PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12Q 1/68

A1

(11) International Publication Number: **WO 95/25179**

(43) International Publication Date: 21 September 1995 (21.09.95)

(21) International Application Number: PCT/US95/02861

(22) International Filing Date: 8 March 1995 (08.03.95)

(30) Priority Data:
08/214,823 17 March 1994 (17.03.94) US

(71) Applicant: UNIVERSITY OF MASSACHUSETTS MEDICAL CENTER [US/US]; 55 Lake Avenue North, Worcester, MA 01655 (US).

(72) Inventors: SINGER, Robert, H.; 55 Holman Street, Shrewsbury, MA 01545 (US). TANEJA, Krishan, L.; 106 Francis Avenue, Shrewsbury, MA 01545 (US).

(74) Agent: CREASON, Gary, L.; Fish & Richardson P.C., 225 Franklin Street, Boston, MA 02110-2805 (US).

(81) Designated States: AU, CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: DETECTION OF TRINUCLEOTIDE REPEATS BY *IN SITU* HYBRIDIZATION

(57) Abstract

Disclosed is a method of detecting a trinucleotide repeat expansion by *in situ* hybridization. The disclosed method uses a sample of nucleated cells, a labelled trinucleotide repeat-specific probe and detection of the hybridized probe by a means whose sensitivity distinguishes between the signal from probes hybridized to an expanded repeat and the signal from probes hybridized to a non-expanded repeat.

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.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

DETECTION OF TRINUCLEOTIDE REPEATS BY IN SITU HYBRIDIZATION

Field of the Invention

The invention relates to detection of genetic abnormalities and diagnosis of genetic diseases.

Background of the Invention

Trinucleotide repeat sequences in transcripts of affected genes have been found in fragile X-syndrome ("Fra-X;" CGG repeats in the 5′ untranslated region), muscular atrophy ("SBMA;" CAG repeat in the coding region), myotonic dystrophy ("DM;" CTG repeat in the 3′ untranslated region), and Huntington's disease ("HD;" CAG repeat in the coding region). These repeats appear to be present in the normal gene as well, but the number of tandem trinucleotide repeats is increased in the disease state ("trinucleotide repeat expansion"). The extent or degree of trinucleotide repeat expansion associated with the disease state varies for different genetic diseases, but in each case there appears to be a relatively consistent number of tandem repeats below which the patient has no symptoms and above which disease symptoms begin to appear. In general, increasing severity of disease symptoms correlates with an increased degree of repeat expansion (i.e., a greater number of tandem trinucleotide repeats) once the maximum normal number is exceeded (Harley *et al.*, 1992, Nature, **355**:545-46; Buxton *et al.*, 1992, Nature, **355**:547; Aslanidis *et al.*, 1992, Nature, **355**:548; Brook *et al.*, 1992, Cell, **68**:799-808; Mahadevan, 1992 Science, **255**:1253-56; Fu *et al.*, 1992, Science, **255**:1256-58; Tsilfidis *et al.*, 1992, Nature Genetics, **1**:192-95).

The mechanisms for trinucleotide repeat expansion are not known, but many of the genetic diseases associated with this phenomenon exhibit "anticipation". That is, the severity of symptoms increases in succeeding generations, suggesting that replication errors may contribute to the repeat expansion.

An example of such a disease is myotonic dystrophy (DM), a human neuromuscular genetic disease inherited in an autosomal dominant fashion. The genetic defect has multisystem effects, including myotonia and weakness, cardiac conduction defects, cataracts, male baldness, hypersomnia, abnormal glucose response and male testicular atrophy as well as abnormalities in other systems. The clinical presentation of myotonic dystrophy is variable and has been well characterized (P.S. Harper, 1989, Myotonic Dystrophy, Saunders, London, and Philadelphia, 2nd ed.). While the genetic bases of the disease are not known, the trinucleotide repeat sequence $(CTG)_n$ has been found in the 3'-untranslated region of myotonic-protein kinase (Mt-PK) mRNA. The severity of the disease may increase from one generation to the next (anticipation) and is related to expansion of the $(CTG)_n$ repeat sequence.

Biochemical studies have not shown any mutated or defective protein associated with myotonic dystrophy, but defects in membrane structure and function have been found. There is also evidence of reduced phosphorylation of membrane proteins in red blood cells (Roses *et al.*, 1973, PNAS, **70**:1855) and sarcolemmal membranes from muscle biopsies of DM patients (Roses and Appel, 1974, Nature, **250**:245). Fu *et al.*, 1993, Science, **260**:235-38 have shown that the amount of Mt-PK mRNA and the corresponding protein decreases with increased repeat expansion in the myotonic dystrophy patient. The regulatory role of protein kinase in development and the

physiological modulation of channel proteins is also reduced in myotonic dystrophy patients (J. Wang *et al*, 1992, Nature, **359**:739; J.W. West, 1991, Science, **254**:866). Fu *et al.*, *supra*, have suggested that the decrease in myotonic protein kinase contributes to the severity of the disease by disrupting signal transduction and amplification pathways. In contrast, another study has shown no difference in mRNA levels in myotonic dystrophy patients (Brook *et al.*, 1992, Cell *supra*).

