



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12Q 1/68, G01N 33/53, C12P 19/34	A1	(11) International Publication Number: WO 98/49345 (43) International Publication Date: 5 November 1998 (05.11.98)
(21) International Application Number: PCT/US98/08616 (22) International Filing Date: 29 April 1998 (29.04.98) (30) Priority Data: 60/045,078 29 April 1997 (29.04.97) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 60/045,078 (CIP) Filed on 29 April 1997 (29.04.97) (71) Applicant (for all designated States except US): TRUSTEES OF BOSTON UNIVERSITY [US/US]; 147 Bay State Road, Boston, MA 02215 (US). (72) Inventor; and (75) Inventor/Applicant (for US only): SMITH, Cassandra, L. [US/US]; 11 Bay State Road #6, Boston, MA 02115 (US). (74) Agents: CARROLL, Peter, G. et al.; Medlen & Carroll, LLP, Suite 2200, 220 Montgomery Street, San Francisco, CA 94104 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, 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), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(54) Title: METHODS AND COMPOSITIONS FOR TARGETED DNA DIFFERENTIAL DISPLAY		
(57) Abstract Methods and compositions are described for distinguishing between the expression of genes in two or more biological samples. In one embodiment, a PCR-based genomic differential display method is described. The method uses PCR suppression to reduce genome complexity and to amplify genomic fragments containing a targeted interspersed repeat sequence. The PCR products can be displayed by size to create a genomic fingerprint or cloned to create a library of fragments containing the targeted repeat sequence and their flanking sequences. The primers and method of the present invention allows for the accurate identification of differentially expressed gene(s) in various cell types.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon	KR	Republic of Korea	PL	Poland		
CN	China	KZ	Kazakstan	PT	Portugal		
CU	Cuba	LC	Saint Lucia	RO	Romania		
CZ	Czech Republic	LI	Liechtenstein	RU	Russian Federation		
DE	Germany	LK	Sri Lanka	SD	Sudan		
DK	Denmark	LR	Liberia	SE	Sweden		
EE	Estonia			SG	Singapore		

METHODS AND COMPOSITIONS FOR TARGETED DNA DIFFERENTIAL DISPLAY

FIELD OF THE INVENTION

5 The present invention relates to the identification of expressed genes, and in particular, methods and compositions for distinguishing between the expression of genes in two or more biological samples.

BACKGROUND

10 Identifying differences between biological samples is not trivial. The first approach involved the production of a so-called "subtracted cDNA library." A subtracted cDNA library contains cDNA clones corresponding to mRNAs present in one sample and not present in another (*e.g.*, present in a particular species, tissue or cell and not present in another species, tissue or cell). *See generally, Current Protocols in Molecular Biology*, Section 5.8.9 (1990).
15 In the protocol, cDNA containing the gene(s) of interest ["+cDNA"] is prepared with *EcoRI* ends and the cDNA not containing the gene(s) of interest ["-cDNA"] is prepared with blunt ends. The +cDNA is mixed with a 50-fold excess of -cDNA inserts and the mixture is heated to make the DNA single-stranded. Thereafter, the mixture is cooled to allow for hybridization. Annealed cDNA inserts are ligated to a vector and transfected. In theory, the
20 only +cDNA likely to be double-stranded with an *EcoRI* site at each end are those not hybridized to something in the -cDNA preparation; in other words, where a complementary sequence is in the -cDNA preparation, the sequence will not be transfected. Thus, only sequences unique to the +cDNA preparation will be cloned and amplified.

25 The subtraction approach is tedious. Moreover the hybridizations and library production with a small amount of cDNA are technically artful.

 A second approach to identifying differences involves the differential display of mRNAs using arbitrarily primed polymerase chain reaction (DDRT-PCR). The polymerase chain reaction is described by Mullis *et al.* in U.S. Patents Nos. 4,683,195, 4,683,202 and 4,965,188, hereby incorporated by reference. Briefly, the PCR process consists
30 of introducing a molar excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence. The two primers are complementary to their respective strands of the double-stranded sequence. The mixture is denatured and then allowed to hybridize. Following hybridization, the primers are extended with a thermostable DNA polymerase so as

to form complementary strands. The steps of denaturation, hybridization, and polymerase extension can be repeated as often as needed to obtain a relatively high concentration of a segment of the desired target sequence.

In the case of DDRT-PCR, the target is mRNA; the mRNA is, however, treated with reverse transcriptase in the presence of oligo(dT) primers to make cDNA prior to the PCR process. The PCR is carried out with random primers in combination with the oligo(dT) primer used for cDNA synthesis. In theory, since only mRNA is (indirectly) amplified, only the expressed genes are amplified. Where two samples are to be compared, the amplified products are placed in side-by-side lanes of a gel; following electrophoresis, the products can be compared or "differentially displayed."

DDRT-PCR, while an improvement over subtractive hybridization, has a number of drawbacks. First, the use of arbitrary random primers can cause faint banding at essentially every position of the gel. Secondly, the process is generally biased toward high-copy number genes.

There have been some attempts to remedy these problems. For example, E. Haag *et al*, *Biotechniques* 17:226-228 (1994) describes an improved DDRT-PCR method, whereby the use of the standard oligo-dT primer in the PCR step is omitted to decrease the faint banding at essentially every position of the electrophoresis gel. Instead, a second arbitrary primer was utilized in PCR. Another example is O.C. Ikononov *et al*, *Biotechniques* 20:1030-1042 (1996); this paper describes the use of a modified DDRT-PCR protocol to increase bias towards moderate to low abundance transcripts. The authors utilized experimentally selected primer pairs directed at known coding sequences that avoid amplification of highly abundant ribosomal and mitochondrial transcripts. While such efforts have improved DDRT-PCR, the process remains unsatisfactory because of the continued amplification of material that is not of interest.

What is needed is a convenient method for distinguishing between the expression of genes in two or more biological samples.

SUMMARY OF THE INVENTION

The present invention relates to the identification of expressed genes, and in particular, methods and compositions for distinguishing between the expression of genes in two or more biological samples. In some embodiments, the present invention employs oligonucleotide

primers targeting CAG repeats. Such repeats are known to be important in an increasing number of neurological diseases.

In one embodiment, the present invention contemplates first and second oligonucleotide primers, said first oligonucleotide primer comprising a CTG repeat and said second oligonucleotide primer comprising a CAG-repeat. The primer containing a CTG repeat is contemplated for amplifying unique sequences flanking the 5'-side of the CAG repeat sequence in the target. The primer containing the CAG repeat is contemplated for amplifying the 3'-flanking sequences.

It is not intended that the present invention be limited by the number of CTG or CAG repeats in the primers. A variety of primers are contemplated. In one embodiment, an oligonucleotide comprising $(CAG)_n$ or $(CTG)_n$ is contemplated, wherein n is a whole number between 1 and 15, and more preferably, between 2 and 12, and still more preferably, between 6 and 12.

It is also not intended that the present invention be limited by the number of additional non-repeating bases in the primers. A variety of primers are contemplated. In one embodiment, oligonucleotides comprising the general formula $X_p(CAG)_nX_p$ or $X_p(CTG)_nX_p$ are contemplated, wherein X is selected from the group consisting of A, T, C or G and p is a whole number between 1 and 15, and n is a whole number between 2 and 12, and still more preferably, between 6 and 12.

Furthermore, it is not intended that the present invention be limited to the use of CAG and CTG repeats, as other repeating and non-repeating nucleotide sequences will find use as primers with the present invention.

It is not intended that the present invention be limited by the nature of the sample. The terms "sample" and "specimen" in the present specification and claims are used in their broadest sense. On the one hand they are meant to include a specimen or culture. On the other hand, they are meant to include both biological and environmental samples. These terms encompasses all types of samples obtained from humans and other animals, including but not limited to, body fluids such as urine, blood, fecal matter, cerebrospinal fluid (CSF), semen, and saliva, cells as well as solid tissue (including both normal and diseased tissue). These terms also refers to swabs and other sampling devices which are commonly used to obtain samples for culture of microorganisms.

It is also not intended that the invention be limited by the particular purpose for carrying out the biological reactions. In one medical diagnostic application, it may be desirable to differentiate between normal and diseased tissue.

In one embodiment, the present invention contemplates a method of analyzing nucleic acid in a sample, comprising: a) providing: i) a sample containing nucleic acid, ii) a first oligonucleotide primer comprising a CTG repeat, iii) a second oligonucleotide primer comprising a CAG-repeat, and iv) a polymerase and PCR reagents; b) preparing said nucleic acid from said sample under conditions so as to produce amplifiable nucleic acid; c) amplifying said nucleic acid with said first and second primers, said polymerase and said PCR reagents under conditions such that amplified product is generated; d) detecting said amplified product. In one embodiment, adapter oligonucleotides are employed. In a preferred embodiment, PCR-suppression is employed in conjunction with the primers described above.

It is not intended that the present invention be limited by the means of detection. In one embodiment, said detecting comprises gel electrophoresis. Furthermore, it is not intended that the present invention be limited to the use of PCR amplification as a means of characterizing the samples. In some embodiments, the present invention contemplates targeting methods to detect differences in nucleic acid samples, whereby the sample are identified through the use of DNA arrays (*See e.g.*, the methods of Chee *et al.*, Science 274, 610 [1996]; DeRisi *et al.*, Nat. Genet. 14, 457 [1996]; Gress *et al.*, Oncogene 13, 1819 [1996]; Maskos and Southern, Nucleic Acids Res. 21, 4663 [1993]; Pietu *et al.*, Genome Res. 6, 492 [1996]; Schena *et al.*, Science 270, 467 [1995]; and Schena *et al.*, Proc. Natl. Acad. Sci. 93, 10614 [1996]).

The present invention can be used with particular success when comparing samples. In one embodiment, the present invention contemplates a method of analyzing expressed genes in biological samples. Clinical samples are specifically contemplated within the scope of the present invention.

The present invention contemplates the primers of the present invention as unique compositions. The present invention also contemplates kits containing these novel compositions. In one embodiment, the kit comprises a first primer comprising a CTG repeat and said second oligonucleotide primer comprising a CAG-repeat.

DEFINITIONS

To facilitate understanding of the invention, a number of terms are defined below.

"Nucleic acid sequence" and "nucleotide sequence" as used herein refer to an oligonucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand.

5 As used herein, the terms "complementary" or "complementarity" are used in reference to "polynucleotides" and "oligonucleotides" (which are interchangeable terms that refer to a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "C-A-G-T," is complementary to the sequence "G-T-C-A." Complementarity can be "partial" or "total." "Partial" complementarity is where one or more nucleic acid bases is not matched
10 according to the base pairing rules. "Total" or "complete" complementarity between nucleic acids is where each and every nucleic acid base is matched with another base under the base pairing rules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which
15 depend upon binding between nucleic acids.

The terms "homology" and "homologous" as used herein in reference to nucleotide sequences refer to a degree of complementarity with other nucleotide sequences. There may be partial homology or complete homology (*i.e.*, identity). A nucleotide sequence which is partially complementary, *i.e.*, "substantially homologous," to a nucleic acid sequence is one
20 that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid sequence. The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the
25 binding (*i.e.*, the hybridization) of a completely homologous sequence to a target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a
30 partial degree of complementarity (*e.g.*, less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

Low stringency conditions comprise conditions equivalent to binding or hybridization at 42°C in a solution consisting of 5X SSPE (43.8 g/l NaCl, 6.9 g/l NaH₂PO₄•H₂O and 1.85

g/l EDTA, pH adjusted to 7.4 with NaOH), 0.1% SDS, 5X Denhardt's reagent [50X Denhardt's contains per 500 ml: 5 g Ficoll (Type 400, Pharmacia), 5 g BSA (Fraction V; Sigma)] and 100 µg/ml denatured salmon sperm DNA followed by washing in a solution comprising 5X SSPE, 0.1% SDS at 42°C when a probe of about 500 nucleotides in length is employed.

