single copy sequence.

29. The method of claim 28, said purifying step comprising carrying out PCR.

30. The method of claim 26, including the step of labeling said hybridization probe.

31. The method of claim 26, target nucleic acid sequence and said probe being DNA.

WO 01/88089

PCT/US01/15674

-56-

32. The method of claim 26, said hybridization probe having at least about 80% sequence identity with said single copy sequence.

33. The method of claim 32, said hybridization probe being complementary to single copy sequence.

34. A DNA sequence having at least about 80% sequence identity with a member selected from the group consisting of SEQ ID Nos. 429-446.

35. The DNA sequence of claim 34, said sequence having at least about 90% sequence identity.

36. A DNA sequence selected from the group consisting of SEQ ID Nos. 429-446 and 480-613.

37. A DNA sequence having at least about 80% sequence identity with a PCR-amplified sequence produced using as PCR primers adjacent pairs of the sequences identified as SEQ ID Nos. 429-446 and 480-613, beginning with SEQ ID No. 429.

38. The sequence of claim 37, there being at least about 90% sequence identity.

39. A method of determining the existence of previously unknown repeat sequence families in a genome, comprising the steps of reacting a labeled, putative single copy nucleic acid probe with the genome, causing the probe to hybridize, and ascertaining if the probe hybridizes to the genome at more than three different locations as a determination of a new repeat sequence family.

40. The method of claim 39, including the step of ascertaining if the probe hybridizes at more than ten different locations.

WO 01/88089

PCT/US01/15674

-57-

41. The probe of claim 1, said nucleic acid derived from a duplicon or triplicon sequence interval.

42. The method of claim 13, including the step of selecting a single copy nucleic acid which will hybridize to a duplicon or triplicon sequence domain.

43. The method of claim 26, said determining step comprising the step of selecting said single copy sequence from a duplicon or triplicon sequence domain.

44. A method of determining a chromosome breakpoint, comprising the steps of:

providing a pair of separate, labeled, single copy nucleic acid probes of predetermined sequence and designed to respectively hybridize on opposite sides of said breakpoint;

reacting said pair of probes with a chromosomal target sequence containing said breakpoint, and causing the probes to hybridize to the target sequence; and

detecting said hybridized probes as a way of ascertaining the location of said breakpoint.

45. The method of claim 44, including the step of providing a plurality of separate, labeled, single copy nucleic acid probes of predetermined sequence designed to hybridize on one side of said breakpoint, and providing a plurality of separate, labeled, single copy nucleic acid probes of predetermined sequence designed to hybridize on the other side of said breakpoint.