The CTG trinucleotide repeat sequence is polymorphic in the normal population and undergoes various degrees of expansion in myotonic dystrophy patients (Brook *et al.*, 1992, Cell, **68**:799). The average number of CTG repeats in normal cells is about 5 (48%) - 27, but DM patients have at least 50 copies to several hundred copies. More severe cases are associated with higher number of repeats. One possible explanation for the expansion of the trinucleotide repeat may be errors in DNA replication during meiotic cell division or in the rapidly dividing cells of the early embryo. That is, replication of five trinucleotide repeat alleles may be stable, whereas duplication or triplication may occur when 27 repeat alleles are involved due to error in the DNA replication step from one generation to the next. The CTG trinucleotide repeat is transcribed from the gene and is located about 500 bp upstream of the poly(A) signal in the mRNA. The gene is expressed in many tissues of the myotonic dystrophy patient and encodes a protein (Mt-PK) having a strong homology with the protein kinase gene family. Normal Mt-PK protein is encoded by a gene having a genomic sequence of 11.5 kb. The gene contains 14 exons and has been mapped to chromosome 19. It is not known at the present time whether expansion of the trinucleotide repeat affects transcription, transport or function of the mRNA.

Many genes and RNAs contain sequences similar or identical to the trinucleotide repeats known to be expanded in genetic diseases. Probes and primers directed to the repeat sequence hybridize to these sequences, which are unrelated to the genetic disease of interest, creating smears on Northern and Southern blots or producing non-specific target amplification. For example, both the rRNA genes and the histone genes are GC rich and can be expected to hybridize to trinucleotide repeat probes. Probes to the trinucleotide repeat sequences have therefore previously been used only for hybridization to isolated nucleic acid segments, such as for screening cDNA libraries (Li *et al.*, 1992, Am. J. Hum. Genet., **51**:(4 Suppl.), A41; Riggins *et al.*, 1992, Am. J. Hum. Genet., **51**:(4 Suppl.), A41). Because a variety of short trinucleotide repeat sequences can be found all over the genome, previous methods for detection and analysis of repeat expansion in uncloned DNA have focused on the use of probes and primers which hybridize to unique sequences flanking or otherwise closely linked to the trinucleotide repeat of interest. This approach for specific analysis of repeat expansion in a gene of interest has been applied to diagnosis of Huntington's Disease (The Huntington's Disease Collaborative Research Group, 1993, Cell, **72**:971-983; Goldberg *et al.*, 1993, Human Molec. Genet., **2**:635-636), X-linked spinal and bulbar muscular atrophy (SBMA) (Yamamoto *et al.*, 1992, Biochem. Biophys. Res. Commun., **182**:507-513) and to identify polymorphisms in cloned sequences containing trinucleotide repeats (Riggins *et al.*, 1992, *supra*). Warner *et al.* 1993, Molec. Cell Probes, **7**:235-239 have reported a polymerase chain reaction (PCR) assay for detection of the trinucleotide repeat associated with Huntington's disease. This PCR method employs one primer which spans the repeat and a GC rich region of the gene,

but retains amplification specificity by directing the second primer to a unique flanking sequence. Repeat-specific oligonucleotides have also been used to detect expanded repeats in the genome by Repeat Expansion Detection (RED - Schalling *et al.*, 1993, Nature Genet., **4**:135-39). RED is similar to the ligase chain reaction in that repeat-specific oligonucleotides are cyclically hybridized to repeats in the genome, ligated and denatured. Only long repeats in the target DNA can serve as templates for adjacent annealing of multiple complementary oligonucleotides, reportedly eliminating detection of non-expanded trinucleotide repeats elsewhere in the genome. Several of these prior art methods have been applied to diagnosis of myotonic dystrophy. See, for example, Fu *et al.*, 1992, Science, *supra*; Mahadevan *et al.*, 1992, Science, *supra*; Hecht *et al.*, 1993, Clin. Genet., **43**:276-285; Brook *et al.*, 1992, Cell, **68**:799-808; WO 93/17104; WO 93/16196.

As discussed above, it is possible that the trinucleotide repeat sequence does not affect the transcription of the DM gene, but rather interferes with mRNA processing or transport to the cytoplasm. This would explain the reduction of Mt-PK mRNA and protein, and a similar abnormality has been reported in the double sex mutant of Drosophila, in which the repeat sequence binds to the protein involved in mRNA processing (Nagoshi, *et al.*, 1990, Genes Dev., **4**:89). Alternatively, the Mt-PK mRNA may be transported normally but may be dysfunctional in the cytoplasm. It is not possible, using prior art methods, to determine which of these mechanisms is operative.

Summary of the Invention

We have discovered that expansions of normally-occurring trinucleotide repeats in the human genome, which expansions have been associated with various genetic diseases (e.g., fragile X-syndrome, muscular atrophy, myotonic dystrophy and Huntington's disease) can be specifically detected by *in situ* hybridization.

Accordingly, the invention features a method for detecting an expansion of a trinucleotide repeat sequence by *in situ* hybridization using a sample of nucleated cells,

a labelled trinucleotide repeat-specific oligonucleotide probe, and detection of the hybridized probe by a means whose sensitivity distinguishes between the signal from probes hybridized to an expanded repeat and the signal from probes hybridized to a non-expanded repeat (e.g., the normal allele of the affected gene).