The art knows well that numerous equivalent conditions may be employed to comprise low stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, *etc.*) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol), as well as components of the hybridization solution may be varied to generate conditions of low stringency hybridization different from, but equivalent to, the above listed conditions. In addition, the art knows conditions which promote hybridization under conditions of high stringency (*e.g.*, increasing the temperature of the hybridization and/or wash steps, the use of formamide in the hybridization solution, *etc.*).

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe which can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe which can hybridize (*i.e.*, it is the complement of) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids using any process by which a strand of nucleic acid joins with a complementary strand through base pairing to form a hybridization complex. Hybridization and the strength of hybridization (*i.e.*, the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementarity between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids.

As used herein the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen

bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (*e.g.*, C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized to a solid support [5 *e.g.*, a nylon membrane or a nitrocellulose filter as employed in Southern and Northern blotting, dot blotting or a glass slide as employed in *in situ* hybridization, including FISH (fluorescent *in situ* hybridization)].

As used herein, the term " T_m " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl [10 *see e.g.*, Anderson and Young, Quantitative Filter Hybridization, in *Nucleic Acid Hybridization* (1985)]. Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m . 15

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. "Stringency" typically occurs in a range from about $T_m - 5^\circ\text{C}$ (5°C below the T_m of the probe) to about 20°C to 25°C below T_m . 20 As will be understood by those of skill in the art, a stringent hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences.

As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids which may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template." 25

As used herein, the term "sample template" refers to nucleic acid originating from a sample which is analyzed for the presence of a target sequence of interest. In contrast, "background template" is used in reference to nucleic acid other than sample template which 30 may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

"Amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction technologies well known in the art [Dieffenbach CW and GS Dveksler (1995) *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview NY]. As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis U.S. Patent Nos. 4,683,195 and 4,683,202, hereby incorporated by reference, which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. The length of the amplified segment of the desired target sequence is determined by the relative positions of two oligonucleotide primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

With PCR, it is possible to amplify a single copy of a specific target sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ³²P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications.

Amplification in PCR requires "PCR reagents" or "PCR materials", which herein are defined as all reagents necessary to carry out amplification except the polymerase, primers and template. PCR reagents normally include nucleic acid precursors (dCTP, dTTP etc.) and buffer.

As used herein, the term "primer" refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, (*i.e.*, in the presence of nucleotides and an inducing agent such as DNA polymerase and at a suitable temperature and pH). The primer is preferably single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is first treated to separate its strands before being used to prepare extension products.

Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source of primer and the use of the method.

5 As used herein, the term "probe" refers to an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, which is capable of hybridizing to another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification and isolation of particular gene sequences. It is contemplated that any probe used in the present invention will be labelled with any
10 "reporter molecule," so that it is detectable using any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is not intended that the present invention be limited to any particular detection system or label.

15 As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial enzymes, each of which cut double-stranded DNA at or near a specific nucleotide sequence.

DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one
20 mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring. An end of an oligonucleotide is referred to as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of another mononucleotide pentose ring. As used herein, a nucleic acid sequence,
25 even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements which direct transcription of a linked gene are generally located 5' or upstream of the coding
30 region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

As used herein, the terms "nucleic acid molecule encoding," "DNA sequence encoding," and "DNA encoding" refer to the order or sequence of deoxyribonucleotides along a strand of deoxyribonucleic acid. The order of these deoxyribonucleotides determines the order of amino acids along the polypeptide (protein) chain. The DNA sequence thus codes for the amino acid sequence.

The term "isolated" when used in relation to a nucleic acid, as in "an isolated oligonucleotide" refers to a nucleic acid sequence that is identified and separated from at least one contaminant nucleic acid with which it is ordinarily associated in its natural source. Isolated nucleic acid is nucleic acid present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA which are found in the state they exist in nature.

As used herein, the term "purified" or "to purify" refers to the removal of undesired components from a sample.

As used herein, the term "substantially purified" refers to molecules, either nucleic or amino acid sequences, that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated. An "isolated polynucleotide" is therefore a substantially purified polynucleotide.

As used herein, the term "gene" means the deoxyribonucleotide sequences comprising the coding region of a structural gene and including sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb on either end such that the gene corresponds to the length of the full-length mRNA. The sequences which are located 5' of the coding region and which are present on the mRNA are referred to as 5' non-translated sequences. The sequences which are located 3' or downstream of the coding region and which are present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns" or "intervening regions" or "intervening sequences." Introns are segments of a gene which are transcribed into heterogenous nuclear RNA (hnRNA); introns may contain regulatory elements such as enhancers. Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a nascent polypeptide.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences which are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (these flanking sequences are located 5' or 3' to the non-translated sequences present on the mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers which control or influence the transcription of the gene. The 3' flanking region may contain sequences which direct the termination of transcription, posttranscriptional cleavage and polyadenylation.

The term "sample" as used herein is used in its broadest sense and includes environmental and biological samples. Environmental samples include material from the environment such as soil and water. Biological samples may be animal, including, human, fluid (e.g., blood, plasma and serum), solid (e.g., stool), tissue, liquid foods (e.g., milk), and solid foods (e.g., vegetables).

DESCRIPTION OF THE DRAWINGS

Figure 1 is a display of fluorescein-labeled *Mse* I CAG-containing DNA fragments from three pairs of monozygotic twins. Genomic DNAs were digested with the restriction enzyme *Mse* I, ligated to adapter oligonucleotides of known sequence and hybridized to an immobilized single-stranded probe containing (CAG)₁₂ repeats. Captured DNAs were amplified by PCR with fluorescein-labeled primer 3 (Table 1) and displayed on an ALF sequencing instrument. DNAs were from three pairs of monozygotic twins (pair 1, lanes 15 and 16; pair 2, lanes 17 and 18; pair 3, lanes 19 and 20). The elution time (min) is shown on the x-axis, and the fluorescence intensity is shown on the y-axis (arbitrary units). The size range of the displayed fragments is ~230 to ~350 bp.

Figure 2 shows HD-sequence information used in experiments. A partial sequence of the first exon of the HD gene (Genbank accession no. L34020) is shown along with the locations of HD primers (oligonucleotides 9-12; Table 1; horizontal arrows) and *Sau*96 I restriction sites (vertical arrows) surrounding a CAG repeat (bold).

Figure 3 is a display of different HD alleles from three related individuals (lanes 1, HD-A; 2, HD-B; and 3, HD-C) and an anonymous unrelated control (lane 4) in a *Sau*96 I CAG-containing genome subset. Since there were many differences between samples, a number of control experiments was done to determine the size of the HD alleles, to identify HD-containing fragments, and to understand the primary structure of the PCR products using

different primers. A subset of these control experiments is shown in (A) and (B), whereas the true HD genomic display is shown in (C). (A) HD-specific amplification of genomic DNA using primers 9 (fluorescein-labeled) and 12 (Table 1) measured the size of the HD-alleles. (B) The HD-fragments, generated from genomic DNA using primers 11 and 12 (Table 1 and Figure 2A), were digested with *Sau96* I, tagged with *Sau96* I-adapters and amplified by PCR using primer 5 and fluorescein-labeled CTG-repeat primer 13 (Table 1). This experiment measured the size of the HD-containing fragments when the PCR used adapter and repeat primers. (C) Genomic DNAs were digested with the restriction enzyme *Sau96* I, ligated to adapter oligonucleotides of known sequence and hybridized to a (CAG)₁₂-containing oligonucleotide probe. The captured fragments were amplified by PCR using CAG repeat-containing primer 14 and fluorescein-labeled primer 5 (Table 1). This experiment displays *Sau96* I CAG-containing genome subsets *en masse*. The differentially displayed fragments eluting at ~227 min hybridized to an HD-specific capture probe (data not shown). Also not shown are the results of experiments using adapter primers only. These experiments detected the same differentially displayed fragments eluting at ~227 min as shown in (C).one embodiment of the primers of the present invention (a "K primer") partially hybridized to one strand of a denatured double-stranded template.

Figure 4 is a schematic showing the PCR-based genomic differential display method used to amplify interspersed repeats and flanking unique sequences.

Figure 5 is a genomic differential display targeting different (CAG)_n-containing genome subsets. DNA sample was amplified first with a fluorescein-labeled T-primer and an A-primer (oligonucleotides 11 and 3, respectively, Table 2). Then, a second PCR amplification was done using the same fluorescein-labeled T-primer (oligonucleotide 11, Table 2) and various A-primers; oligonucleotide 5 (lane 1), 9 (lane 2), 8 (lane 3), or 6 (lane 4) (Table 2). No PCR products were generated if fluorescein-labeled T-primer (oligonucleotide 11, Table 2) was used alone (lane 5). Fluorescence intensity versus fragment size is shown. Each lane is independently auto-scaled.

Figure 6 shows a comparison of CAG repeat-containing genomic subsets from monozygotic twins amplified by PCR using 3'-(A) or 5'-terminally anchored (B) T-repeat primers. DNA from two pairs of monozygotic twins was digested with *Hae* III, ligated to adapter (oligonucleotides 1 and 2, Table 2) and PCR amplified. The products were fractionated by size on an ALF sequencing instrument. (A) Both PCR amplifications used a 3'-terminally anchored fluorescein-labeled T-primer (oligonucleotide 12, Table 2). The A-

primers were oligonucleotides 3 and 8 (Table 2) in the first and second PCR reactions, respectively. (B) Both PCR amplifications used a 5'-anchored fluorescein-labeled T-primer (oligonucleotide 11, Table 2). The A-primers were oligonucleotides 3 and 8 in the first and second PCR, respectively (Table 2).

5

DESCRIPTION OF THE INVENTION

The present invention relates to the identification of expressed genes, and in particular, methods and compositions for distinguishing between the expression of genes in two or more biological samples. Specifically, the present invention contemplates methods and
10 compositions for differential display with genomic DNA, cDNA, and other nucleic acid species. For genomic DNA, genome complexity is reduced by the methods of the present invention, and focus is provided by targeting genome subsets containing specific interspersed repeats. In one embodiment, the targeted genomic subsets contain a CAG trinucleotide repeating sequence. Using primers described to such repeats, surrounding unique sequences are
15 identified.

Other interspersed repeat-containing fragments can be targeted in the same manner, *e.g.*, fragments containing SINEs, LINEs, LTRs (long terminal repeats), sequences coding for particular protein motifs or *cis*-acting sequence elements. Differential display of cDNAs could also be enhanced by a similar use of interspersed repeats to target interesting cDNA
20 subsets.

The display method of the present invention will detect polymorphisms that may arise in the unique sequences surrounding the repeat. However, each displayed polymorphism must be characterized individually in order to understand its origin.

The informativeness of these and other differential display methods will be maximized
25 when large amounts of data can be analyzed automatically, such as by high throughput automated analysis using signal processing methods. The display method can be readily applied to assess the variation within monozygotic twin pairs. At a minimum, such experiments should provide quantitative information on genome stability, and they have the potential to reveal interesting facets of twin biology.

