US006828097B1

(12) **United States Patent**　　(10) **Patent No.:**　**US 6,828,097 B1**

Knoll et al.　　(45) **Date of Patent:**　**Dec. 7, 2004**

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

(75) Inventors: **Joan H. M. Knoll**, Overland Park, KS (US); **Peter K. Rogan**, Overland Park, KS (US)

(73) Assignee: **The Childrens Mercy Hospital**, Kansas City, MO (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/573,080**

(22) Filed: **May 16, 2000**

(51) **Int. Cl.**[7] **C12Q 1/68**; C07H 21/02; C07H 21/04

(52) **U.S. Cl.** **435/6**; 536/24.3; 536/24.31; 536/24.32; 536/23.1; 536/25.4; 435/91.2

(58) **Field of Search** 435/6; 536/24.3, 536/24.31, 24.32, 23.1, 25.4

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,447,841	A		9/1995	Gray et al. 435/6
5,721,098	A	*	2/1998	Pinkel et al. 435/6
5,756,696	A		5/1998	Gray et al. 435/6
5,811,231	A	*	9/1998	Farr et al. 435/6
6,040,140	A	*	3/2000	Croce et al. 435/6
6,222,029	B1	*	4/2001	Edwards et al. 536/24

OTHER PUBLICATIONS

DeRisi et al. Nature Genetics. Dec. 1996. 14: 457–460.*
Sakuma, GenBank Accession No. ABOO1872, Feb. 13, 1998.*
Urano, GenBank Accession No. AB002058, Aug. 29, 1997.*
Inoue et al. GenBank Accession No. AB002135, Jul. 23, 1999.*
Watanabe, GenBank Accession No. AB003723, Feb. 25, 1998.*
Kamei, GenBank Accession No. AB003592, Jul. 19, 1999.*
Ohno, GenBank Accession No. AB000114, Feb. 5, 1999.*
Shimomura, GenBank Accession No. AB000095, Mar. 4, 1998.*

Satoh, GenBank Accession No. AB000276, Nov. 6, 1997.*
Fukunaga, GenBank Accession No. AB000409, Feb. 5, 1999.*
Yokouchi, GenBank Accession No. AB000520, Sep. 26, 1997.*
Fukuta, GenBank Accession No. AB003791, Feb. 14, 1998.*
Okumoto, GenBank Accession No. AB004546, Jul. 11, 1998.*
Ikeda, GenBank Accession No. AB000812, Feb. 20, 1999.*
Ishihara, GenBank Accession No. AB003333, Feb. 26, 1999.*
Rockman et al. Australian Journal of Medical Science. May 1994. 15: 56–57.*

* cited by examiner

Primary Examiner—Carla J. Myers
(74) *Attorney, Agent, or Firm*—Hovey Williams LLP

(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.