The invention may be practiced with a single probe. If a DNA denaturation step is not included in the *in situ* hybridization procedure, an antisense probe must be used, i.e., to detect RNA transcripts of the gene. If a DNA denaturation step is included, the probe may be either a sense probe or an antisense probe. Alternatively, the invention may be practiced with two different probes for simultaneous differential detection of the coding strand of a trinucleotide repeat expansion-containing gene and transcripts from that gene.

As used herein, "antisense probe" means a probe that hybridizes to RNA transcripts and to the non-coding strand of the (genomic) DNA of the gene of interest. For example, an antisense probe for detecting the trinucleotide repeat expansion associated with myotonic dystrophy comprises the trinucleotide repeat $(CAG)_n$ (which is complementary to the $(CTG)_n$ found in the DM mRNA).

As used herein, "sense probe" means a probe that hybridizes only to the coding strand of the (genomic) DNA of the gene of interest, and not to RNA transcripts of that gene.

As used herein, "coding strand" means the genomic DNA strand that is the template for RNA production, i.e., the coding strand is complementary to mRNA.

As used herein, "non-coding strand" means the genomic DNA strand whose sequence corresponds to the sequence of RNA produced by the process of gene transcription.

Other features and advantages of the invention will become apparent from the detailed description that follows, and from the claims.

Brief Description of the Drawings

Figs. 1A - 1F: Detection of Trinucleotide Repeat (CTG) Sequences. Fibroblasts from normal patients and patients with congenital DM containing an unstable expanded CTG repeat were plated on gelatin coated coverslips and either fixed with 4% paraformaldehyde or prepared for a cytogenetic analysis. The preparation was heated in 70% formamide, 2X SSC at 70° for two minutes to denature the chromosomal DNA and quenched in cold 70% ethanol for five minutes and then cold 100% ethanol for five minutes. Cells were air dried before hybridization. Two synthetic probes, sense (CTG-30; SEQ ID NO: 15) and antisense (CAG-30; SEQ ID NO: 14) to trinucleotide repeat (CTG) sequence, were labelled with separate fluorochromes. Normal and DM fibroblasts and the cytogenetic preparation were hybridized to the fluorescein labelled CAG-30 (SEQ ID NO: 14) or Texas red labelled CTG-30 (SEQ ID NO: 15), and visualized by epifluorescence microscopy.

Fig. 1A and Fig. 1B are DM fibroblast cells isolated from patients which have been hybridized with fluoresceinated antisense, CAG-30 probe (SEQ ID NO: 14), showing a number of discrete foci in the nucleus (arrowheads) (A = Patient No. 3132, B = Patient No. 3755).

Fig. 1C is the same as Fig. 1A, but Texas red labelled sense CTG-30 probe (SEQ ID NO: 15) was used. No signal was detected.

Fig. 1D is a cytogenetic preparation from 3132 which was hybridized with fluoresceinated antisense CAG-30 probe (SEQ ID NO: 14). The nuclei contain a variable number of discrete foci. Chromosomal signal was not detected. The signal is due to transcripts containing expanded CTG repeat sequences.

Fig. 1E is the same as Fig. 1D, but the Texas red labelled sense, CTG-30 probe (SEQ ID NO: 15) was used. No signal was detected in the nucleus and chromosomes.

Fig. 1F is a photograph of human diploid fibroblast cells which were hybridized with fluoresceinated antisense CAG-30 probe (SEQ ID NO: 14). No signal was detected.

Figs. 2A - 2C The probes CAG-30 (SEQ ID NO: 14) and CTG-30 (SEQ ID NO: 15) were labelled with digoxigenin using terminal transferase and dig-11-dUTP. Digoxigenin labelled probes were hybridized to the paraformaldehyde fixed fibroblast cells from normal and DM patients. The hybridized probe was detected with anti-digoxigenin alkaline phosphatase conjugate.

Fig. 2A is a photograph of cells (from 3132) which were hybridized with antisense CAG-30 probe (SEQ ID NO: 14). mRNA signal can be seen in the cytoplasm perinuclearly.

Fig. 2B is the same as Fig. 2A, but the sense CTG-30 probe (SEQ ID NO: 15) was used as a control for the CAG-30 probe (SEQ ID NO: 14). No signal was detected.

Fig. 2C is the same as Fig. 2A, hybridized with antisense CAG-30 probe (SEQ ID NO: 14), but normal (Detroit 551) cells were used as an undiseased control. No signal was detected.

Fig. 3 is a cytogenetic preparation which was hybridized to fluorescein labelled antisense CAG-30 probe (SEQ ID NO: 14). After hybridization, images were digitized by a CCD camera. Cells (n = 152) from patient 3132 were counted for number of discrete foci per nucleus. Most of the nuclei contained 2-6 foci. Patient 3755 showed similar results.

Fig. 4A - 4C. Fixed tissue sections taken from DM patient and a normal individual were hybridized to the fluoresceinated (Fig. 4A and Fig. 4B) or digoxigenated antisense (CAG-30; SEQ ID NO: 14) probe (Fig. 4C). Digoxigenin labelled probe was detected with antidigoxigenin alkaline phosphatase conjugate. After hybridization, tissues were counterstained with DAPI and visualized either by fluorescence or bright field microscopy.

Fig. 4A is a photograph of tissue from DM patient which showed the diagnostic feature of the triple repeats: a number of discrete foci in the nucleus (arrowheads).