30 The methods and compositions of the present invention provide general means of targeting a variety of DNA sources to allow comparison of samples, to target samples to reduce complexity, to focus analysis on DNA fragments containing specific sequences, and to target rare nucleic acid species. One skilled in the art will recognize many applications of the

methods and compositions of the present invention. For example, the method can be generalized to amplify genome subsets containing a variety of targets including various consensus sequences coding for cis-acting elements or protein motifs. Several examples of such applications are provided below to illustrate uses of the present invention. However, the present invention is not limited to these examples.

In some embodiments, the methods of the present invention use a sequence specific capture step and PCR amplification to focus analysis on genome subsets containing target sequences. Adaptor-tagged genomic restriction fragments (i.e., restriction fragments ligated to known oligonucleotides) containing a target sequence are captured by hybridization to an immobilized complementary single stranded probe. The captured fragments are amplified by PCR using primers to the adaptor sequences alone, or in the presence of a primer complementary to the targeted sequence. The amplified and labeled fragments can then be characterized on a sequencing gel or other means known in the art.

In one embodiment, the method allows the isolation and analysis of genome subsets containing targeted repeat sequences. In a preferred embodiment (see Description of Preferred Embodiments), the method takes advantage of PCR-suppression ["PS;" Siebert *et al.*, *Nucleic Acids Res.* 23:1087-1088 (1995); Lukyanov *et al.*, *Anal. Biochem.* 229:198-202 (1995)] to amplify targeted repeat-containing sequences along with a unique flanking sequence directly from genomic DNA. Previously, PS was used to selectively amplify genomic DNA sequences adjacent to known sequences and to improve subtractive hybridization protocols [Lukyanov *et al.*, *supra*; Diachenko *et al.*, *Proc. Natl. Acad. Sci. USA* 93:6025-6030 (1996)]. Here, the focus is on displaying genomic fragments containing trinucleotide repeating sequences because of their importance in an increasing number of human diseases [for review see Sutherland and Richards, *Proc. Natl. Acad. Sci. USA* 92:3636-3641 (1995); Huddart *et al.*, *Br. J. Cancer* 72:642-645 (1995); Wooster *et al.*, *Nature Genet.* 6:152-156 (1994)]. A similar PCR approach was recently reported for microsatellites isolation from less complex cloned DNA samples (YAC's, P1, cosmids, bacteriophages or plasmid clones) using primers complementary to a dinucleotide repeat and a vector sequence [Lench *et al.*, *Nucleic Acids Res.* 24:2190-2191 (1996)].

Principles Of PCR Suppression

The outline of a PCR-based genomic differential display method is shown in Figure 4. The template is genomic DNA digested with a restriction enzyme and ligated to known

adapter sequences. A 40-base adapter sequence is used to promote the annealing of the fragment ends to each other. The annealed complementary ends suppress annealing or extension of shorter PCR primers (A-primers) complementary to the same sequences. This effect has been termed PS for PCR suppression [Siebert *et al.* (1995); Lukyanov *et al.* (1995);
5 Diachenko *et al.* (1996)].

Under PS conditions, efficient PCR amplification is achieved only when two primers are used. The A-primer corresponds to the self-complementary end adapter sequences. The second primer, termed the T-primer, is complementary to a genomic target sequence located in the single-stranded section of the end-annealed genomic fragments. Single-stranded PCR
10 products, produced by extension of annealed T-primers, no longer have complementary ends and will not be subjected to PS. Genomic fragments that do not contain the target sequence will remain end-annealed and cannot be extended by the A-primer. Occasional extension of an annealed A-primer to the original template will produce a single-stranded fragment that is still subject to PS because of its complementary ends. Thus, PS ensures that only fragments
15 containing the targeted sequence are efficiently amplified by PCR. The T-primers used here target CAG-repeat sequences. A T-primer containing a CTG-repeat amplifies unique sequence flanking the 5'-side of the CAG repeat sequence, whereas the 3'-flanking sequence is amplified with the T-primer containing a CAG-repeat sequence.

20 Preparation of RNA

The nucleic acid content of cells consists of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA). The DNA contains the genetic blueprint of the cell. RNA is involved as an intermediary in the production of proteins based on the DNA sequence. RNA exists in three forms within cells, structural RNA (*i.e.*, ribosomal RNA "rRNA"), transfer
25 RNA ("tRNA"), which is involved in translation, and messenger RNA ("mRNA"). Since the mRNA is the intermediate molecule between the genetic information encoded in the DNA, and the corresponding proteins, the cell's mRNA component at any given time is representative of the physiological state of the cell. In order to study and utilize the molecular biology of the cell, it is therefore important to be able to purify mRNA, including
30 purifying mRNA from the total nucleic acid of a sample.

The preparation of RNA is complicated by the presence of ribonucleases that degrade RNA (*e.g.*, T. Maniatis *et al.*, Molecular Cloning, pp. 188-190, Cold Spring Harbor Laboratory [1982]). Furthermore, the preparation of amplifiable RNA is made difficult by

the presence of ribonucleoproteins in association with RNA. (See, R. J. Slater, *In: Techniques in Molecular Biology*, J.M. Walker and W. Gaastra, eds., Macmillan, NY, pp. 113-120 [1983]).

Typically, the steps involved in purification of nucleic acid from cells include 1) cell
5 lysis; 2) inactivation of cellular nucleases; and 3) separation of the desired nucleic acid from the cellular debris and other nucleic acid. Cell lysis may be achieved through various methods, including enzymatic, detergent or chaotropic agent treatment. Inactivation of cellular nucleases may be achieved by the use of proteases and/or the use of strong denaturing agents. Finally, separation of the desired nucleic acid is typically achieved by extraction of
10 the nucleic acid with phenol or phenol-chloroform; this method partitions the sample into an aqueous phase (which contains the nucleic acids) and an organic phase (which contains other cellular components, including proteins). Commonly used protocols require the use of salts in conjunction with phenol (P. Chomczynski and N. Sacchi, *Anal. Biochem.* 162:156 [1987]), or employ a centrifugation step to remove the protein (R.J. Slater, *supra*). While useful, phenol
15 extraction is time consuming and creates a serious waste disposal problem.

Once the nucleic acid fraction been isolated from the cell, the structure of the mRNA molecule may used to assist in the purification of mRNA from DNA and other RNA molecules. Because the mRNA of higher organisms is usually polyadenylated on its 3' end ("poly-A tail" or "poly-A track"), one means of isolating RNA from cells has been based on
20 binding the poly-A tail with its complementary sequence (*i.e.*, oligo-dT), that has been linked to a support such as cellulose. Commonly, the hybridized mRNA/oligo-dT is separated from the other components present in the sample through centrifugation or, in the case of magnetic formats, exposure to a magnetic field. Once the hybridized mRNA/oligo-dT is separated from the other sample components, the mRNA is usually removed from the oligo-dT.
25 However, for some applications, the mRNA may remain bound to the oligo-dT that is linked to a solid support.

A wide variety of solid supports with linked oligo-dT have been developed and are commercially available. Cellulose remains the most common support for most oligo-dT systems, although formats with oligo-dT covalently linked to latex beads and paramagnetic
30 particles have also been developed and are commercially available. The paramagnetic particles may be used in a biotin-avidin system, in which biotinylated oligo-dT is annealed in solution to mRNA. The hybrids are then captured with streptavidin-coated paramagnetic particles, and separated using a magnetic field. In addition to these methods, variations exist,

such as affinity purification of polyadenylated RNA from eukaryotic total RNA in a spun-column format. These approaches allow for hybridization of poly A mRNA, but vary in efficiency and sensitivity.

It is not intended that the present invention be limited by the source of RNA; a variety of sources is contemplated, including but not limited to mammalian (*e.g.*, liver tissue), plant (*e.g.*, tobacco leaves) and microbial (*e.g.*, yeast). In one embodiment, the present invention contemplates the isolation of PolyA+ RNA from extracts, including direct isolation from crude extracts.

EXPERIMENTAL

The following examples serve to illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); gm (grams); mg (milligrams); μ g (micrograms); L (liters); ml (milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); $^{\circ}$ C (degrees Centigrade); Ci (Curies); MW (molecular weight); OD (optical density); EDTA (ethylenediamine-tetracetic acid); PAGE (polyacrylamide gel electrophoresis); UV (ultraviolet); V (volts); W (watts); mA (milliamps); bp (base pair); CPM (counts per minute).

Nonphosphorylated oligonucleotides (Table 1) were from Operon Technologies (Alameda, CA). DNA samples were from an anonymous healthy donor, a Huntington's disease (HD)-affected kindred and monozygotic twins.

Capture Of Targeted Genome Subsets

Genomic DNA (100 ng), digested with the restriction enzyme *Sau*3A I or *Mse* I, was ligated with 50 pmoles of corresponding adapters (oligonucleotides 1 and 2, or and 3 and 4 for *Mse* I, respectively; Table 1) in a 10-20 μ l total reaction volume overnight at 14 $^{\circ}$ C with 40 units of T4 DNA ligase (New England Biolabs). Each pair of oligonucleotides was first annealed by cooling the mixture from 70 $^{\circ}$ C to 10 $^{\circ}$ C in a 1-hr period. Ligase was inactivated by heating at 75 $^{\circ}$ C for 10 min, and a fill-in reaction was done at 72 $^{\circ}$ C for 10 min after the addition of dNTPs (100 μ M each) and 0.5 units of AmpliTaq DNA polymerase (Perkin

Elmer). DNA was phenol extracted, precipitated with ethanol, washed with 70% ethanol, dried and dissolved in TE buffer (10 mM Tris-HCl, pH 8.0/1 mM EDTA). A biotinylated oligonucleotide (10 pmol) containing a (CAG)₁₂ or (CCG)₁₂ sequence (oligonucleotide 7 or 8, respectively; Table 1) was mixed with 50 ng of ligation products in 50 µl of TE buffer containing 2 µM of the corresponding adapter oligonucleotides to prevent annealing of the fragment ends to each other. After the addition of mineral oil, the sample was heated to 95°C, slowly cooled to room temperature, added to 100 µg of prewashed streptavidin-coated magnetic beads M-280 [as directed by Dynal (Oslo, Norway)] /using a 3-fold molar excess of biotin-binding capacity over biotinylated oligonucleotides/ and incubated at room temperature for 1 hr with gentle rotation. The beads were collected with a magnet, washed twice at 55-60°C for 20 min with 3x SSC (1x SSC: 0.15 M NaCl/ 15 mM sodium citrate)/ 0.5% SDS, and, at room temperature, twice each, with TE containing 1 M NaCl and with TE alone. Beads with captured DNA were stored in TE buffer at 4°C.