8 Claims, 6 Drawing Sheets

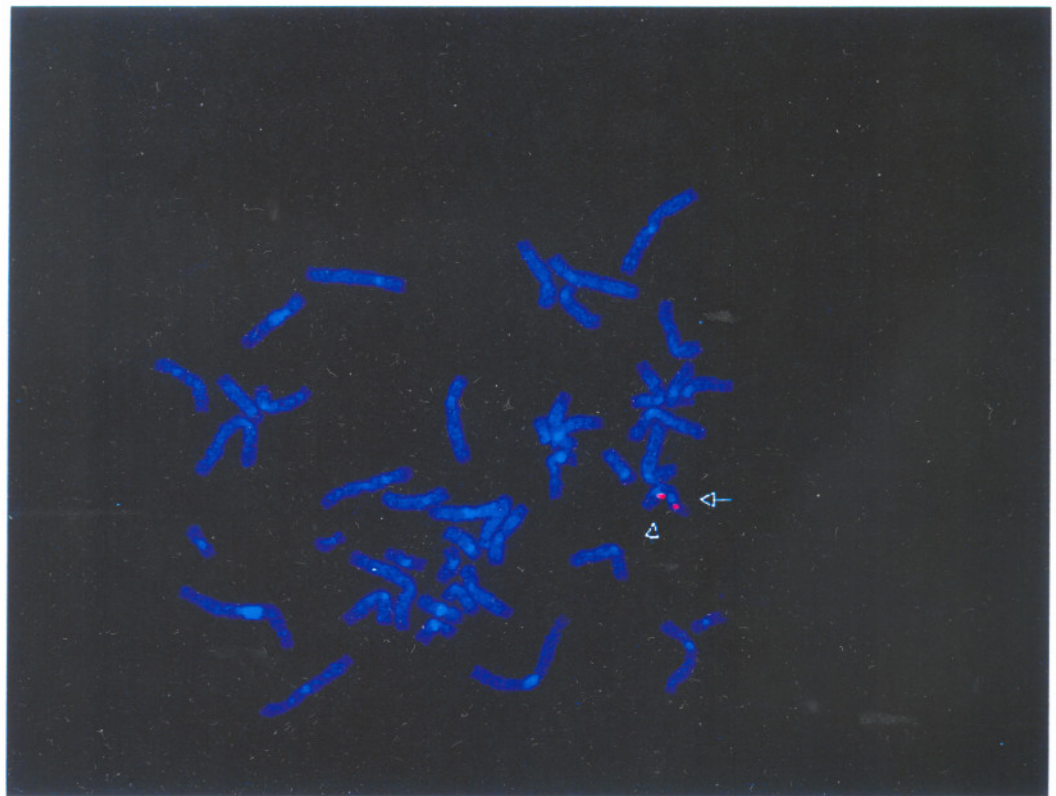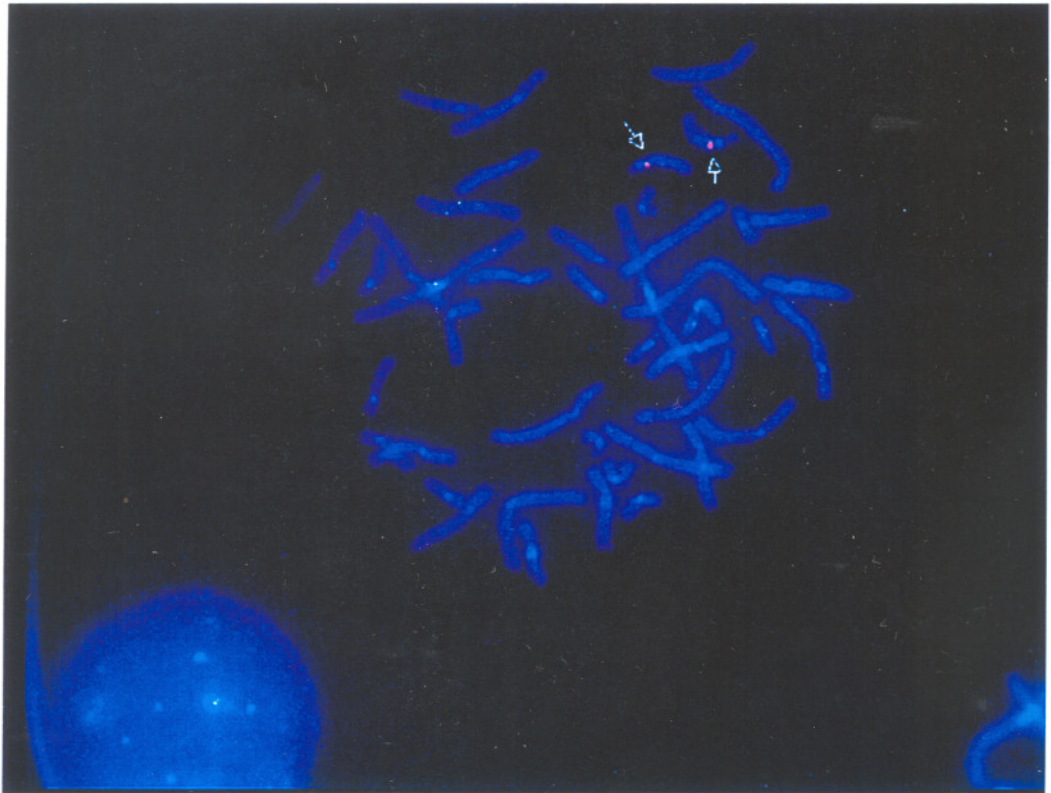