Fig. 4B is a control with normal tissue which did not show any signal in the nucleus.

Fig. 4C is a photograph showing that signal was also present in the cytoplasm of DM with a digoxigenin labelled CAG-30 (SEQ ID NO: 14).

Detailed Description of the Invention

This invention provides a simple, rapid and

reliable method for specifically detecting the expansion of a trinucleotide repeat by *in situ* hybridization. Although the oligonucleotide probe hybridizes to both expanded and non-expanded repeats, an expanded repeat accommodates a larger number of probes hybridized in tandem. Due to its larger number of hybridized probes, an expanded repeat yields a stronger signal than does a non-expanded repeat. This difference in signal strength is advantageously exploited in the present invention. The hybridization detection sensitivity is designed so that the stronger signal from an expanded trinucleotide repeat is clearly distinguishable from the much weaker signal from a non-expanded repeat. Preferably, the signal from a non-expanded repeat falls below the threshold of detection, so that there is no need to compare two detectable signals.

The operation of the method of this invention is well-illustrated by its application to detect the trinucleotide repeat expansion found in the Mt-PK gene. In cells from human patients with myotonic dystrophy, the Mt-PK gene contains an expanded trinucleotide repeat, i.e., typically 50-2,000 repeats. Since normal Mt-PK genes (and other genetic loci) contain non-expanded trinucleotide repeats, i.e., 5-27 repeats, a repeat-specific probe can hybridize to normal genes (and mRNAs), as well as to those containing an expanded repeat. The specificity of the method of this invention derives as a matter of signal strength and detection threshold. Preferably, the invention is practiced with a fluorescently-labeled oligonucleotide probe which includes about 10 trinucleotide repeats (i.e., about 30 nucleotides). A normal (non-expanded) repeat therefore hybridizes to one to three probe molecules in tandem, at most. The resulting fluorescence signal falls below the

threshold of detection. An expanded repeat hybridizes to as many as 200 probe molecules in tandem, yielding a concomitantly greater fluorescence signal, which is well above the detection threshold. Detection specificity for the expanded repeat thereby results. In contrast, a probe hybridizing to Mt-PK regions other than the trinucleotide repeat would yield essentially the same signal strength, regardless of the presence or absence of an expanded trinucleotide repeat.

While the foregoing illustration involves the trinucleotide repeat expansion found in myotonic dystrophy, it will be appreciated that the principles applied in the that example may be applied to detect similarly any other trinucleotide repeat expansion. Accordingly, the invention has broad applicability.

Preferably, the oligonucleotide probe used is approximately 15 to 45 nucleotides in length. More preferably, the probe is about 30 nucleotides in length. Hybridization of the probe to the expanded trinucleotide repeat sequences is detected by a direct or indirect label on the probe, preferably by the attachment of a fluorescent moiety that can be detected by fluorescence microscopy. Hybridization of the probe to a single non-expanded trinucleotide repeat yields a signal substantially weaker than the signal from a single expanded trinucleotide repeat.

In one embodiment of the invention, an antisense probe is used, and there is no DNA denaturation step included in the *in situ* hybridization procedure. In that situation, the probe hybridizes only to trinucleotide repeat-containing mRNA molecules. The threshold of detection is such that the large number of probes hybridized in tandem at a repeat expansion (e.g., 5-200) yields a detectable signal, while the relatively small number of probes (e.g., 2-5) hybridized to transcripts of

a normal allele containing a non-expanded repeat yields a weak or non-detectable signal.

In another embodiment of the invention, a sense probe is used, and a DNA denaturation step is included in the *in situ* hybridization procedure. In that situation, the probe hybridizes to trinucleotide repeat-containing genomic DNA. The threshold of detection is such that the large number of probes hybridized in tandem at a repeat expansion (e.g., 5-200) yields a detectable signal, while the relatively small number of probes hybridized at a non-expanded repeat (e.g., 2-5) yields a weak or non-detectable signal.

In a preferred embodiment of this invention, used for DM diagnosis, an antisense probe contains CAG repeats. In another preferred embodiment, used for DM diagnosis, a sense probe contains CTG repeats. Preferably the number of repeats is between six and fifteen. More preferably the number of such repeats is ten. Use of probes within the preferred size range causes detection of the hybridized probe to indicate the presence of expanded CTG or CAG trinucleotide repeats, which in turn indicates a genetic disease. Positive *in situ* hybridization results generally will appear as 2 or more detectable foci in a cell, and this will be indicative of an increased probability of a genetic disease such as myotonic dystrophy.

The preferred threshold of detection is such that an expanded number of repeats will be detectable in the cells, but normal numbers of repeats in the genome will be below the threshold of detection.

Unexpectedly, the present *in situ* hybridization method, using probes specific for CTG (or CAG) trinucleotide repeats, was found to be highly specific for detection of the expanded trinucleotide repeats associated with myotonic dystrophy. The method did not

result in detectable non-specific hybridization throughout the genome, as would have been expected by its high GC content, and as has been seen with other methods (e.g., Southern blots). The present detection methods are particularly well-suited for diagnosing myotonic dystrophy, because the trinucleotide repeat expansions associated with myotonic dystrophy are larger (i.e., contain greater numbers of repeats) than the expansions associated with some other genetic diseases. For example, repeat expansion in Huntington's disease patients was not detected with the conditions used for the successful DM assay, presumably because the degree of expansion in Huntington's disease (20-40 repeats) was below the limit of sensitivity of the assay. As described below, however, sensitivity of the assay may be adjusted to allow detection of smaller repeat expansions.