TABLE 1

Oligonucleotide Sequences

No.	Description	Sequence (5' → 3') ^a
1	<i>Sau</i> 3A I-adapter-24	CGGGAATTCTGGCTCTGCGACATG
2	<i>Sau</i> 3A I-adapter-10	GATCCATGTC
3	<i>Mse</i> I-adapter-20	TCTCCAGCCTCTCACCGCAT
4	<i>Mse</i> I-adapter-11	TAATGCGGTGA
5	<i>Sau</i> 96 I-adapter-22	AGCACTCTCCAGCCTCACCGCG
6	<i>Sau</i> 96 I-adapter-11	GNCCGCGGTGA
7	CAG-12	b-GATGATCCGACGCAT(CAG) ₁₂
8	CCG-12	b-GATGATCCGACGCAT(CCG) ₁₂
9	HD2F	ATGAAGGCCTTCGAGTCCCTCAAGTCCTT
10	HD1R	GGAAGGACTTGAGGGACTCGAAGGCCTTCAT
11	HD1F	ACGGCCGCTCAGGTTCTGCTTTTAC
12	HD2R	GGCGGCTGAGGAAGCTGAGGA
13	CTG-4	A(A/T/C) (CTG) ₄
14	CAG-4	T(A/T/C) (CAG) ₄

DNA Amplification And Labeling By PCR

One-fifth of the captured DNA was amplified by PCR in a PTC-100 thermal cycler (MJ Research, Inc.) as described in Lisitzyn *et al.*, *Science* 259:946-951(1993). The 50-µl

reaction contained 67 mM Tris-HCl, pH 8.8/ 4 mM MgCl₂/ 16 mM (NH₄)₂SO₄/ 10 mM 2-mercaptoethanol/ 300 µM of each dNTP/ 2 units of AmpliTaq DNA polymerase/ 5 µM fluorescein-labeled adapter primer (i. e., in the absence of a repeat primer). The samples were incubated at 94°C for 3 min and subjected to 20-23 cycles, each consisting of 1 min at 94°C and 3 min at 72°C, and a final incubation at 72°C for 5 min (e.g., Fig. 1).

In other experiments, captured DNA was amplified by PCR using the appropriate adapter primer (2.5 µM) and a primer complementary to the repeat (5 µM, oligonucleotides 13 or 14; Table 1). PCR conditions were as described above, except that the annealing temperature was 45°C. The primer labeled with fluorescein varied in different experiments.

Display Of Targeted Genome Subsets

Amplified PCR products (1-2 µl) were denatured for 5 min at 90°C in 4 µl of a stop solution containing 6 mg/ml of dextran blue and 0.1% SDS in deionized formamide, loaded onto a 6% denaturing polyacrylamide gel and analyzed on an ALF DNA Sequencer (Pharmacia-Biotech., Sweden). The results were displayed using Fragment Manager software provided with the instrument. The size standard was a fluorescein-labeled 100-base pair (bp) ladder (Gibco-BRL). The electrophoresis conditions fractionated fragments from 80 to 800 bp.

Cloning And Analysis Of Captured Sequences

The CAG-containing double-stranded fragments obtained after capture and PCR amplification with adapter primers were cloned using a TA cloning kit (Invitrogen, San Diego, CA). Randomly chosen clones were sequenced using a Sequenase 2.0 kit (Pharmacia-Biotech., Sweden) and an ALF sequencer.

Quantitation Of CAG Repeats In The Human Genome

One µg of human genomic DNA was digested to completion with *Hind* III (New England Biolabs, MA). The 5' ends were labeled with ³²P by a T4 polynucleotide kinase exchange reaction as recommended by New England Biolabs. Radiolabeled DNA (5.8 pmoles) was hybridized to 5 pmoles of a biotinylated (CAG-12) probe (oligonucleotide 7; Table 1) and captured on streptavidin-coated magnetic beads. Hybridization and washing conditions were as described above for purification of genome subsets. The amount of

radioactive DNA retained on the beads was determined by Cerenkov counting in a Beckman scintillation counter.

Detection Of Specific Allelic Differences In CAG-containing Genome Subsets

5 These control experiments were done to demonstrate that known allelic differences in the HD locus could be detected in a CAG-enriched genome subset. The samples compared were from non-identical genomes. Thus, additional experiments were also done to determine the length of the HD-alleles using locus-specific PCR and to identify the HD allele sizes produced by PCR with adapters and repeat primers.

10 A standard HD-specific PCR amplification (see Fig. 2 for relevant HD sequence and primer and restriction site locations) with oligonucleotides 9 and 12, or 11 and 12 (Table 1) was done to determine the HD-allele sizes. Each 50- μ l reaction contained 100 ng DNA/ 0.5 μ M of each primer/ 20 mM Tris-HCl, pH 8.4/ 50 mM KCl/ 200 μ M dNTPs/ 2 mM $MgCl_2$ / 3.5% formamide/ 15% glycerol/2.5 units of AmpliTaq DNA polymerase. Cycling conditions
15 were 94°C for 3 min, followed by 35 cycles at 94°C for 1 min, 64°C for 1 min, and 72°C for 1 min, and a final incubation at 72°C for 7 min.

 In other experiments, gel-purified HD-fragments (~150 ng), obtained from PCR amplification of genomic DNA with primers 11 and 12 (Table 1), were digested with *Sau*96 I and ligated to *Sau*96 I adapters (oligonucleotides 5 and 6; Table 1). The resulting HD-
20 fragments were then amplified by PCR using adapter primer 5 and fluorescein-labeled CTG-containing repeat primer 13 (Table 1). PCR conditions and analysis were as described above.

 CAG repeat-containing genomic subsets, captured from *Sau*96 I-digested and tagged DNAs, were amplified by PCR using fluorescein-labeled primer 5 and a CAG-containing repeat primer (primer 14) or CTG-containing repeat primer 13 (Table 1). Two- μ l aliquots
25 were displayed as described above. The remaining PCR product (~100 ng) was hybridized overnight at 37°C to an immobilized HD-specific capture probe in 6x SSC/ 5x Denhardt's solution/ 0.5% SDS/ 100 μ g/ml of herring sperm DNA/ 100 pmol each of oligonucleotides (CAG)₆ and (CTG)₆. The capture probe, generated by PCR using oligonucleotide 10 and biotinylated oligonucleotide 11 (Table 1), was a 173-bp fragment upstream from the CAG
30 repeat in the first exon of the HD gene. The gel-purified PCR product (175 ng, 1.5 pmol) was immobilized on streptavidin-coated magnetic beads and treated with alkali, as recommended by the manufacturer, Dynal. After hybridization, the beads were washed once in 1x SSC/ 0.5% SDS at 65°C for 2 hrs and rinsed twice with TE at room temperature.

Captured fragments were released from the beads by boiling for 5 min in stop solution and displayed as described above.

RESULTS

5 Principles Of The Method

One goal of this work was the development of a genomic differential display method targeting interspersed repeat-containing sequences, although the present invention finds use with non-genomic DNA (*e.g.*, cDNA), among other nucleic acid species. Major considerations were simplicity and the use of a minimal amount of DNA (200 ng/experiment). The procedure takes advantage of previously developed sequence-specific capture methods [Ito *et al.* (1992) *Nucleic Acids Res.* 20:3624; Broude *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:3072-3076; Kandpal *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:88-92; Brown *et al.* (1995) *Mol. Cell. Probes* 9:53-58; Rothuizen & van Raak (1994) *Nucleic Acids Res.* 22:5512-5513; Parimoo *et al.* (1995) *Anal. Biochem.* 228:1-17]. CAG- (and CGG-) trinucleotide repeating sequences were targeted because of the importance of these repeats in neurodegenerative diseases [Ross *et al.* (1993) *Trends Neurosci.* 16:254-260; Sutherland & Richards (1995) *Proc. Natl. Acad. Sci. USA* 92:3636-3641]. It should be noted that only ten capture probes are needed to profile all trinucleotide repeat sequences.

In brief, genomic DNA is cleaved with a restriction enzyme which cuts outside of the targeted repeat sequence. The restriction fragments are tagged at their ends by ligation to adapters, *i. e.*, oligonucleotides of known sequence. The adapters permit subsequent amplification and labeling of the fragments by PCR. The fragments are denatured, hybridized to a biotinylated single-stranded oligonucleotide probe containing a sequence, (CAG)₁₂ [or (CCG)₁₂], complementary to the targeted repeat sequence, and captured on streptavidin-coated magnetic beads. Captured fragments are then amplified by PCR using an adapter primer alone or in combination with a primer complementary to the targeted repeat. Either the adapter or the repeat primer was fluorescein-labeled to allow detection of the products on a high-resolution automated fluorescent DNA sequencing instrument. Others have also used an automated DNA sequencing instrument for analysis for differential display [Ito *et al.* (1994) *FEBS Lett.* 351:231-236; Kato, K. (1995) *Nucleic Acids Res.* 23:3685-3690].

Efficacy Of The Procedure

Captured fragments containing CAG- (or CCG-) repeating sequences were amplified by PCR using an adapter primer and cloned. The clones were hybridized to the targeted repeat sequence. As expected, most (~90% and ~60% of the putative CAG and CCG clones, respectively) hybridized to the corresponding probes. Sequencing of four randomly-selected CAG-containing clones revealed the presence of four different CAG repeated sequences, *i.e.*, (CAG)₃, (CAG)₄, (CAG)₆CCAGAGCCAG, and (CAG)₂ACAGCA. These results showed that the capture procedure generated genome subsets enriched for targeted repeat-containing sequences.

The completeness of the capture CAG-containing genome subset was evaluated by determining the total number of ³²P end-labeled *Hind* III restriction fragments captured by hybridization to an immobilized oligonucleotide containing (CAG)₁₂ (oligonucleotide 7; Table 1). A total of ~0.5% of the fragments (or ~5 x 10³ fragments) were captured. A similar number (~2.5 x 10³) of CAG repeats were predicted when human DNA sequences in Genbank were analyzed [Stallings (1994) Genomics 21:116-121]. This suggests that the capture procedure enriches for a particular genome subset with little loss of targeted sequences.

Reproducibility And Sensitivity Of The Method

Display of CAG repeat-containing fragments from three pairs of monozygotic twins is shown in Fig. 1. In this experiment genomic DNAs were digested with *Mse* I, ligated to an adapter consisting of oligonucleotides 3 and 4 (Table 1) and hybridized to an immobilized single-stranded oligonucleotide probe containing (CAG)₁₂ (oligonucleotide 7; Table 1). The captured DNA was amplified by PCR using a fluorescein-labeled adapter primer (oligonucleotide 3; Table 1). The fluorescence intensity versus elution time profiles, shown in Fig. 1, represents size-fractionated fluorescein-labeled fragments.

Although it was shown that the capture step is very efficient (see above), it is difficult to estimate the total number of displayed fragments. It is clear that the patterns are very complex, and many peaks are detected. Many peaks likely contain multiple fragments. Given the expected complexity of the captured genome subset and the expected average size of genomic *Mse* I fragments (*i.e.*, ~160 bp), it is not likely that the entire CAG-containing *Mse* I genome subset is shown. Large fragments will not be efficiently amplified by PCR or, even if amplified, they may be outside the size range analyzed. The annealing of the inverted terminal repeats at the ends of the fragments may prevent PCR amplification of some

fragments [Siebert *et al.* (1995) *Nucleic Acids Res* 23:1087-1088; Lukyanov *et al.* (1995) *Anal. Biochem.* 229:198-202]. The hairpin structures formed by long CAG repeats [Mariappan *et al.* (1996) *Nucleic Acids Res.* 24:775-783; Gacy *et al.* (1995) *Cell* 81:533-540] are likely to interfere with annealing to complementary repeat capture probes or PCR primers [see below and Walsh *et al.* (1992) *PCR Meth. Appl.* 1:241-250; Demers *et al.* (1995) *Nucleic Acids Res.* 23:3050-3055]. Nevertheless, the results show many clear differences between monozygotic twin pairs (Fig. 1). As expected, differences between pairs were much greater than differences within pairs. Fewer differences between pairs were observed when a fluorescein-labeled adapter primer was used in combination with a repeat primer (data not shown). Differences between samples were enhanced when the complexity of the captured fragments was reduced further by selective PCR using adapter primers anchored at their 3'-end to unique sequences flanking the repeat sequence, or by cleaving the captured sample with a second restriction enzyme before PCR labeling (data not shown). These results are reproducible in replicated experiments.