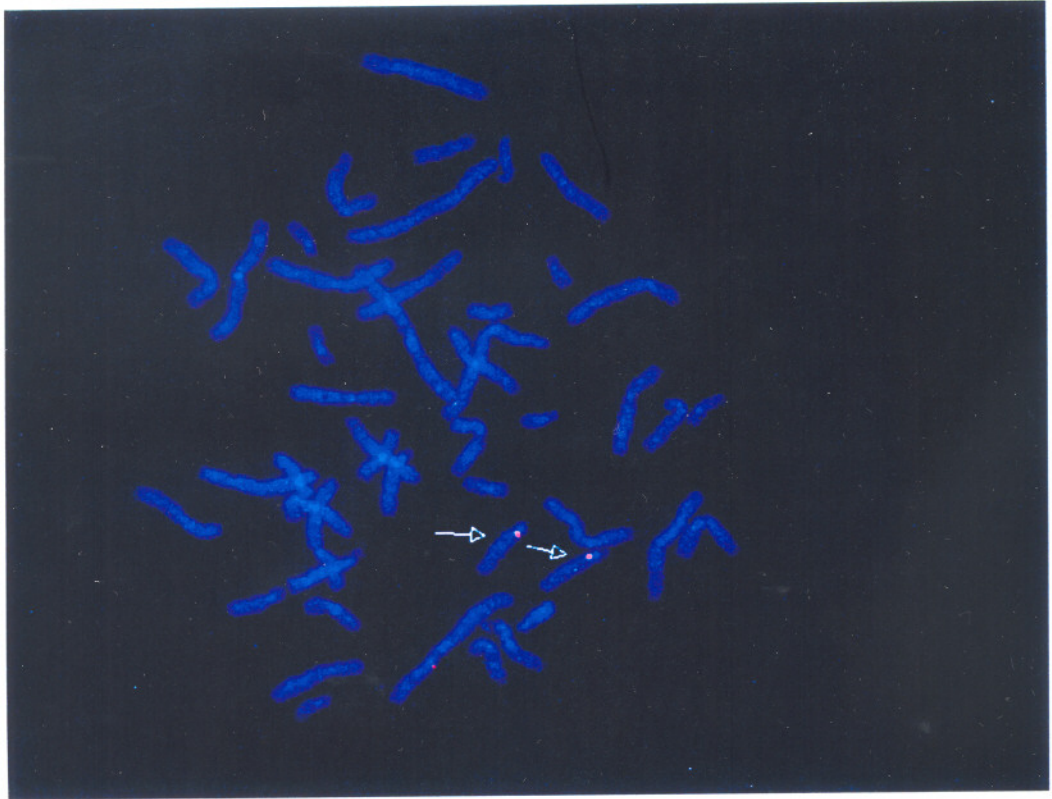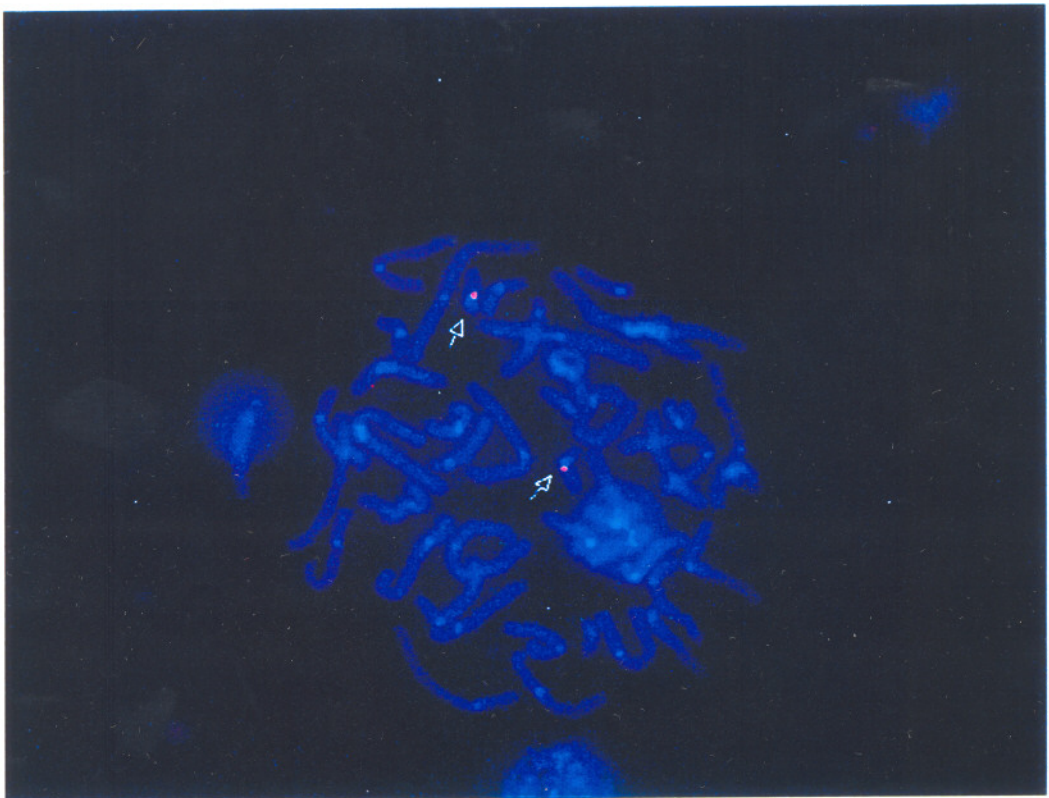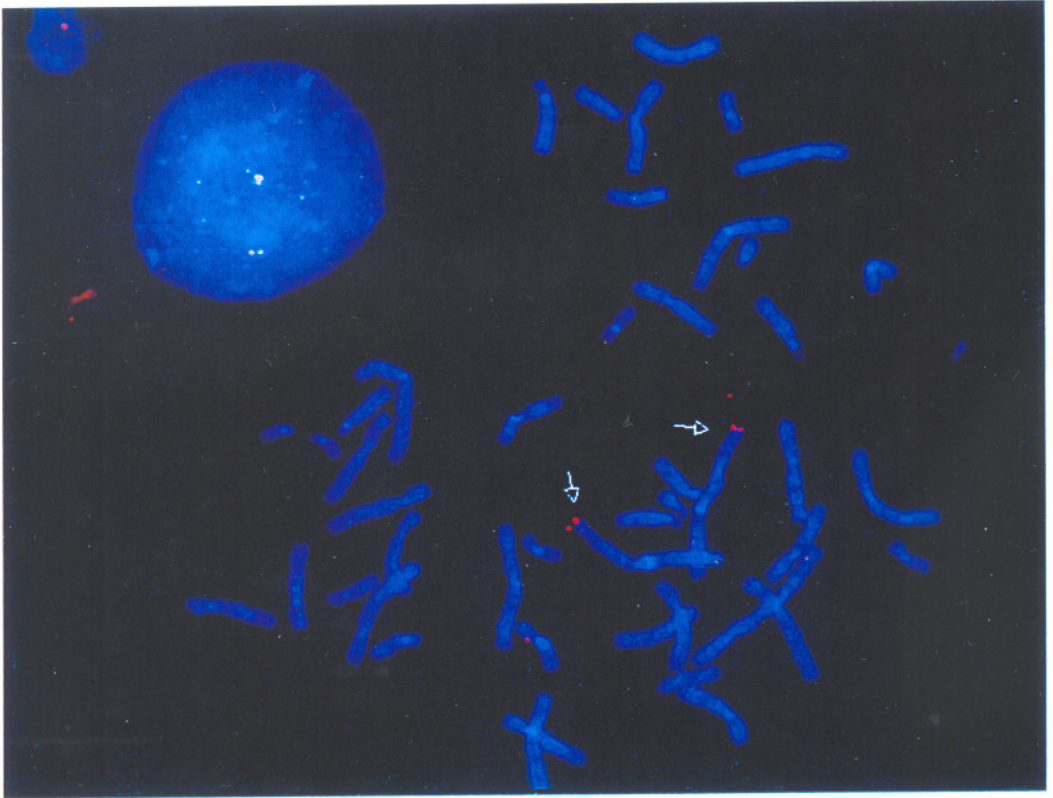