The methods of the invention do not require that a patient is known or even suspected to have expanded trinucleotide repeats. The patient tested may be asymptomatic. The methods of the invention are suitable for clinical screening, e.g., in the general population or for prenatal diagnosis. For example, white blood cells, epithelial cells, fetal cells from amniotic fluid or chorionic villus sampling, and myocytes or fibroblasts from biopsy samples may be used.

If the sample used in this invention is a blood sample, the white blood cells need not necessarily be isolated, because there will likely be a sufficient number of nucleated cells in a standard blood smear to provide assay results. Alternatively, a sample of epithelial cells, e.g., from a cheek scraping or from amniotic fluid, may be used in the practice of this invention. Any type of nucleated cell may be used, because in embodiments involving a DNA denaturation step, it is not necessary that the cells of the sample

transcribe RNA from the expanded repeat-containing gene, in order for the trinucleotide repeat expansion to be specifically detected. Analysis of myocytes from a muscle biopsy sample may be used, in a dual-probe embodiment of this invention to obtain information on transcription of a myotonic dystrophy gene and subcellular localization of the DM transcripts.

In one embodiment of this invention, two differently-labelled oligonucleotide probes are used. An oligonucleotide probe complementary to the expanded repeat region of the coding strand of the DNA and a differently labelled probe complementary to the mRNA transcripts (and non-coding strand of the DNA) can be used together to detect the presence of an expanded trinucleotide repeat, or to determine whether the expansion-containing gene is being actively transcribed.

This invention includes embodiments capable of distinguishing between the coding strand of the gene and the transcripts from the gene. Such embodiments include labeling each probe with a different detectable label (e.g., colored fluorescing labels), denaturing the DNA in the nucleus, and incubating the probes *in situ* under conditions appropriate to allow them to hybridize with the nucleic acids of the cell. Detection of a sense probe hybridized to the gene only indicates that an expanded trinucleotide repeat in is the genome of the patient. Detection of an antisense probe hybridized to mRNA indicates that the gene has been or is being transcribed in the cell type used for the assay.

In preferred embodiments of this invention, used to detect the genetic defect associated with myotonic dystrophy, the oligonucleotide probes include CAG (antisense probes) or CTG (sense probes) trinucleotide repeats. For either type of DM diagnostic probe (i.e., antisense or sense), the number of trinucleotide repeats

in the probe is preferably between five and fifteen, and most preferably is ten.

If the labelled probes of the double-labelling hybridization co-localize, it indicates that transcription is ongoing in the cells. If the two probes are visible, but do not co-localize, it indicates that transcription has occurred and mRNAs have been transported, but that transcription is not currently ongoing. This method may be repeated periodically to monitor the progression of the disease, or to monitor a course of therapy being administered to the patient. If transcription decreases, it indicates that the applied therapy is effective at reducing production of repeat-containing transcripts. Such periodic testing is greatly facilitated by the ability to sample blood or epithelium as the source of cells for the assay, since these sources are inexpensive and accessible.

The greater the number of copies of probe hybridized to the target, the greater the signal from the label on the probe. As one of skill in the art will appreciate, the limits of detection of the signal depend on the type of label, number of labeled moieties per probe, and the means/conditions used to detect the label. For example, a radioactive label given a long exposure/detection time would be expected to yield a detectable signal for even a very low number of probe molecules. For discriminating between expanded repeats and non-expanded repeats, a shorter exposure/detection time would be preferred, so as to put the signal from low numbers of hybridized probes beneath the limit of detection. The number of labeled atoms or moieties per probe molecule and/or the detection conditions may be varied, depending on the genetic disease being diagnosed. For example, the number of expanded trinucleotide repeats in Huntington's disease is between around 42 - 86 (normal

individuals have between 16 and 36 repeats). Therefore, greater detection sensitivity would be required for detection of the Huntington's trinucleotide repeat expansion than would be required for detection of a myotonic dystrophy trinucleotide repeat expansion. Likewise, Fra-X genes exhibit between around 54 and 200 (normal between 6 - 54), and SBMA, between 40 and 52 (normal between 17 and 26). It is within ordinary skill in the art to appropriately vary labeling and detection conditions to allow detection of trinucleotide repeat expansions, while keeping the signal level for non-expanded repeats near or below the threshold of detection.

It will be appreciated by those of skill in the art that the 5′ and 3′ terminal sequences of the probes need not be a full iteration of a particular trinucleotide repeat, and that the terminal sequences of the probes may contain a small amount of non-repeat-specific DNA. It will also be appreciated that a particular trinucleotide repeat may be described or represented by any one of three triplets, depending on the reading frame selected.

The antisense repeat-specific probes, as well as the sense gene-specific probes, may be cloned and isolated from a recombinant vector as a restriction fragment or, preferably, produced by chemical synthesis using synthetic methods known in the art. The probes will be preferably about 15-75 nucleotides in length, more preferably 30-50 nucleotides in length.