Detection Of Specific Allelic Differences In The CAG-Containing Genome Subset

The genomic differential display method was tested for its ability to distinguish different HD alleles (Fig. 3). The HD sequence relevant to these experiments is shown in Fig. 2. Many differences were expected between the samples used in these experiments.

Thus, a number of control experiments were used to identify the HD-containing fragments.

First, the length of the HD CAG repeat was determined in three members of an HD-affected kindred (HD-A, -B, and -C) and an unrelated (control) individual (Fig. 3A). The control sample had two normal alleles, i. e., both alleles had <30 repeats, as did the HD-A sample. Both the HD-B and HD-C samples had two expanded alleles (>40 repeats); these HD homozygotes were reported previously [Wexler *et al.* (1987) *Nature* (London) 326:194-197]. The HD-B and HD-C samples also contained small amounts of fragments with shorter repeats. This type of HD mosaicism has been seen by others [Goldberg *et al.* (1993) *Nature Genet.* 5:174-179].

The next experiments (Fig. 3B) tested the ability of the method to distinguish between normal and expanded HD alleles in the absence of other genomic restriction fragments. The experiments took advantage of known *Sau96* I restriction enzyme recognition sites located near the HD-CAG repeat sequence. First, HD-specific PCR with primers (oligonucleotides 11 and 12; Table 1) flanking the repeat was used to generate HD-containing fragments with

Sau96 I recognition sites proximal to the repeat. The PCR products were digested with *Sau96* I, ligated to *Sau96* I adapters (oligonucleotides 5 and 6; Table 1) and amplified by PCR using a *Sau96* I adapter primer and a fluorescein-labeled repeat primer (oligonucleotides 5 and 13, respectively; Table 1). Clear differences were detected in the samples containing normal and expanded alleles (Fig. 3B). A heterogenous ~130 bp product, eluting at ~105 min (Fig 3B), was present in samples with normal HD alleles, but this fragment was absent in HD-B, and its amount was substantially reduced in HD-C. The expected size of the normal HD allele-containing fragments was ≥ 175 bases (the distance from the *Sau96* I site to the distal 3' end of the CAG repeat). An HD-containing fragment of about 130 bp long could contain 4-6 CAG repeats. This result suggests that during PCR variable length CAG repeats with a low number of repeats (<40) were converted to a constant length equal to that contained in the repeat primer [Weising *et al.* (1995) PCR Methods Appl. 4:249-255].

Since the repeat primer does not amplify across the repeat, one must explain why there are no HD fragments in the samples with expanded alleles (Fig. 3B). CAG repeat-containing fragments are known to form hairpin structures with stabilities that increase with repeat length [Mariappan *et al.* (1996) Nucleic Acids Res. 24:775-783; Gacy *et al.* (1995) Cell 81:533-540]. These structures appear to inhibit PCR amplification of expanded alleles as seen in these experiments and as reported before [Walsh *et al.* (1992) PCR Meth. Appl. 1:241-250; Demers *et al.* (1995) Nucleic Acids Res. 23:3050-3055].

A broad faint peak eluting at ~115 min (~150 bp) was seen in the samples with expanded HD alleles and in control experiments in the absence of templates. This peak is apparently generated by primer interactions when there is no competing complementary sequence, supporting the notion that the long HD repeats were not available for annealing to the repeat primers.

Finally, experiments similar to those shown in Fig. 1 were done to demonstrate that the genomic differential display procedure can distinguish between normal and expanded HD alleles even when multiple templates were present. These experiments analyzed *Sau96* I CAG-containing genome subsets expected to include the HD alleles (see above and Figs. 2 and 3C). The same basic procedure was used as that described for the experiments shown in Fig. 1, except for the change in restriction enzyme and the inclusion of the repeat primer. As expected, many differences were detected when the genomic DNAs from different individuals were compared. Since the fluorescent label was on the *Sau96* I adapter primer, two different HD-containing PCR products could be displayed. The longer fragment (~330 bp, elution time

~225 min) should be the PCR product formed by the adapter primer only. The shorter fragment (~155 bp, elution time 115 min) should be the PCR product formed by the adapter and repeat primer. Fragments eluting at ~225 min in HD-A and control samples were absent in the HD-B and HD-C samples (Fig. 3C). Small peaks at ~115 min were detected in the HD-A and control samples but not in the HD-B and HD-C samples (data not shown). As expected, PCR with adapter primers alone also amplified the ~330 bp fragments in the HD-A and control samples but not in the HD-B and HD-C samples (data not shown).

Additional experiments were done to prove that the ~330 bp fragments contained the normal HD allele. A 173-base sequence adjacent to, but not including, the CAG-repeat region within the HD gene (see above and Fig. 2) was used as an HD-specific capture probe. The CAG-containing genome subsets shown in Fig. 3C were hybridized to the HD-specific capture probe and then displayed (data not shown). As expected, the fragments captured and displayed from the control and HD-A samples eluting at ~227 min were not present in the HD-B and HD-C samples. Thus, the results in Fig. 3 show that our genomic differential display method is effective in distinguishing normal vs expanded CAG allele lengths.

This work characterized the different products that were generated when PCR reactions used adapter primers alone or adapter primers in combination with repeat primers. The greatest number of peaks was displayed when fluorescein-labeled adapter primers were used in combination with a repeat primer. In this case each genomic fragment can be potentially displayed twice; once as a PCR product of the adapter primers alone and second as a fragment amplified between the repeat and the adapter primer. The use of the adapter primers alone allows the simultaneous assessment of many short CAG-repeat lengths, whereas the use of adapter primers in combination with a repeat primer permits studies to focus on sets of unique sequences flanking trinucleotide repeats.

DESCRIPTION OF PREFERRED EMBODIMENTS

An example of genomic differential display using a two-step PCR protocol that targeted CAG-repeat containing genomic *Hae* III fragments is shown in Fig. 5. In these experiments, genomic DNA digested with *Hae* III, and ligated to adapters (oligonucleotides 1 and 2, Table 2) was amplified in a two-step PCR protocol. In the first PCR amplification, the A-primer was a 21-base oligonucleotide 3 (Table 2) corresponding to the outermost part of the ligated adapter and the T-primer was a fluorescein-labeled oligonucleotide 11 (Table 2) composed of 3' CTG-repeating sequence plus two unique bases at the 5'-end. The two bases

at the 5'-end were used to anchor the primer to unique sequences adjacent to the 3'-end of the CAG repeat sequence. PCR primers with 5'-anchors are not as selective as PCR primers and 3'-anchors [Broude *et al.* (1997); Ziethiewicz *et al.* (1994)].

The products of the first PCR were diluted and used as templates in a second PCR. The first and second PCRs used the same T-primers but different A-primers. The A-primer in the second PCR reaction was a 26- or 27-base oligonucleotide containing a 5'-terminal 22 base segment complementary to the innermost adapter sequence plus a 3'-terminal tetranucleotide (CCTT) or pentanucleotide (CCTTA, CCTTG, or CCTTT) sequence. The CC dinucleotide corresponded to the remainder of the *Hae* III recognition sequences. The 3' terminal TT, TTA, TTG and TTT bases (lanes 1, 2, 3 and 4, respectively, Fig. 5) anchored the A-primers to complementary genomic sequences adjacent to the *Hae* II recognition site. The dinucleotide and trinucleotide anchors should reduce the total genomic *Hae* III fragment complexity by, approximately, 16-fold and 64-fold, respectively.

Different genome subsets were displayed (Fig. 5) when the same T primer was used in combination with A-primers having the same adaptor sequence but with varying anchor sequences. PCR amplification depended on the presence of both an A- (data not shown) and a T-primer, since no PCR products were detected when these primers were used alone (Fig. 5, lane 5). PCR using an A-primer with a 2 base 3' anchor amplified several hundred fragments within the size range of ~100 to ~80 base pairs (Fig. 5, lane 1). Lengthening of the 3' terminal anchor reduced the number of amplified fragments (Fig. 5, lanes 2-4).

A complexity reduction was also achieved using 3' terminal anchors on the A- and T-primers simultaneously. These experiments used DNA samples from monozygotic twins. In one set of experiments (Fig. 6A), PCR amplification was done with a 3'-terminal 3 base anchored A-primer and a 3'-terminal 2 base anchored T-primer (*e.g.*, oligonucleotides 8 and 12, respectively, Table 2). In another set of experiments (Fig. 6B), PCR amplification was done using a 3'-terminal 3 base anchored A-primer and 5'-terminal 2 base anchored T-primer (oligonucleotides 8 and 11, respectively, Table 2). Clear differences were detected when DNAs from different twin pairs were compared. As expected, differences between pairs were much greater than differences within pairs. These experiments, along with a large number of control experiments, showed that T-primers with 3'- or 5'-anchors generated reproducible displays. Although the total lengths of the anchor sequences were the same for all the experiments shown in Fig. 6, many more peaks were observed with the 5'-terminally anchored T-primer (Fig. 6B). Complexity reduction should be directly proportional to the

length of the unique anchor sequence and the frequency of occurrence of the particular unique bases used in the anchor sequences. The experiments shown in Fig. 6A and 6B used the same 3'-anchored A-primer (oligonucleotide 8, Table 2). Therefore, it should be the 5'-anchored T-primer (Fig. 6B) which primes from many more sites than 3'-anchored T-primer (Fig. 6A). The 5'-anchored T-primer contained one degenerate and one unique base in the anchor, while the 3'-anchored T-primer contained two unique bases (compare oligonucleotides 11 and 12, respectively, Table 2). This difference can partially explain the greater number of peaks in Fig. 6A than in 3B. In addition, the weaker specificity expected for 5'-anchors can also contribute to the greater number of peaks in the experiment shown in Fig. 6B.

Nevertheless, both protocols gave highly reproducible results.

Reproducibility was assessed to establish the robustness of our genomic fingerprinting method. In control experiments, CAG-containing subsets were obtained from genomic DNA templates isolated several times from the same blood sample. PCR amplifications were done with 5' terminally anchored T-primer (oligonucleotide 11, Table 2) and different A-primers. PCR products displayed on the same PAAG were compared using methods under development for automatic analysis of displayed genomic subsets (Graber *et al.*, unpublished results). In brief, peak to peak comparisons were tabulated using a conditioned signal function calculated from dividing individual signal peak differences by the individual peak means after high pass filtering (to remove slowly varying backgrounds), low pass filtering (to remove rapidly varying background), and overall temporal alignment (and scaling) and normalization. The results showed that the PCR amplification itself was most critical for insuring reproducibility (versus DNA isolation, ligation and electrophoresis). PCR variability was minimized if the samples were amplified and compared simultaneously. Nevertheless, it is best to do each PCR amplification at least in duplicate and to load each sample in parallel on the same gel.