**(6 of 6 Drawing Sheet(s) Filed in Color)**
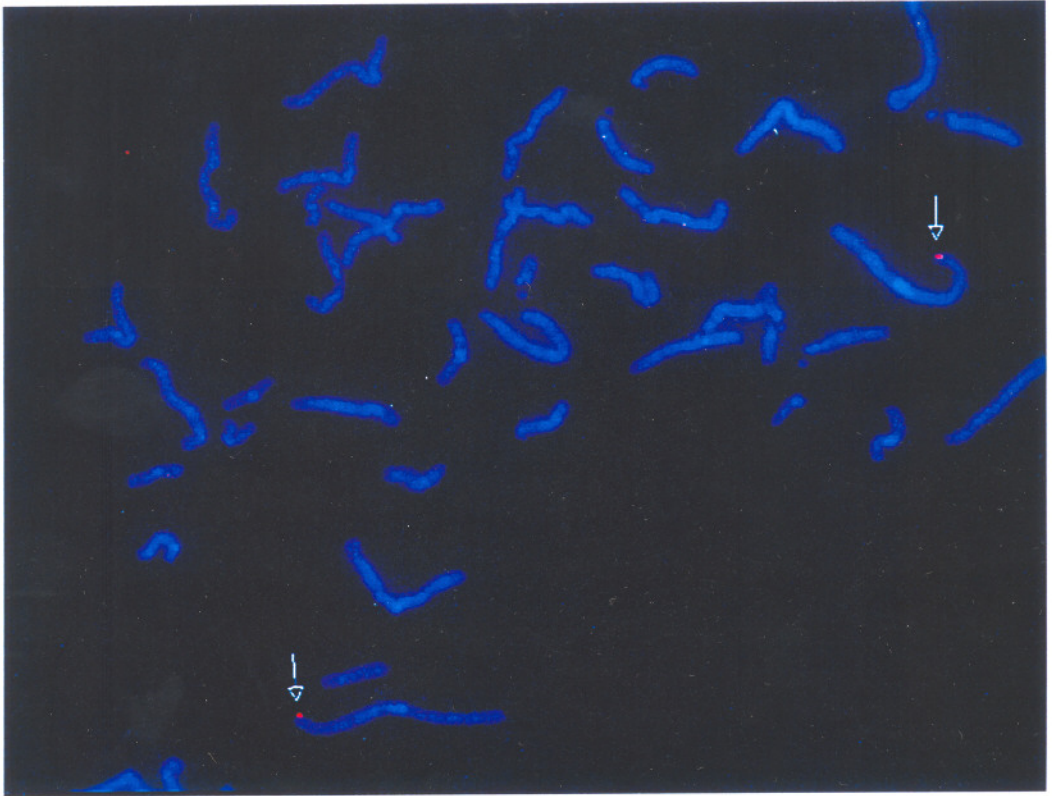
Fig. 1