We have discovered that *in situ* hybridization of antisense repeat-specific oligonucleotide probes to nucleic acids in cells from patients with expanded trinucleotide repeats results in at least one signal focus per cell generated by the detectable label associated with the probe. Cells from normal patients

contain no such signal foci after *in situ* hybridization to the antisense repeat-specific probes. The methods of this invention may be used to detect accumulations of transcripts (in the examples herein, DM-specific transcripts) containing the expanded repeats (e.g., CTG repeats). As used herein, the term DM-specific is intended to indicate transcripts which are present in DM cells but not in normal cells. The functions, if any, of the RNAs detected in the multiple foci are not known at the present time, nor is it known whether or not these transcripts are directly involved in the pathology or etiology of DM. However, as positive signals are detected only in cells from patients with expanded repeats, the instant methods are considered expanded repeat-specific and are therefore useful for detecting a degree of trinucleotide repeat expansion associated with disease. The methods are therefore useful for predicting predisposition to e.g., DM, or diagnosing such a disease prior to or after the onset of symptoms.

The size of the repeat expansion associated with a genetic disease to be detected is a factor in the practice of the invention. Trinucleotide repeat expansions of fewer than about 40 - 50 repeats (e.g., Huntington's disease) require relatively high sensitivity for probe detection. The appropriate adjustment of probe detection sensitivity is within ordinary skill in the art. The signal intensity is proportional to the number of repeats, assuming transcription and transport rates remain roughly equivalent.

This invention allows detection of trinucleotide repeat expansions in transcribed coding mRNA, non-coding mRNA regions, or in genomic DNA. The invention allows a qualitative measure of the severity of the disease (the brighter the foci, the more likely that there is a higher number of repeats, which is correlated with the severity

of these diseases and often indicates an earlier age of symptomatic onset).

Sample Preparation

The sample of cells containing nucleic acids for *in situ* hybridization to the gene- and/or repeat-specific probes may be cells in a tissue section, individual cells in suspension, or plated cells. Although fixation is not necessary, prior to hybridization with the repeat-specific probes, the cells in the sample can be fixed briefly with a fixative which maintains the morphological integrity of the cell but does not cross-link or precipitate cellular proteins so extensively that penetration of probes and other reagents is prevented. Harsh pretreatment with protease is therefore generally avoided after fixation. Either cross-linking or precipitating fixatives, as are known in the art, may be used in the practice of the invention. Examples include 4% paraformaldehyde, 2% glutaraldehyde, ethanol/acetic acid fixatives, Carnoy's fixative (acetic acid, ethanol, chloroform), 1% osmium tetroxide, Bouin's fixative (1.21% picric acid, 11% formaldehyde, 5.6% acetic acid), Zenker's fixative (5.0% mercuric chloride, 2.5% potassium dichlorate, 5.0% acetic acid, 1.0% sodium sulfate), and acetic acid/methanol fixatives. The preferred fixative for use in the invention is 1-4% paraformaldehyde, which is preferably used to treat the cells or tissues for about 1 min to 1 hr. It has been found that this brief fixing with paraformaldehyde often allows penetration of probes and other reagents into the cells without the need for the destructive protease treatment.

Following fixation, the nucleic acids in the cells are hybridized *in situ* to the repeat-specific probes and detected by means of a detectable label associated with the probe. That is, the fixed cells or tissues are

exposed to the labeled probe under reaction conditions appropriate for specific *in situ* hybridization of the probe to the trinucleotide repeat sequences, if present, or gene. Several appropriate *in situ* hybridization methods are known in the art. Preferred fixation and *in situ* hybridization methods are described by (Lawrence *et al.*, 1988, Cell, **52**:51-61; Lawrence *et al.* 1989, Cell, **57**:493-502) and in U.S. Patent No. 4,888,278. These are hereby incorporated by reference.

If a tissue section containing cells for *in situ* hybridization is embedded in paraffin, the paraffin is removed prior to fixation by treatment with xylene as is known in the art. These tissue sections may have been previously fixed for other purposes (e.g., in the pathology laboratory) without regard to penetration of hybridization reagents, i.e., they may have been fixed for substantially longer than 1 hr. Often, penetration of reagents, including probes, into the cells of the tissue section is still satisfactory in spite of extensive fixation, possibly due to the thinness of the section. However, in some cases penetration of the cells by hybridization reagents may be prevented or significantly reduced by prior extensive fixation. If so, it is preferred that the tissue section be treated with protease or heat prior to *in situ* hybridization to improve penetration of reagents into the cells of the tissue sample. Frozen tissue sections which have not been previously fixed may be fixed as described above and do not require prior protease or heat treatment.

Detectable Labels for Probes

Hybridization of the repeat-specific probes to the expanded trinucleotide repeat sequence and 5′ probes may be detected by means of a detectable label associated with the probe. The detectable label is a moiety which