Sequence Analysis Of PCR Products

The sequences of ten randomly chosen cloned products from a *Sau96* I genome subset targeted to contain CAG-repeating sequences were determined (Table 3). In this experiment, the first PCR amplification used oligonucleotides 3 and 11, and the second PCR amplification used oligonucleotides 13 and 11 (Table 3) as the A- and fluorescein-labeled T-primers, respectively. All of the clones contained 6-8 perfectly, or imperfectly, repeated CAG trinculeotides at one termini. Clones 3, 4, 5, 8 and 10 also contained multiple scattered CAG

repeats near the tandem repeat. Scattered CAG repeats are commonly found near tandem CAG repeats [Armour *et al.*, *Hum. Mol. Genet.* 3:599-605 (1994)]. Six clones contained the exact repeat-primer sequence and four lacked the 5'-distal anchoring base (Table 3). The sequences shown in Table 3, minus the adapter sequences, were filtered and compared to DNA sequences in Genbank using BLAST. Filtering removed tandem repeat sequences from the analysis to avoid spurious alignments based on low complexity sequences. Clones 3 and 4 matched previously identified sequences containing CAG-repeats [Armour *et al.* (1994)]. Clone 10 contained a gene fragment with a *SAU96* I restriction site upstream of the CAG-repeat sequence [Chiba *et al.*, *Nucleic Acids Res.* 22:1815-1820 (1994)]. These results confirmed that the targeted CAG repeat-containing genome subset was obtained by PCR amplification directly from genomic DNA.

The prevalence of short repeat lengths in the cloned sequences correlates with the predominance of short, 5-7 CAG, repeats in the genome [Stallings, *Genomics* 21:116-121 (1994); Gastier, *Genomics* 32:75-85 (1996); Neri *et al.*, *Hum. Mol. Genet.* 5:1001-1009 (1996)]. However, the results are also consistent with the instability of long repeat lengths in *Escherichia coli* [Wells, *J. Biol. Chem.* 271:2875-2878 (1996)] and the preferential PCR amplification of short repeat lengths [Broude *et al.* (1997); Demers *et al.*, *Nucleic Acids Res.* 23:3050-3055 (1995); Haddad *et al.*, *Hum. Genet.* 97:808-812 (1996)]. Increasingly stable hairpin structures form from CAG- or CTG-repeating sequences of increasing length [for review see Wells (1996)]. Hence, inhibition of PCR amplification by these hairpins increases with repeat length. This means that short repeat sequences are preferentially amplified [Broude *et al.* (1997); Demers *et al.* (1995); Haddad *et al.* (1996)], and are distinguishable from long repeat sequences [Broude *et al.* (1997); Haddad *et al.* (1996)].

The complexity of genome subsets analyzed by the method described here is modulated by the total number of occurrences of the targeted repeat sequence, the restriction fragment size distribution, the sensitivity of the restriction enzyme to methylation, and the location (5'- or 3'-terminal), composition and length of the anchor sequences on the primers. The use of anchored PCR primers allows for a controlled complexity reduction (Fig. 5 and 6) so that particular fragments of interest can be isolated from gels containing low complexity subsets.

Restriction fragment length polymorphisms may be due to polymorphisms in the repeat sequence or in the unique flanking sequences. Since the T-primer can be anchored with unique sequence at its 3'- or 5'-end, sample comparisons can be focused solely on

unique sequences flanking a targeted repeat sequence or on unique sequence flanking the repeat plus the repeat sequence itself, respectively.

There are a large number of potential applications of this approach. For instance, ongoing experiments are searching for genomic differences between tumors and normal tissue and between monozygotic twins. Furthermore, this approach can be used to focus conventional differential display of cDNAs to analyze cDNA containing targeted common sequence.

Methods

Nonphosphorlated oligonucleotides (Table 2) were purchased from Operon Technologies (Alameda, CA). Genomic DNA was isolated from blood lymphocytes from anonymous donors and monozygotic twin samples obtained from E. Fuller Torrey by a standard phenol-based extraction procedure.

For genomic differential display, isolated genomic DNA was digested with a restriction enzyme and ligated to oligonucleotides of known sequence as described [Broude *et al.* (1997)]. Briefly, human genomic DNA (300 ng) was digested with 10 units of *Hae* III or *Sau*96 I at 37 C overnight in a 100 µl total volume reaction. The recessed *Sau*96 I ends were filled in using AmpliTaq DNA polymerase (Perkin-Elmer). DNA was precipitated by ethanol, dissolved in 20 µl of sterile water, and blunt-end ligated to an excess (2 µM) of each adapter oligonucleotide (oligonucleotides 1 and 2, Table 2) at 16 C overnight in a 30 µl final volume, containing 50 mM Tris-HCl, pH 7.6/10 mM MgCl₂/0.5 mM ATP/10 mM dithiothreitol/ and 5 units of T4 DNA ligase (Life Technologies, Gaithersburg, MD). This reaction was done with uneven adapter lengths to insure all the adapter oligonucleotides were ligated to the genomic fragments with the same polarity. The ligation reactions produced genomic restriction fragments with 5' twenty-six base single-stranded overhangs. The ligation was terminated by incubation at 75 C for 5 min. DNAs were then purified from excess primers by passing the samples through Wizard DNA purification columns (Promega, Madison, WI). DNA was eluted into 50 µl of sterile water.

The prepared DNA (3-5 ng) was amplified by PCR in 50 µl reaction volume in PCR buffer II (10 mM Tris-HCl, pH 8.3/ 50 mM KCl) from Perkin-Elmer, plus 2.5 mM MgCl₂, 250 µM of each dNTP and 2.5 units of AmpliTaq DNA polymerase (Perkin-Elmer). Hot start PCR was performed at 94 C by adding 0.2 µM each of the adapter-primer (A-primer)

and fluorescein-labeled CTG-containing target-primers (T-primers), *e.g.*, oligonucleotides 3 and 11 or 12, respectively (Table 2).

5 PCR mixtures were subjected to 20 - 25 amplification cycles consisting of incubations at 94 C for 3 sec, 65 C for 20 sec and 72 C for 30 sec in the PTC-100™ Temperature Cycler (MJ Res. Inc., MA). The products of this first PCR were diluted 1000-fold and used as templates for a second PCR amplification. T-primers in the first and second PCR were oligonucleotides 3 and one of the oligonucleotides designated 4-10, respectively (Table 2). The PCR products (1-2 µl) were denatured for 3 min at 90 C in a stop solution (Pharmacia-Biotech, Sweden) containing 6 mg/ml of dextran blue and 0.1% sodium dodecyl sulfate in
10 deionized formamide, loaded onto a 6% denaturing polyacrylamide gel (PAAG) and analyzed on the ALF DNA sequencing instrument (Pharmacia-Biotech, Sweden). The results were visualized using the Fragment Manager software provided with the instrument. Fluorescein-labeled 50 - 500 base pair (bp) ladder (Pharmacia-Biotech, Sweden) was used as a size marker.

15 The specificity of the PCR amplification was investigated by cloning and sequencing randomly chosen amplification products obtained from *Sau*96 I-digested DNA. Oligonucleotides 11 and 13 (Table 2) were used as T- and A-primers, respectively. The PCR products from the second PCR amplification were cloned using a TA cloning kit (Invitrogen, San Diego, CA). Plasmid DNAs were isolated and sequenced using a Sequenase 2.0 kit
20 (Pharmacia-Biotech, Sweden) and an ALF sequencing instrument.

TABLE 2

Oligonucleotides Used (5'-3')

No.	Description	Sequence
1	adaptor-40	TGTAGCGTGAAGACGACAGAAAGGGCGTG- GTGCGGAGGGCGGT
2	adaptor-11	ACCGCCCTCCG
3	5'-A primer	TGTAGCGTGAAGACGACAGAA
4	3'-A primer	AGGGCGTG GTGCGGAGGGCGGT
5	<i>Hae</i> III 3'-A primer TT	AGGGCGTG GTGCGGAGGGCGGTCCTT
6	<i>Hae</i> III 3'-A primer TTT	AGGGCGTG GTGCGGAGGGCGGTCCTTT
7	<i>Hae</i> III 3'-A primer TTC	AGGGCGTG GTGCGGAGGGCGGTCCTTC
8	<i>Hae</i> III 3'-A primer TTA	AGGGCGTG GTGCGGAGGGCGGTCCTTA
9	<i>Hae</i> III 3'-A primer TTG	AGGGCGTG GTGCGGAGGGCGGTCCTTG
10	<i>Hae</i> III 3'-A primer5 TTCTG	AGGGCGTG GTGCGGAGGGCGGTCCTTCG
11	T-primer1	A(A/C/T)(CTG) ₆
12	T-primer2	(CTG) ₆ AC
13	<i>Sau</i> 96 I 3'-A primer GA	AGGGCGTG GTGCGGAGGGCGGTGCCCCGA

TABLE 3

DNA Sequences Of Randomly Chosen Fragments
Amplified By Two-Step PCR From Genomic DNA

Clone No.	Length Of Insert (bp)	DNA Sequence ^a
1	133	AT(CTG) ₇ TGGAACAAACACTCCAGGATTTAGC...
2	153	AT(CTG) ₆ CAATTATTATGAA...
3 ^{b, c}	209	ACTGTTGCTGTGGAAACTGTAGCTGTTGCTG- TGAAAACCTGATGCGGATGGGCTTG(CTG) ₅ TTGCTG...
4 ^d	255	AA(CTG) ₆ CCAG(CTG) ₂ GGTCAGGATAGGAGAGGCTG...
5	193	A(CTG) ₆ CCATAACTGCTTCCACTCCAAAGGC...
6	147	AA(CTG) ₆ TGGAACAAACACTCCAGGATC...
7	120	C(CTG) ₆ GCCTGAAGACCACATTCTGAGAACCCCTGA...
8	216	AA(CTG) ₆ GTGGGG(CTG) ₉ GTGCTGGTGGTGGGG(CTG) ₇ ...
9	126	AT(CTG) ₆ GCCTGACTACTGACCACA...
10 ^e	102	A(CTG) ₆ TGTTGGCTGTGCTGGTTGTTGCTG...

^a Only the terminal sequence including the repeat sequence in the T-primer is shown. Genbank accession numbers for the entire sequences for clones 1-10 are U92822-U92824, U92826-U92832, respectively. Sequence 10 is a fragment of transcriptional activator hSNF2a gene (*see below*).

^b Clone 3 did not contain repeat sequence of the T-primer.

^c BLAST homology score is 4xe-28 to accession number X76572.