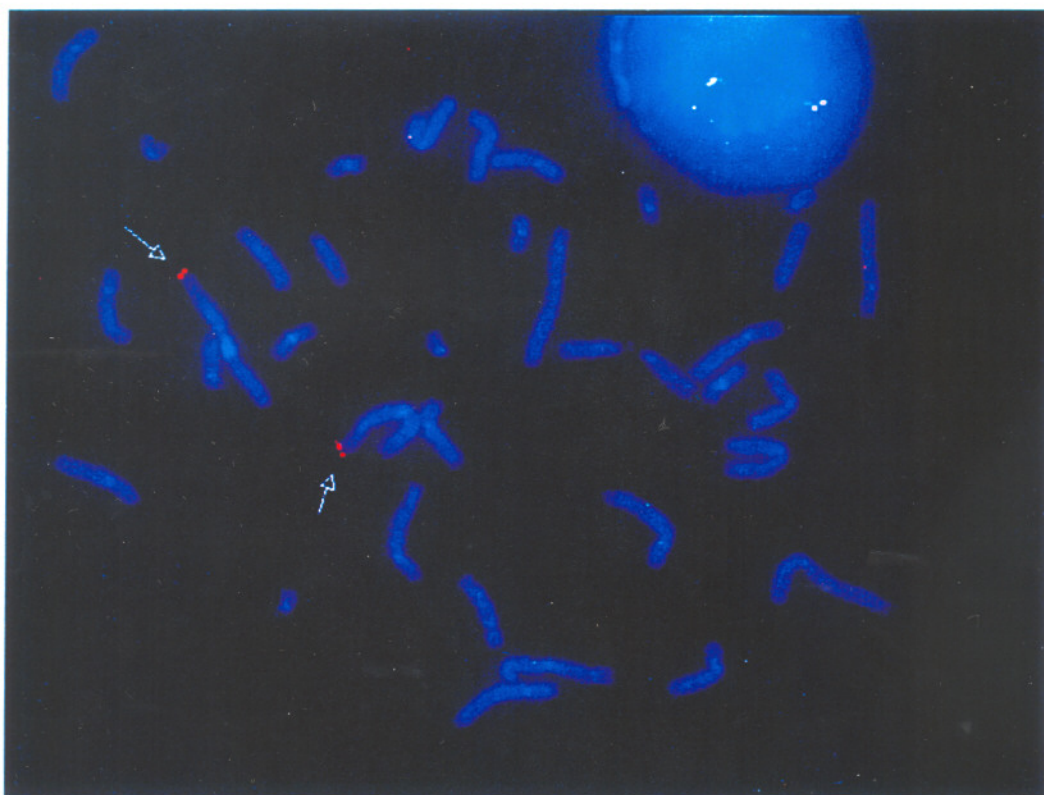Fig. 2

Fig. 3

Fig. 4

Fig. 5
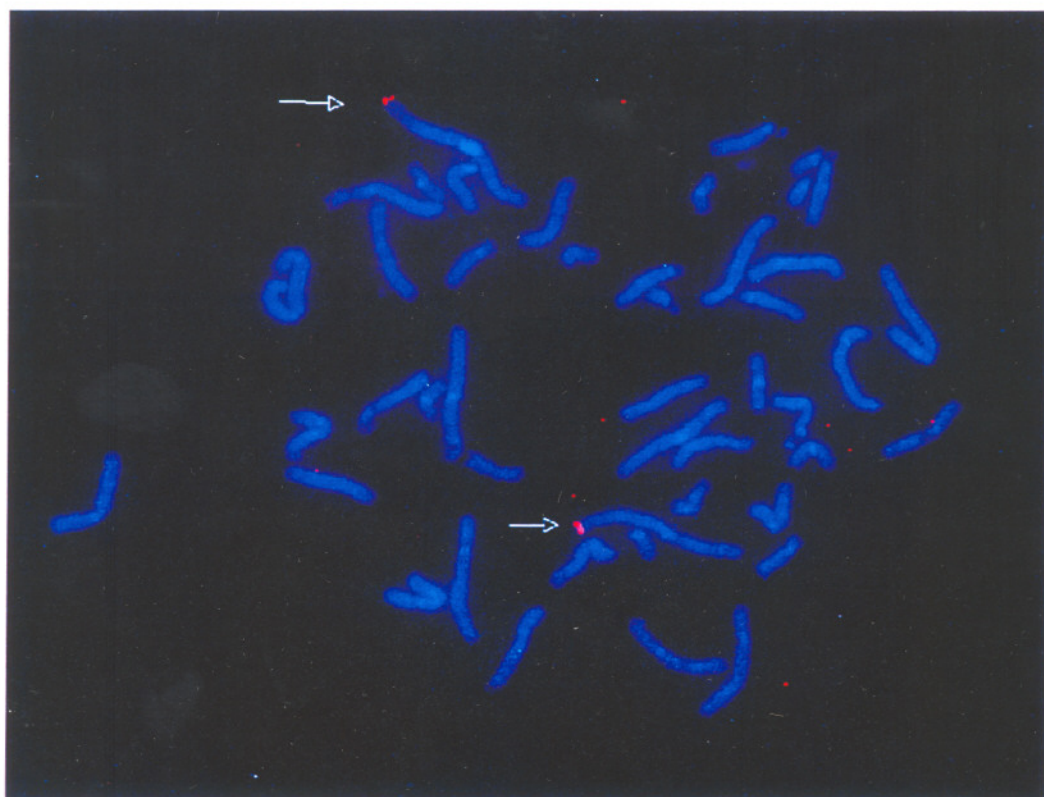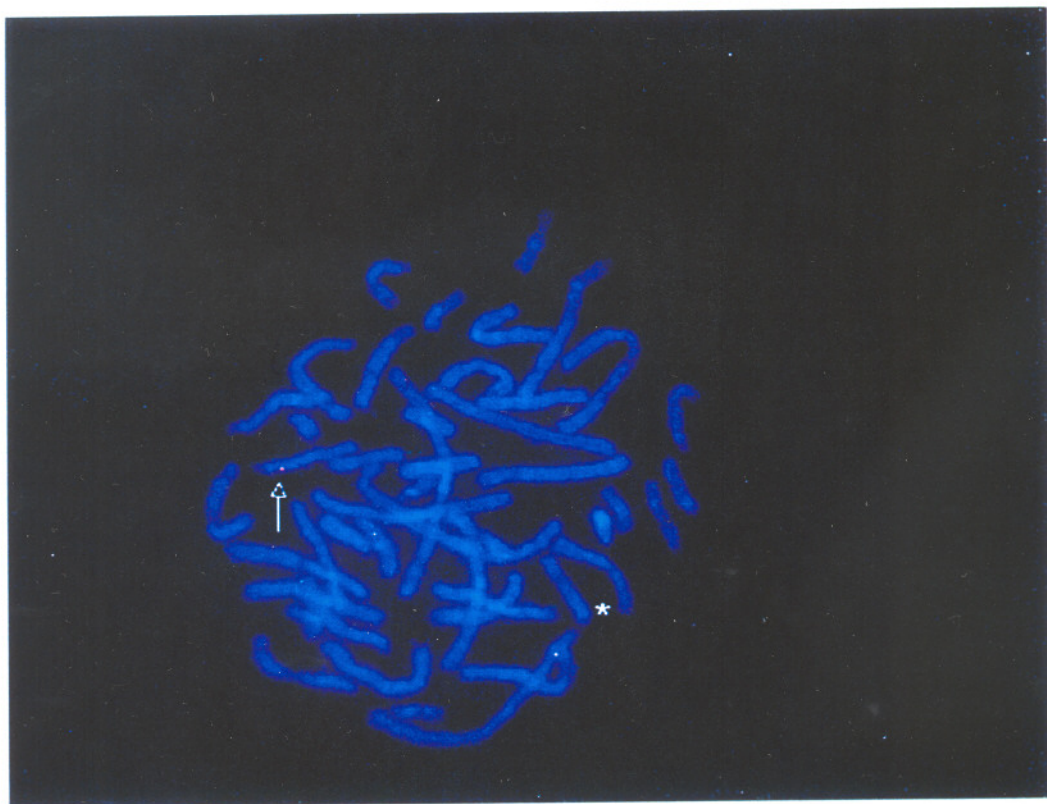