Fig. 1

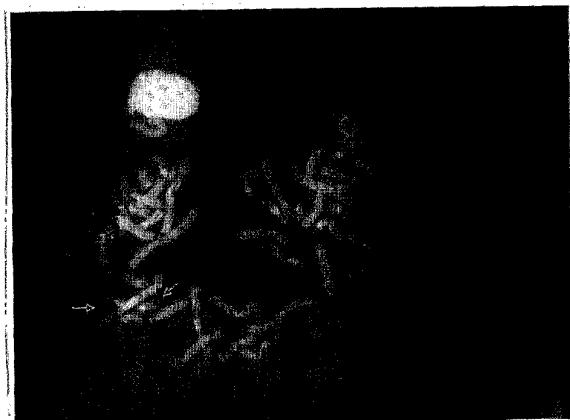


Fig. 2



Fig. 3



Fig. 4

WO 01/88089

PCT/US01/15674

3/8

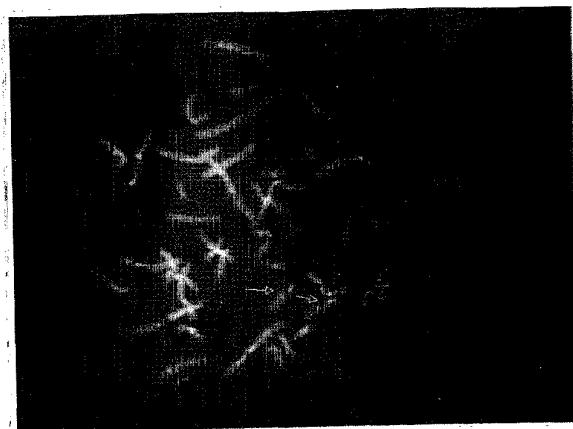


Fig. 5



Fig. 6

WO 01/88089

PCT/US01/15674

4/8

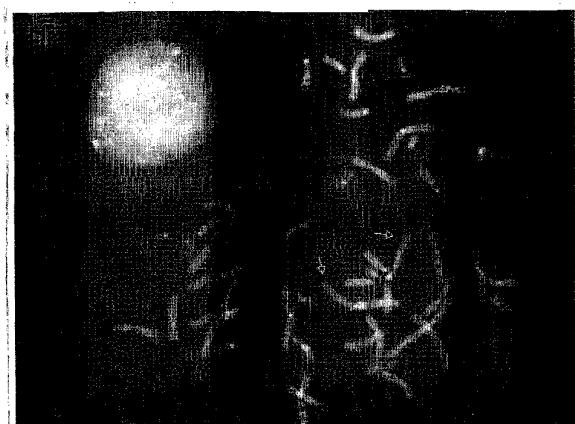


Fig. 7

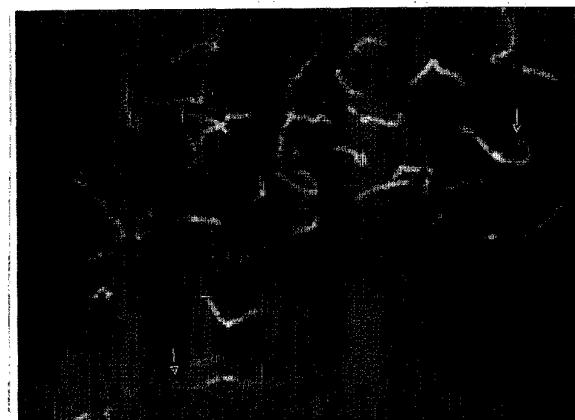


Fig. 8



Fig. 9

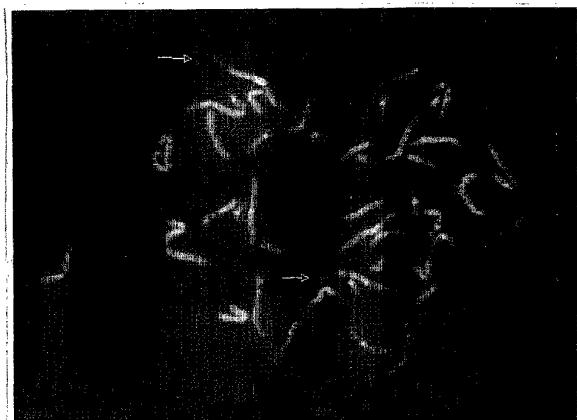


Fig. 10

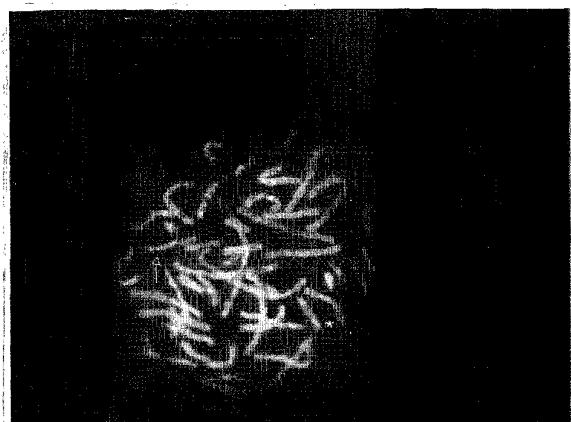


Fig. 11

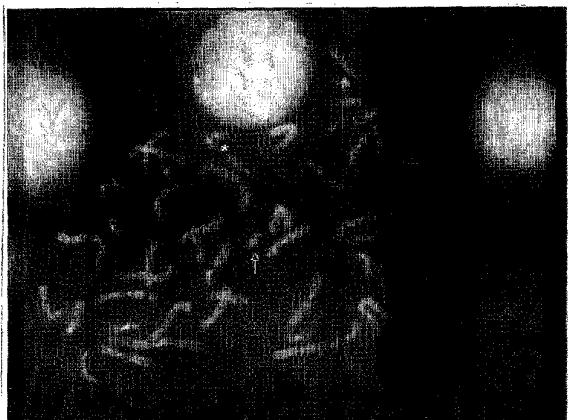
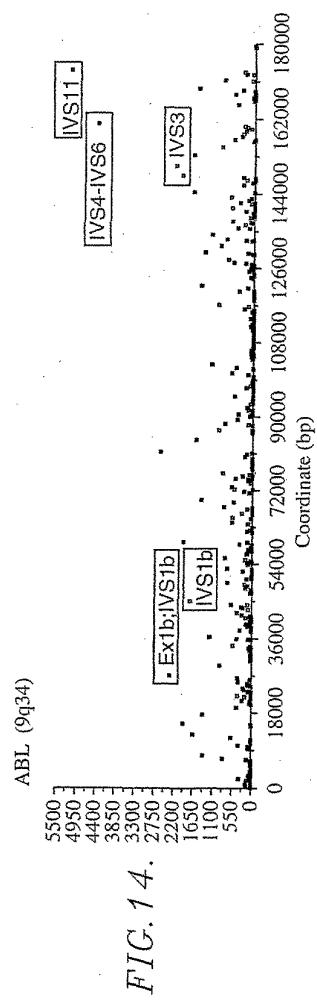
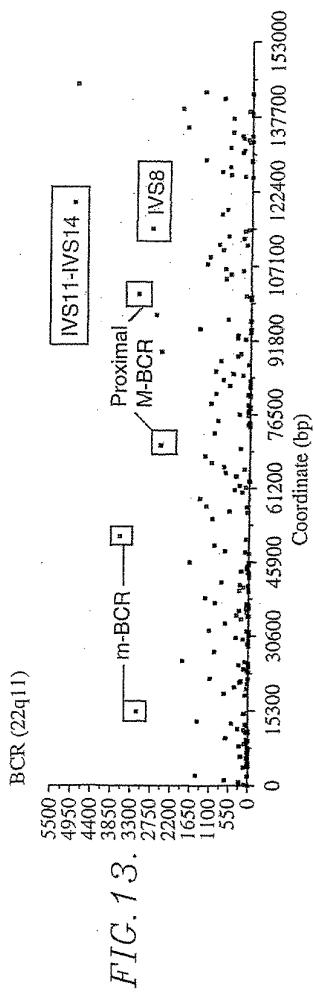


Fig. 12

WO 01/88089

PCT/US01/15674

7/8



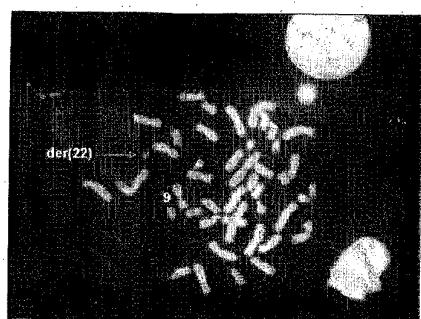


Fig. 15

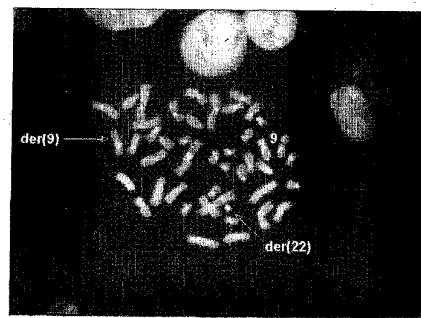


Fig. 16

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
22 November 2001 (22.11.2001)

PCT

(10) International Publication Number
WO 01/088089 A3(51) International Patent Classification⁵: C12Q 1/68, C12P 19/34, C07H 21/02, 21/04

(21) International Application Number: PCT/US01/15674

(22) International Filing Date: 15 May 2001 (15.05.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/573,080 16 May 2000 (16.05.2000) US
09/854,867 14 May 2001 (14.05.2001) US

(71) Applicant (for all designated States except US): CHILDREN'S MERCY HOSPITAL [US/US]; 2401 Gillham Road, Kansas City, MO 64108 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): KNOLL, Joan H., M. [US/US]; 12226 Gillette Road, Overland Park, KS 66213 (US); ROGAN, Peter K. [US/US]; 12226 Gillette Road, Overland Park, KS 66213 (US); CAZARRO, Patricia, M. [US/US]; 9005 Accuff Lane, Lenexa, KS 66215 (US).