generates, or can be made to generate, a signal which can be detected, for example visually or by instrumentation. The label may be incorporated into the probe e.g., by nick translation or random priming if the probe is cloned in a recombinant vector. Alternatively, the label may be covalently linked to the probe by end-labelling, incorporated during chemical synthesis of the probe by introduction of label-derivatized nucleotides, or coupled to the probe by covalent attachment to reactive groups on the nucleotides (e.g., aliphatic amino groups). Many directly and indirectly detectable labels are known in the art for use with oligonucleotide probes. Directly detectable labels include those labels which do not require further reaction to be made detectable, e.g., radioisotopes (^{32}P, ^{33}P, ^{125}I or ^{35}S), fluorescent moieties and dyes. Fluorescent labels such as fluorescein isothiocyanate (FITC), Texas red, or Cy3 are preferred as directly detectable labels. Indirectly detectable labels include those labels which must be reacted with additional reagents to be made detectable, e.g., enzymes capable of producing a colored reaction product, biotin, avidin, digoxigenin, antigens, haptens or fluorochromes. The signal from enzyme labels is generally developed by reacting the enzyme with its substrate and any additional reagents required to generate a colored, insoluble, enzymatic reaction product. For example, alkaline phosphatase (AP) is stable and has been used extensively for labeling in tissues and cells. The presence of AP may be detected by reaction with a substrate, the preferred substrates being Vector Red/Vector Blue (Vector Labs, CA), 5-bromo-4-chloro-3-indolyl phosphate (BCIP)/nitro blue tetrazolium (NBT) (Sigma Chemical Company, St. Louis, MO) or Nuclear Fast Red (Sigma Chemical Company). Vector Red has the added advantage of fluorescence, allowing visualization of a positive signal

either by conventional light microscopy or by fluorescence microscopy. Methods for developing the colored reaction product of AP with these substrates are known in the art. Biotin labels may be detected by binding to labeled (enzyme or fluorochrome) avidin or labeled anti-biotin antibodies. Digoxigenin and hapten labels are usually detected by specific binding to a labeled (enzyme or fluorochrome) anti-digoxigenin or anti-hapten antibody. Enzymes are preferred for use as indirectly detectable labels in the present invention.

The label of the hybridized repeat-specific probe is detected as an indication of the presence of expanded trinucleotide repeat sequences in the cells. This may require the addition of reagents to the cells to develop the signal of an indirectly detectable label such as AP, biotin or digoxigenin. Microscopic analysis of the cells is preferred when the detectable label is an enzyme or fluorescent moiety. Microscopic analysis may be either by visual observation of the cells or tissues (fluorescence or light microscopy), or automated image analysis using instruments such as DISCOVERY (Becton Dickinson Cellular Imaging Systems, Leiden, Holland) to evaluate the number and signal intensity of the spots, which indicate the presence of the repeats and thus positive cells. In addition to fluorescence microscopy, use of a directly detectable fluorescent label allows fluorescence analysis of cells in suspension by flow cytometry (e.g., FACSCAN, Becton Dickinson Immunocytometry Systems, San Jose, CA). A shift in peak fluorescence to the right on a plot of cell number vs. fluorescence intensity is indicative of an increased number of cells containing expanded trinucleotide repeat sequences. Conversely, a shift in peak fluorescence to the left on the plot is indicative of a reduced number of cells containing expanded trinucleotide repeat sequences.

This technique could be useful diagnostically, to detect differences in gene "dose" (e.g., relative number of trinucleotide repeats) between individuals.

This invention may be modified to detect lower copy numbers in order to diagnose diseases such as Huntington's which are characterized by fewer expanded trinucleotide repeats. By changing the probe's label or the conditions under which it is detected, one can increase the sensitivity of the assays described herein. Adjustment of the detection sensitivity may be accomplished by means including, but not limited to, the following: making the label a high energy isotope; adding a greater number of labels/probe (e.g., using more fluorochromes/probe or more than one labelled dNTP in the synthesis of the probe); lengthening exposure time of the hybridized cell to the detection medium (e.g., x-ray film, phototransduction elements); or any other methods known in the art to detect low copy numbers of a labelled probe.

Experimental Information

Preparation of Oligonucleotide Probes

All oligonucleotide probes were synthesized on a DNA synthesizer (Applied Biosystems, Foster City, CA). Repeat specific oligonucleotides were synthesized with two amino modified dT (Glen Research, Sterling, VA) at either end. In addition, thirteen oligonucleotide probes (40-45 bases long) to the 5′ end of the transcript (DM1-DM13; SEQ ID NOs: 1-13) were synthesized with five amino modified dT about 10 bases apart (Kislauskis *et al*, 1993, J. Cell Biol., **123**:165-72). Sequences were obtained from Genbank (accession L00727; Caskey *et al*, 1992, Science, **256**:784-88), and unpublished results of David Housman and Mila McCurrach. After deprotection, probes were purified by gel electrophoresis. The purified probes were labelled with fluorescein isothiocyanate or the red dye

CY3 (Biological Detection Systems, Pittsburgh, PA) or Texas Red (Molecular Probes) by using 0.1 M $NaHCO_3/Na_2CO_3$, pH 9.0, overnight at room temperature in the dark (Aggrawal *et al*, 1986, Nucleic Acids Res., **14**:6227-45; hereby incorporated by reference). Reaction products were passed twice through Sephadex G-50 (using a 25 ml disposable pipette). Fractions were combined, lyophilized and further purified on a 10% polyacrylamide native gel. Purified probe was then extracted from the gel by soaking overnight in 1M triethyl ammonium bicarbonate at 37°C. The supernatant was passed through a C18 Sep-Pak cartridge and the DNA eluted in 30% acetonitrile/10mM triethyl ammonium bicarbonate (TEAB). The probes were 3'-end labelled with digoxigenin using dig-11-dUTP and terminal deoxy transferase (Boehringer Mannheim Biologicals, Indianapolis, IN).