Fig. 6

Fig. 7

Fig. 8

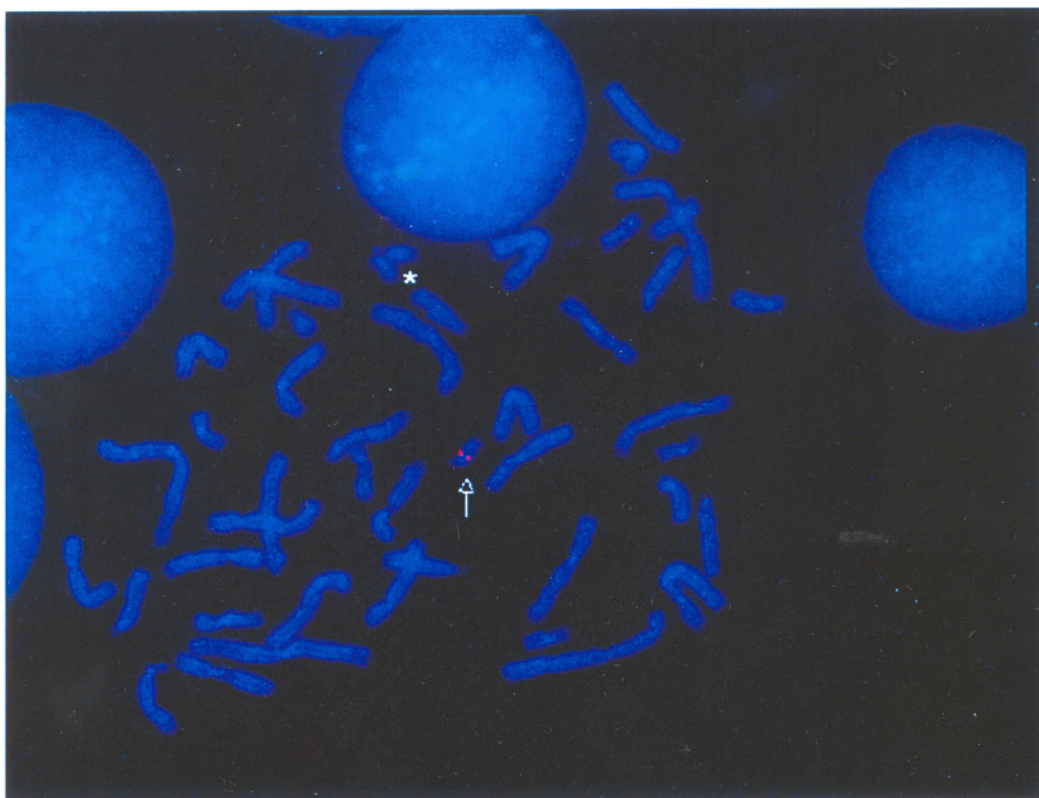Fig. 9

Fig. 10

Fig. 11

Fig. 12

SINGLE COPY GENOMIC HYBRIDIZATION PROBES AND METHOD OF GENERATING SAME

SEQUENCE LISTING

A printed Sequence Listing accompanies this application, and has also been submitted with identical contents in the form of a computer-readable ASCII file on a floppy diskette.

BACKGROUND OF THE INVENTION

1. 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 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.

2. Description of the Prior Art

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.

Hybridization probes have been developed in the past for chromosome analysis and diagnosis of abnormalities. The probes comprise clone or amplified genomic sequences or cDNA. For example, U.S. Pat. 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 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 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.*, 100:472476 (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 U.S. Pat. Nos. 5,447,841 and 5,756,696.