(74) Agent: COLLINS, John, M.; Hovey, Williams, Timmons & Collins, Suite 400, 2405 Grand Boulevard, Kansas City, MO 64108 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MR, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SL, SK, SL, TI, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(88) Date of publication of the international search report:
4 July 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/088089 A3

(54) Title: SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING SAME

(57) Abstract: Nucleic acid (e.g., DNA) hybridization probes are described which comprise a labeled, single copy nucleic acid which hybridizes to a deduced single copy sequence interval in target nucleic acid of known sequence. The probes, which are essentially free of repetitive sequences, can be used in hybridization analyses without adding repetitive sequence-blocking nucleic acids. This allows rapid and accurate detection of chromosomal abnormalities. The probes are preferably designed by first determining the sequence of at least one single copy interval in a target nucleic acid sequence, and developing corresponding hybridization probes which hybridize to at least a part of the deduced single copy sequence. In practice, the sequences of the target and of known genomic repetitive sequence representatives are compared in order to deduce locations of the single copy sequence intervals. The single copy probes can be developed by any variety of methods, such as PCR amplification, restriction or exonuclease digestion of purified genomic fragments, or direct synthesis of DNA sequences. This is followed by labeling of the probes and hybridization to a target sequence.

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US01/15674
A. CLASSIFICATION OF SUBJECT MATTER		
IPC(7) : C12Q 1/68; C12P 19/34; C07H 21/02, 21/04 US CL : 435/6, 91.2; 536/23.1, 23.5, 24.1, 24.2, 24.3, 24.31, 24.33 According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) U.S. : 435/6, 91.2; 536/23.1, 23.5, 24.1, 24.2, 24.3, 24.31, 24.33		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) Please See Continuation Sheet		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X, P	US 6,222,029 B1 (EDWARDS et al) 24 April 2001, columns 21, 25-27, 31-32 and 49-50.	1-15, 17-33, 41-43
X	US 6,040,140 A (CROCE et al) 21 March 2000, column 9.	1-9, 11-13, 15-21, 23-33, 41-45
Y		-----
X	US 5,811,231 A (FARR et al) 22 September 1998, columns 6, 7 and 25.	10, 14, 22
Y	US 5,721,098 A (PINKEL et al) 24 February 1998, column 31.	1, 2, 5-14, 17-33, 41-43
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
<p>* Special categories of cited documents:</p> <ul style="list-style-type: none"> *T* later document published after the international filing date or priority date and in conflict with the application but cited to understand the principle or theory underlying the invention *A* document defining the general state of the art which is not considered to be of particular relevance *B* earlier application or patent published on or after the international filing date *I* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed 		
Date of the actual completion of the international search 25 March 2002 (25.03.2002)	Date of mailing of the international search report 09 MAY 2002	
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Carlis Myers Telephone No. 703-305-0196 	

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT	International application No. PCT/US01/15674
Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet) This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons: 1. <input type="checkbox"/> Claim Nos.: because they relate to subject matter not required to be searched by this Authority, namely: 2. <input type="checkbox"/> Claim Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically: 3. <input type="checkbox"/> Claim Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).	
Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet) This International Searching Authority found multiple inventions in this international application, as follows: Please See Continuation Sheet	
1. <input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims. 2. <input type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee. 3. <input checked="" type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1-33, 39-45 4. <input type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:	
Remark on Protest <input type="checkbox"/> The additional search fees were accompanied by the applicant's protest. <input checked="" type="checkbox"/> No protest accompanied the payment of additional search fees.	

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998)

INTERNATIONAL SEARCH REPORT	International application No. PCT/US01/15674
<p>BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.</p> <p>Group 1, claim(s) 1-33, drawn to probes, methods of making probes and hybridization methods of using probes.</p> <p>Groups 2-151, claim(s) 34-38, drawn to nucleic acids selected from the group consisting of SEQ ID NO: 429-446 and 480-613, respectively. For example, if group 2 is elected, then claims 34-38 will be examined to the extent that they are limited to nucleic acids consisting of SEQ ID NO: 429. Upon election of one of the groups, please specify the SEQ ID NO corresponding to the requested group.</p> <p>Group 152, claim(s) 39-43, drawn to a method of identifying a previously unknown repeat sequence.</p> <p>Group 153, claims 44 and 45, drawn to a method of determining a chromosome breakpoint using 2 single copy probes.</p> <p>The inventions listed as Groups 1-153 do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Pursuant to 37 C.F.R. 1.475(d), this Authority considers that where multiple products and processes are claimed, the first recited product, method of making the product and method of using the product shall constitute the Main Invention. Further, pursuant to 47 C.F.R. 1.475(d), it considers that any subsequently recited products and methods do not share a special technical feature with the main invention or any such other inventions. In addition, it is noted that each of the nucleic acids of groups 2-151 are distinct from one another and do not share a special technical feature because each nucleic acid consists of a unique nucleotide sequence and thereby each nucleic acid hybridizes to distinct regions of the genome and each nucleic acid has a distinct functional activity. Furthermore, the methods of groups 152 and 153 do not share the same technical feature because each method has a different objective, involves performing different method steps and utilizes different reagents (i.e., the method of group 153 requires the use of 2 probes which hybridize to opposite sides of a breakpoint, whereas the method of group 152 requires the use of a single probe and requires determining whether said probe hybridizes to at least 3 regions of the genome).</p>	
Continuation of B. FIELDS SEARCHED Item 3: WEST-US, EP, JP, WO Patents, DIALOG: Medline, CA, BIOSIS, EMBASE, SCISEARCH search terms: single-copy, repeat sequences, repeat-free, blocking nucleic acids, breakpoint, translocation	