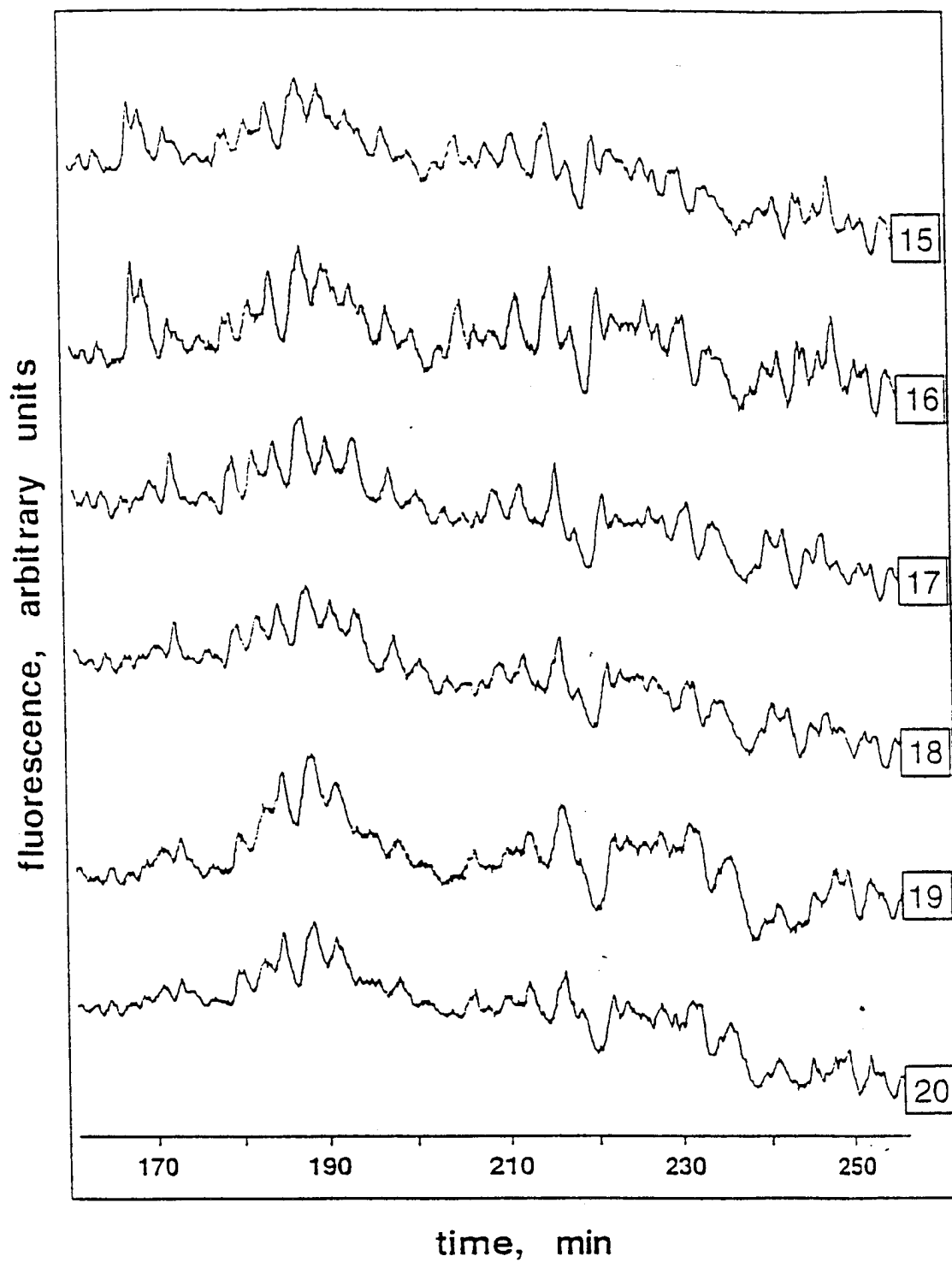
^d BLAST homology score is 5.6xe-72 to accession number X73969.

^e BLAST homology score is 8.4xe-27 to accession number D26155.

CLAIMS

We claim:

1. A method of analyzing nucleic acid in a sample, comprising:
 - a) providing: i) a sample containing nucleic acid, ii) a first oligonucleotide primer comprising a CTG repeat, iii) a second oligonucleotide primer comprising a CAG-repeat, and iv) a polymerase and PCR reagents;
 - b) preparing said nucleic acid from said sample under conditions so as to produce amplifiable nucleic acid;
 - c) amplifying said nucleic acid with said first and second primers, said polymerase and said PCR reagents under conditions such that amplified product is generated; and
 - d) detecting said amplified product.
2. The method of Claim 1, wherein said first oligonucleotide primer has the formula $(CTG)_n$, wherein n is a whole number between 1 and 15.
3. The method of Claim 2, wherein n is between 2 and 12.
4. The method of Claim 3, wherein n is between 6 and 12.
5. The method of Claim 1, wherein said first oligonucleotide primer has the formula $(CAG)_n$, wherein n is a whole number between 1 and 15.
6. The method of Claim 5, wherein n is between 2 and 12.
7. The method of Claim 6, wherein n is between 6 and 12.



9.0 x 7.5 cm, reduction 1:2

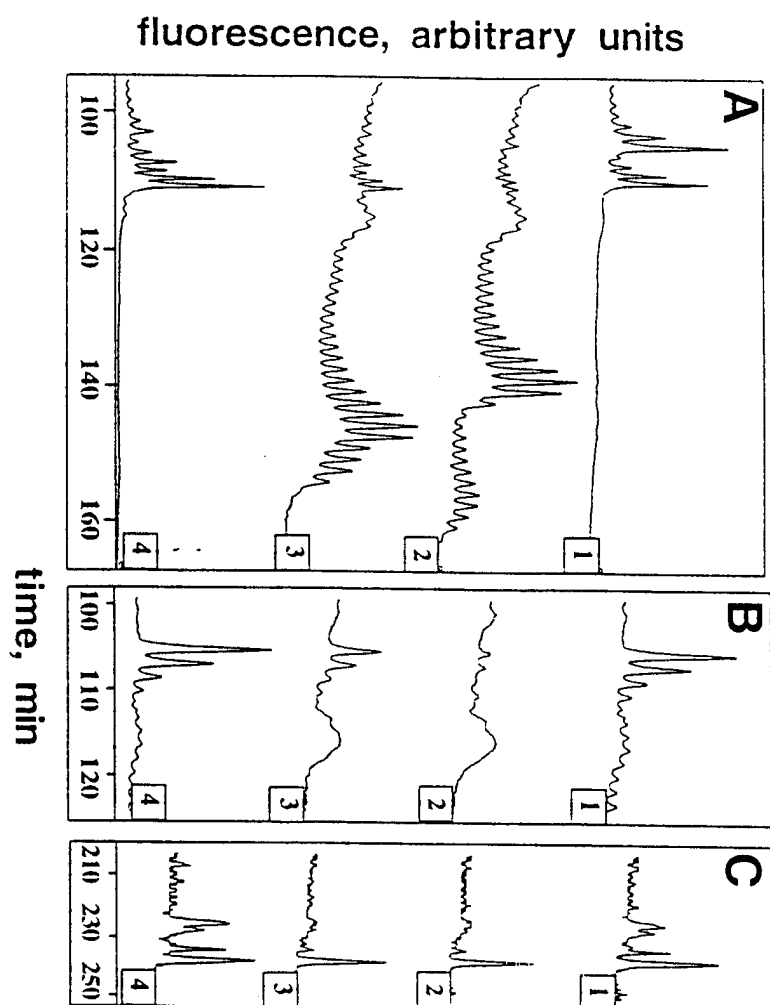
FIG. 1

1 / 6

-----11-----> ↓
3440 GGGACGGGTC CAAGATGGAC GGCCGCTCAG GTTCTGCTTT TACCTGCGGC CCAGAGCCCC 3499
↓
3500 ATTCAATTGCC CCGGTGCTGA GCGGCGCCGC GAGTCGGCCC GAGGCCTCCG GGGACTGCCG 3559
-----9-----
3560 TGCCGGGCGG GAGACCGCCA TGGCGACCCT GGAAAAGCTG ATGAAGGCCT TCGAGTCCCT 3619
<-----10-----
-----9----->
3620 CAAGTCCTTC **CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG CAGCAGCAGC AGCAGCAGCA** 3679
---10-----
-----12-----
3680 **GCAGCAGCAG CAGCAACAGC** CGCCACCGCC GCCGCCGCCG CCGCCGCCTC CTCAGCTTCC 3739
-----12-----
3740 TCAGCCGCCG CCGCAGGCAC AGCCGCTGCT GCCTCAGCCG CAGCCGCCCC CGCCGCCGCC 3799
↓
3800 CCGCGCGCCA CCGGCCCCGG CTGTGGCTGA GGAGCCGCTG CACCGACCGT GAGTTTGGGC 3859