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

BRIEF DESCRIPTION OF THE DRAWINGS

The application file contains at least one drawing executed in color. Copies of this patent and patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

FIGS. 1–12 are respective CCD camera images of FISH experiments wherein 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;

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;

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;

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;

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;

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;

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; and

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.

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 made up 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. 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.

Repetitive sequences occur in multiple copies in the haploid genome. The number of copies can range from two 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); 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, J., et al., *Nucleic Acids Research*, 12(1):387 (1984)), BLASTP, BLASTN and FASTA (Altschul, S. F. et al., *J. Molec. Biol.*, 215:403410 (1990). The BLASTX program is publicly available from NCBI and other sources (BLAST Manual, Altschul, S. et al., NCBI, NLM, NIH, Bethesda, Md. 20894, Altschul, S. F. et al., *J. Molec. Biol.*, 215:403410 (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 bout 500 nucleotides, and 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.

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 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).

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. (USA)*, 85:244 (1988); Altschul et al., 1990) are generally not as comprehensive as the methods of Smith et al. 1981 and Needlemean et al., 1970. However, they are faster than these methods.

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*, at 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 of sequences will decrease as hybridization sensitivity increases. At high sensitivity and low background, smaller and less complex probes can be used.

With current hybridization techniques, it is possible to obtain reliable, easily detectable signals with relatively small probes in accordance with the invention. A readily detectable signal was obtained with a probe on the order of 2.3 kb in length, using FISH technology. This sensitivity of the present method is improved compared to the prior art (U.S. Pat. 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.

One application of the use of multiple fragment probes is in the detection of translocations involving two different chromosomes. Proportionately increasing the complexity of the probe also permits analysis of multiple compact regions of the genome simultaneously. The portion of the probe targeted to one side of the break point can be labeled differently from that targeted to the other side of the break point so that the fused translocated chromosome is detected by both labels and is distinguishable from the intact chromosome.

Once appropriate single copy sequences have been identified, PCR is preferably used for amplifying the appropriate DNA to obtain probes. The 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 (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 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. Pat. 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 primer pairs are generally excluded from consideration as 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.

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, NH_4-acetate, K-glutamate, NH_4Cl, 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 $MgCl_2$, Mg-acetate, and the like. The amount of Mg^{+2}-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 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.

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.

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.

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.

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.

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 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 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. Pat. Nos. 5,612,473; 5,602,756; 5,538,871; and 5,475,610.

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 L E 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, Mass.) 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 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. Pat. 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 μl reaction) utilize endonuclease-free DNA polymerase I (Roche Molecular Biochemicals) and DNaseI (Worthington Chemical). Each fragment is combined with DNA polymerase (20 units/microgram DNA), DNaseI (10 microgram/100 μl reaction), labeled 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), 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 also 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. Pat. Nos. 5,985,549; 5,447,841; 5,756, 696; 5,869,237. Based on early work of Gall and Pardue (PNAS 63:378–383, 1969), isotopic in situ hybridization was established in the 1970s (see Gerhard et al., PNAS 78:3755–3759,1981 and Harper et al., PNAS 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. Pat. No. 5,985,549. The technique relies on the formation of duplex 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.

The present method 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 acid. The sample may be obtained from the fluids or tissues of a mammal, preferably human, which are suspected of being afflicted with a disease or disorder either from a biopsy or post-mortem, or 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 calfserum for 72 hours at 37° C. Ethidium bromide (100 ug/10 ml final) is added 1½ hours prior to harvest. Colcemid (1 ug/10 ml 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., 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 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.

Immediately prior to chromosomal denaturation in the in situ hybridization procedure, the dried or stored chromosome preparations can be pretreated in prewarmed 2×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).

When the target nucleic acid sequence is DNA, DNA in the sample can be denatured by heat or alkali. [See Harper et al., PNAS USA, 78, 4458–60 (1981), for alkali denaturation and Gall et al., PNAS USA, 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×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×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/2×SSC at 39–45° C., then 2×SSC at 3945° C., followed by a 15–30 minute wash at room temperature in 1×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). 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). 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. Pat. 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 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 are provided by way of illustration and nothing therein should be taken as a limitation upon the overall scope of the invention.

EXAMPLE 1

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/seq/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 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-, tetra-, and pentanucleotide 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.

TABLE 1

HIRA position (bp)				Position (bp) in Seq.Listing Corresponding to HIRA Match	
Begin	End	Repeat Family	SEQ ID NO.	Begin	End
798411	798434	(AC)	452	1	24
798983	799395	MLT2A1	444	1	434
801257	801348	CHESHIRE_A	420	132	223
801367	801729	L1ME_ORF2	425	757	1089
801746	802032	Alu-Jb	2	2	289
802033	802308	L1ME_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	MLT1E2	106	198	260
805011	805077	MLT1E1	105	420	484
805110	805211	L1PBA_5	359	103	204
805212	805862	L1PBA_5	359	1089	1738
805933	805989	Alu-J	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

TABLE 1-continued

HIRA position (bp)				Position (bp) in Seq.Listing Corresponding to HIRA Match	
Begin	End	Repeat Family	SEQ ID NO.	Begin	End
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
814780	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	HAL1	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
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