Form PCT/ISA/210 (second sheet) (July 1998)

フロントページの続き

(51) Int.Cl.⁷ F I テーマコード(参考)
 G 01N 33/58 G 01N 33/58 A

(81) 指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(72) 発明者 ローガン ピーター ケイ。
 アメリカ合衆国 カンザス州 66213 オーバーランド パーク ジレット ロード 122
 26

(72) 発明者 カザッコ パトリシア エム。
 アメリカ合衆国 カンザス州 66215 レネクサ アックフ レーン 9005

F ターム(参考) 2G045 AA35 BB24 BB46 BB48 BB50 DA13 FB02 FB05 FB07 FB12
 GC15
 2G054 BB03 BB08 CA22 CE02 EA03 GA04
 4B024 AA11 CA09 HA14
 4B063 QA01 QA12 QA18 QQ42 QR32 QR55 QR62 QS34 QS36 QX02

【要約の続き】

片の制限酵素やエクソヌクレアーゼ処理、または直接核酸を合成する等、様々な方法により調製することができる。その後で、シングルコピー核酸断片は、潜在的に混入している繰り返し配列を除去するために、例えば、電気泳動や高速液体クロマトグラフィーにより精製する；この過程は、関係のないゲノム配列の偽陽性によるハイブリダイゼーションや検出を排除するために行なうことが、特に好ましい。

それから、プローブ断片は、リコンビナントDNAベクターにクローニングしたり、直接ラベルを入れたりできる。プローブは、修飾された、または直接標識されたヌクレオチドを使って、ニックトランスレーションにより、標識されるのが好ましい。その後で、好ましくはマイクロスライド上に固定された染色体のプレパラート、そうでなければ、メンプラン、スライド、DNAチップ、または他の物体上に固相化した精製核酸と、標識されたプローブは変性され、ハイブリダイゼーションする。プローブは、米国特許5985549や5447841号明細書に記載されているように、従来からの蛍光染色in situハイブリダイゼーション(FISH)により、染色体へハイブリダイズする；または、米国特許5110920や5273881号明細書に記載された技術により、固相化した核酸と結合させることもできる。プローブのシグナルは、蛍光、免疫学的、または酵素反応溶液を使用する等、色々な方法により視覚化することができる。

本発明のプローブを使用すると、従来のin situハイブリダイゼーションでは、達成できなかった解像能で、染色体の切断点を、さらに正確に決定することができる。そのような解析を行うに当たり、最初に用いるプローブ群は、切断点の反対側にあると考えられている領域を基に、調製することができる。切断点を決定する最初の実験の後で、シングルコピーの手法を使って、切断点の近傍に新たにプローブを設計することができる。この方法で、切断点の正確な領域を決定することができる。

シングルコピーと推測されるプローブを使って、今まで未知であったゲノム中の繰り返し配列の存在を決定することができることが明らかとなった。従来未知であった繰り返し配列のファミリーは、連続した核酸配列のデータベースの中に追加され、その後、シングルコピープローブの設計に利用することができるようになる。

本発明のプローブは、標的配列の長くなった分、より強固にハイブリダイゼーションできるような、ゲノム中で2倍化または3倍化している配列を含んでいてもよいことが、明らかとなっている。また、このようなデュープリコンまたはトリプリコンは、例えば、シングルコピープローブを使用して確認することができ、上市されているプローブを用いた場合は困難である。