FIG. 2

Reduction 1:2



reduction 1:0.75

FIG. 3

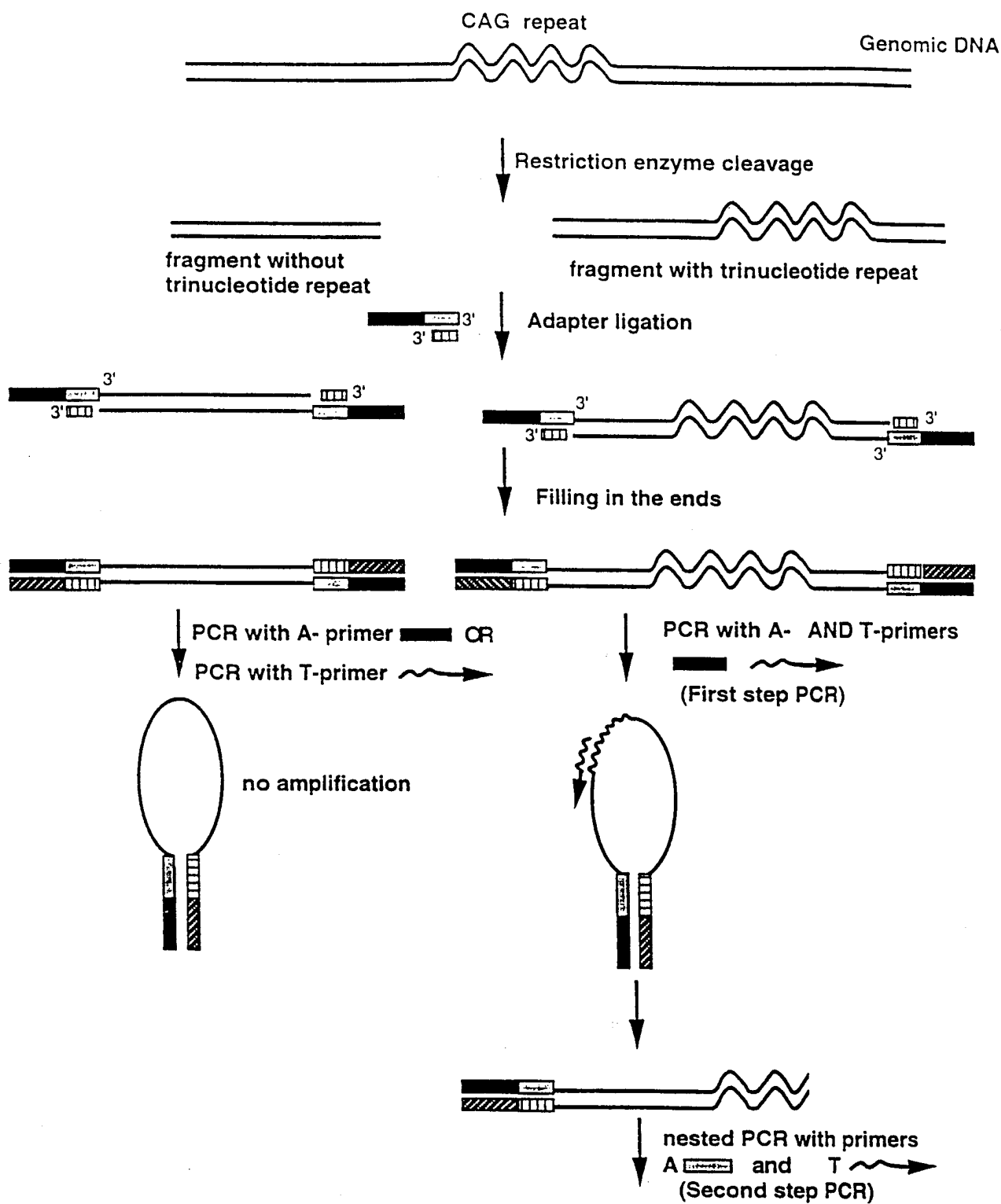


FIG. 4

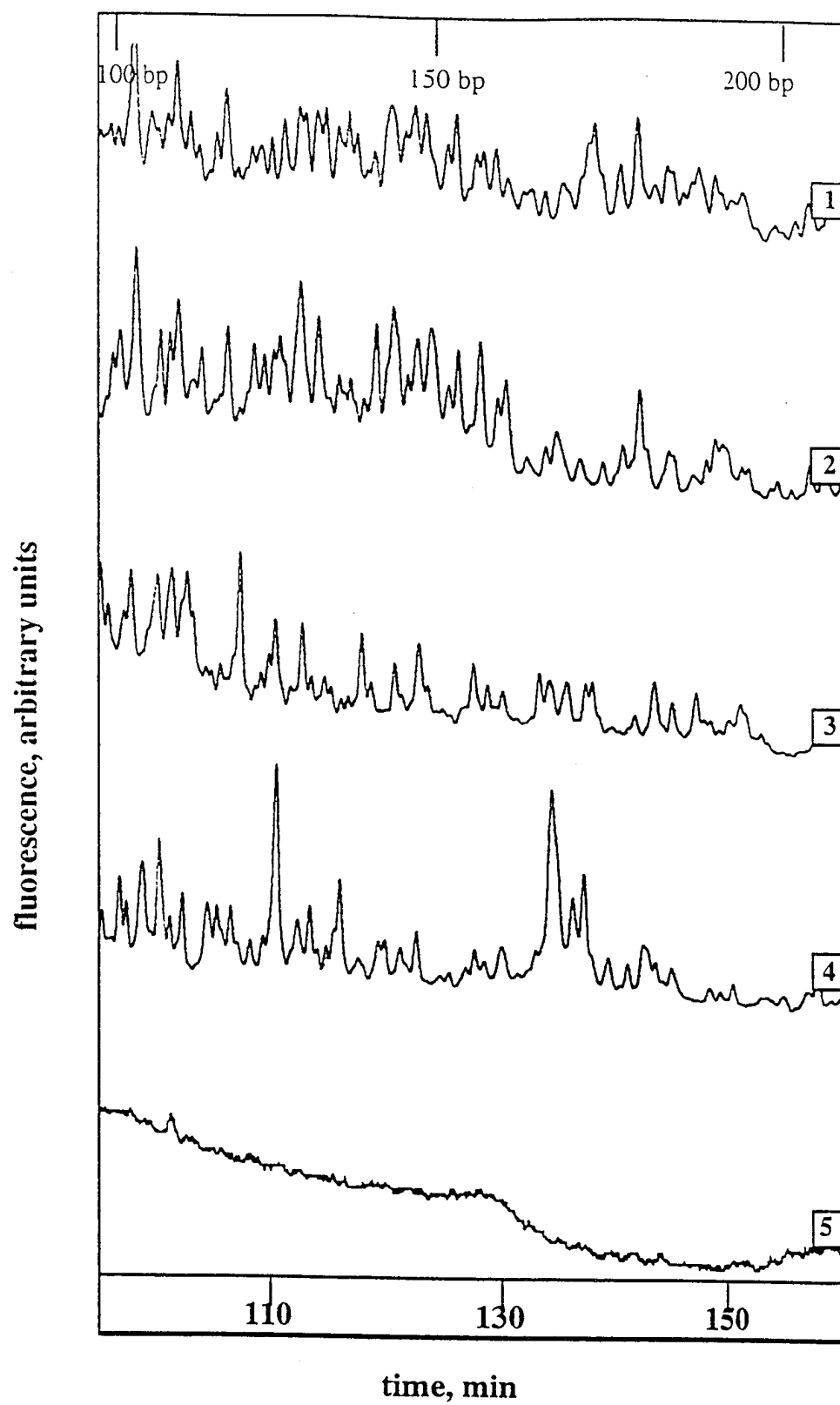


Fig. 5

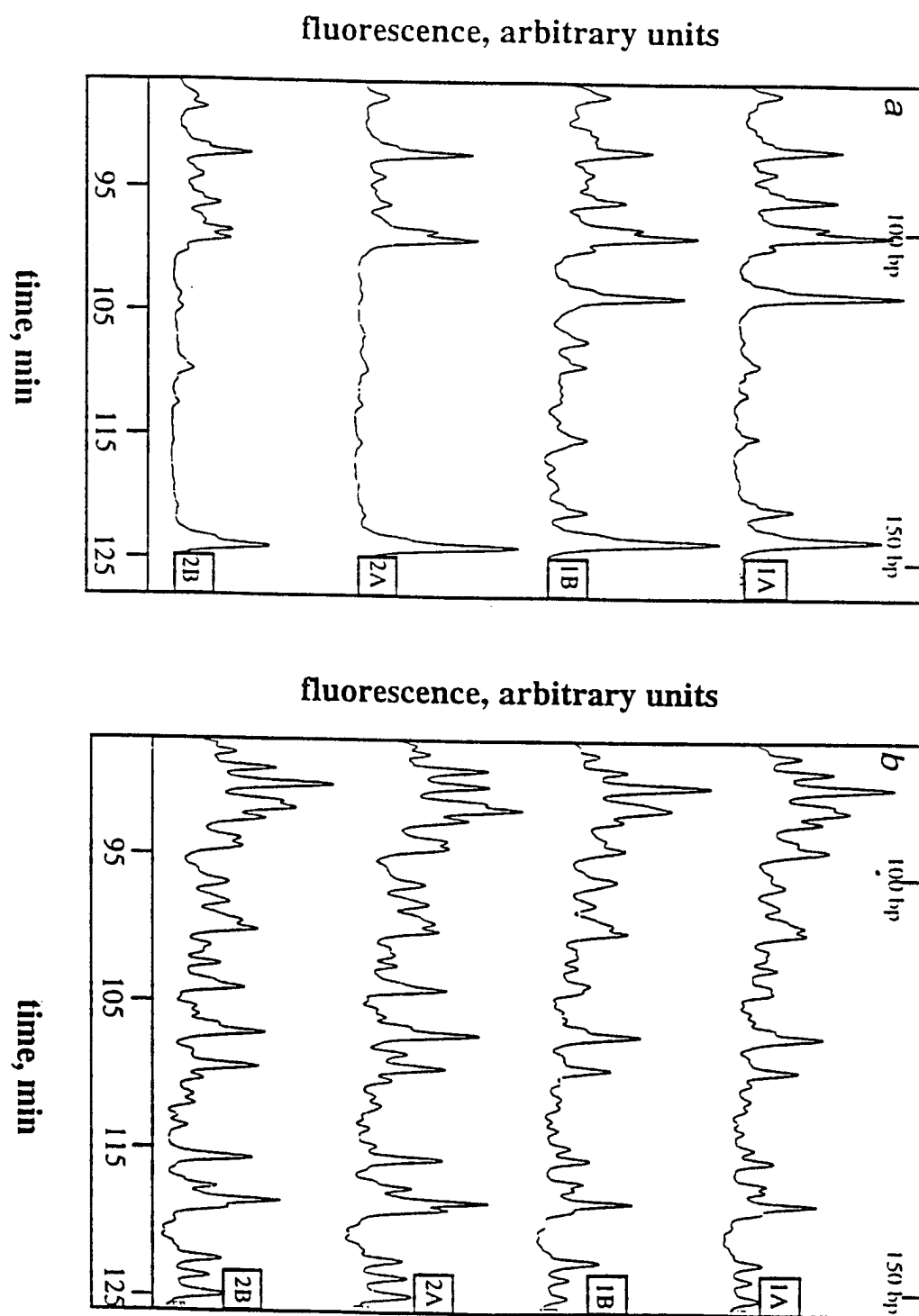


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/08616

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; G01N 33/53; C12P 19/34;

US CL :435/6, 7.1, 91.1, 91.2

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, 7.1, 91.1, 91.2; 536 24.2, 24.3; 935 77, 78

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 Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BROUDE et al. Differential Display of genome subsets containing specific interspersed repeats. Proc. Natl. Acad. Sci. April 1997, Vol. 94, pages 4548-4553, especially, pages 4549-4550.	1-3,5-7
Y	SUTHERLAND, G.R. Simple tandem DNA repeats and human genetic disease. Proc. Natl. Acad. Sci. April 1995, Vol. 92, pages 3636-3641, especially, page 3638.	1-7
Y	US 4,683,202 A (MULLIS) 28 July 1987 (07/28/87), see column 2, lines 37-53.	1-7
Y	WEISING et al. Genomic Fingerprinting by microsatellite-primed PCR: A critical Evaluation. PCR Methods and Applications. 1995, pages 249-255, especially, page 250.	1-7



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not 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	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"B" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" 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)	"&*" document member of the same patent family
"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

03 AUGUST 1998

Date of mailing of the international search report

04 SEP 1998

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

Jeffrey Siew

Telephone No. (703) 305-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US98/08616

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	ZIETKIEWICZ et al. Genome Fingerprinting by Simple Sequence Repeat (SSR)- anchored Polymerase Chain Reaction Amplification. 1994, Vol. 20, pages 176-183, especially, page 177.	1-7
Y, P	US 5,741,645 A (ORR et al) 21 April 1998 (04/21/98), see column 3, lines 1-15.	1-7
Y, P	WO 97/17469 A1 (DANA FARBER CANCER INSTITUTE) 15 May 1997 (05/15/97), see page 8, line 24-35.	1-7
A	MATZ et al. Ordered differential display: A simple method for systematic comparison of gene expression profiles. Nucleic Acids Research. 1997, Vol. 25, No.12, pages 2541-2542, especially, page 2541.	1-7
A	SIEBERT et al. An improved PCR method for walking in uncloned genomic DNA. Nucleic Acids Research. 1995, Vol. 23, No. 6, pages 1087-1088, especially, page 1087.	1-7
A	LUKYANOV et al. Inverted Terminal Repeats Permit the Average Length of Amplified DNA Fragments to be Regulated during Preparation of cDNA Libraries by Polymerase Chain Reaction. Analytical Biochemistry. 1995, Vol. 229, pages 198-202, especially, page 200.	1-7
A	US 5,565,340 A (CHENCHIK et al) 15 October 1996 (10/15/96), see abstract, figure 1.	1-7
A	US 5,599,672 A (LIANG et al) 04 February 1997 (02/04/97), see abstract, column 3, lines 25-54.	1-7
A	US 5,580,726 A (VILLEPONTEAU et al) 03 December 1996 (12/03/96), see abstract, column 2, lines 20-44.	1-7
A	DIATCHENKO et al Suppression subtractive hybridization: A method for generating differentially regulated or tissue specific cDNA probes and libraries. Proc. Natl. Acad. Sci. June 1996, Vol. 93, pages 6025-6030, especially, page 6026.	1-7
A	US 5,459,037 A (SUTCLIFFE et al) 17 October 1995 (10/17/95), see abstract.	1-7

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/08616

B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS

STN CAPLUS, MEDLINE, BIOSIS, WPIDS, WPINDEX, GENBANK, BIOTECHABS, SCISEARCH, EMBASE, CANCERLIT, JPIO, USPATFULL

SEARCH TERMS: PCR OR POLYMERASE CHAIN REACTION SUPPRESS, (CAG OR CTG) REPEAT, TRINUCLEOTIDE REPEAT, GENOMIC DIFFERENTIAL DISPLAY, DIFFERENTIAL DISPLAY, (ANALYSE OR DETECT) POLYMORPHISM,