TABLE 1-continued

	HIRA position (bp)			Position (bp) in Seq.Listing Corresponding to HIRA Match	
Begin	End	Repeat Family	SEQ ID NO.	Begin	End
850114	850249	L1P_MA2	348	5447	5610
850250	850761	L1	59	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	(GAAAAA)	482	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	LIP_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
888460	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	LOR11	366	395	422
888253	888318	LINE2	362	2440	2505
888385	888548	LINE2	362	2579	2739
888865	888893	LOR11	366	394	422
889006	889296	Alu-Jb	2	5	290
889446	889548	Alu-Jo	2	188	290
889549	889677	L1PB3	358	770	897
889842	890133	Alu-Sq	2	1	290
890515	890797	Alu-Sz	2	1	283
890858	890972	L1ME2	341	769	885
890986	891024	LOR11	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

TABLE 1-continued

	HIRA position (bp)			Position (bp) in Seq.Listing Corresponding to HIRA Match	
Begin	End	Repeat Family	SEQ ID NO.	Begin	End
897067	897299	MER20	317	2	217
897492	897624	Alu-FLA	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 Wis.). 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 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 primerpairs. The top ranked primers were selected for synthesis, as set forth in the following Table 2. These primers were commercially produced (Oligos, Etc., Wilsonville, Oreg.).

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	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

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:

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

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.

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 LE agarose gel (FC Bioproducts) in 1×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 5000 g 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 precipitated DNA was then centrifuged, rinsed with cold 70% EtOH, air dried and 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.

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.

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. Pat. 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 purified repetitive DNA by adding 125 ng (or 20 μl) of labeled probe to 10 micrograms of C₀t 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.

Subsequently, the denatured probes with or without purified repetitive DNA (i.e., C₀t 1) were mixed with 1 V prewarmed hybridization solution (comprised of 4×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×SSC. A coverslip was placed over the probe hybridization mixture on 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, 1994). Chromosomal DNA was counterstained with DAPI. The cell preparations on microscope slides were then mounted in antifade solution (Vectashield, Vector Laboratories) 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.

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.

sequences; 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 15q11q13 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.

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.

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 hybridization and detections can be provided to clinical laboratories as kits for detection of these genetic disorders.

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[1]	1p36.3	AL031282	8823/17757	9137/9167	13960/13931	444/443	4823
CDC2L1[1]	1p36.3	AL031282	8823/17757	13028/13057	17752/17720	445/446	4724
NECDIN	15q11-q13	AC006596	94498/99152	94501/94535	98667/98601	439/440	4166
NECDIN	15q11-q13	AC006596	68031/75948	72122/72156	75666/75637	437/438	3544
NECDIN	15q11-q13	AC006596	76249/79221	76608/76639	78898/78867	441/442	2290

[1]Two sets of primers were used to generate two DNA probe fragments which, together, spanned the entire interval.

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 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 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.

The NECDIN probes were also used in a series of FISH experiments, as shown in FIGS. **3–5** and **11**. In FIG. **3**, the 3544 bp probe was used on metaphase cells from a normal individual, with pre-annealing using purified repetitive

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 corresponding to numerous other known genetic abnormalities (including neoplasias) and syndromes besides those examples given above. Hence, the present invention can be useful for producing probes from genomic regions for which probes are not presently commercially available.

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, Canorhabditis elegans, Drosophila melanogaster*, and *Danio rerio.*

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

SEQUENCE LISTING

The patent contains a lengthy "Sequence Listing" section. A copy of the "Sequence Listing" is available in electronic form from the USPTO web site (http://seqdata.uspto.gov/sequence.html?DocID=6828097B1). An electronic copy of the "Sequence Listing" will also be available from the USPTO upon request and payment of the fee set forth in 37 CFR 1.19(b)(3).

We claim:

1. A method of developing a hybridization probe for a target nucleic acid sequence, said method comprising the steps of:

   determining the sequence of at least one single copy sequence in said target nucleic acid sequence computationally, said determining step comprising the steps of ascertaining the nucleotide-by-nucleotide sequence of said target nucleic acid sequence, comparing said ascertained sequences with the sequences of SEQ ID Nos. 1–428 and 447–479 in said target nucleic acid sequence using a computer program, and identifying said single copy sequence from said comparison; and

   developing a hybridization probe comprising a sequence complementary to a non-repetitive portion of the target which hybridizes to at least a part of said identified single copy sequence.

2. The method of claim 1, 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.

3. The method of claim 2, said purifying step comprising carrying out PCR.

4. The method of claim 1, including the step of labeling said hybridization probe.

5. The method of claim 1, target nucleic acid sequence and said probe being DNA.

6. The method of claim 1, said hybridization probe having at least about 80% sequence identity with said single copy sequence.

7. A method of identifying a single copy sequence interval from a target nucleic acid sequence, said method comprising the steps of ascertaining the nucleotide-by-nucleotide sequence of said target nucleic acid sequence, comparing said ascertained sequences with the sequences of SEQ ID Nos. 1–428 and 447–479 in said target nucleic acid sequence using a computer program, and identifying said single copy sequence from said comparison.

8. The method of claim 7, wherein said nucleotide-by-nucleotide sequence of said target sequence is obtained from existing database information.

*   *   *   *   *