Oligonucleotide sequences used in the examples described are SEQ ID NOs 1-15; oligonucleotides for implementing the invention with other genes and products can be readily selected by one of ordinary skill in the art.

In Situ Hybridization

Primary skin fibroblast cells derived from two DM patients (3132 and 3755) and normal human diploid fibroblast cells were grown in dishes containing gelatin coated coverslips at 10^6 cells per 100 mm dish. Cells on the coverslip were washed with HBSS and fixed for 15 minutes at room temperature in 4% paraformaldehyde in PBS (2.7 mM KCl, 1.5 mM KH_2PO_4, 137 mM NaCl, 8 mM Na_2HPO_4) and 5 mM $MgCl_2$. After fixation, cells were washed and stored in 70% ethanol at 4°C. Cells on coverslips were hydrated in PBS and 5 mM $MgCl_2$ for 10 minutes and treated with 40% formamide and 2X SSC for 10 minutes at room temperature. Cells were then hybridized for two hours at 37°C with

fluorochrome labeled oligonucleotide probe (10 ng) in 20 µl volume containing 40% formamide, 2 X SSC, 0.2% BSA, 10% dextran sulfate, 2 mM vanadyl adenosine complex and 1 mg/ml each of E. coli tRNA and salmon sperm DNA. The 5'-end antisense probes (DM 1-13; SEQ IDs 1-13) were used as a mixture of 13 oligonucleotides totalling 40 ng. After hybridization and washing, coverslips were mounted on slides using phenylene diamine (antibleach agent) in 90% glycerol with PBS and the DNA dye DAPI (4,6 Diamidino-2-phenylindole).

Frozen tissue was sectioned at 5 microns and kept frozen until fixation before *in situ* hybridization. Normal control tissue was obtained in the same manner.

Preparation of Nuclei

Cells from patients were incubated in culture with 0.015 µg/ml colcemid for 45 minutes. After incubation, cells were trypsinized with 1% trypsin-EDTA in HBSS. Fresh medium was added to stop the trypsin reaction. The cells were centrifuged and the cell pellet was resuspended in 5 ml of 0.075 M KCl, incubated 17 minutes at 37°C and again centrifuged. Freshly prepared 3:1 methanol:acetic acid (10 ml) was added drop by drop to the cell pellet, with mixing at 25°C for 10 minutes and centrifuged. About 10 ml of methanol:acetic (3:1) was added to the cell pellet and incubated for 10 minutes at 25°C. The cells were centrifuged, the supernatant removed, and the cell pellet was resuspended in 1 ml of methanol:acetic acid. Cells were dropped onto ethanol washed slides from a distance of 2 feet and dried in air overnight. The slides were incubated at 65°C for 10 minutes and stored at -80°C (Johnson *et al*, 1991, Functional Organization of the Nucleus, A Laboratory Guide, Academic Press, Inc., NY, **35**:73-91; hereby incorporated by reference). The slides were then

hybridized as described with oligonucleotide probe. This preparation also resulted in chromosome spreads.

Specific Detection of DM mRNA by Antisense Probes

In situ hybridization to repeat-specific oligonucleotide probes was used to identify trinucleotide repeat transcripts within intact cells. Antisense probes (CAG-30, 30 nucleotides in length, having 10 repeats of the sequence CAG, AGC, or GCA when read in the 5′ to 3′ direction; SEQ ID NO: 14) and sense probes (CTG-30, 30 nucleotides in length having 10 repeats of the sequence CTG, TGC, or GCT when read in the 5′ to 3′ direction; SEQ ID NO: 15) probes to the trinucleotide repeat were synthesized with amino-modified dT on an Applied Biosystems DNA Synthesizer 394. CAG-30 (SEQ ID NO: 14) had two amino-modified dT residues at each end and CTG-30 (SEQ ID NO: 15) had three amino-modified dT residues about 15 nucleotides apart. The purified CAG-30 probe (SEQ ID NO: 14) was labeled with fluorescein isothiocyanate (Molecular Probes, CA) and the CTG-30 probe (SEQ ID NO: 15) was labeled with Cy3 (BDL, PA).

Fibroblast cells from two myotonic dystrophy patients (patient numbers 3132 and 3755) and normal human diploid fibroblasts (Detroit 551 cells) were grown, fixed, and hybridized to the sense and antisense repeat probes as described above (under "*in situ* hybridization" methods). After hybridization and washing, coverslips were mounted on slides using phenylene diamine (antibleach agent) in 90% glycerol and PBS.

For cytogenetic preparations, cells from patients 3132 and 3755 were incubated in culture, prepared, and hybridized as described above (under "preparation of nuclei" methods).

Fig. 1A and Fig. 1B show the epifluorescent signal obtained after *in situ* hybridization of the antisense

probe CAG-30 (SEQ ID NO:14) in the cells of a muscular dystrophy patient. There are a number of discrete foci in the nucleus. In contrast, no signal was seen after *in situ* hybridization with the sense probe CTG-30 (SEQ ID NO: 15; Fig. 1C) or antisense probe in HDF cells. The signal is therefore believed to be due to hybridization of the probe with the CTG repeats present in the mRNA transcript. The number of discrete foci in individual cells was generally between 1 and 13, scattered throughout the nucleus in an apparently random pattern. This was unexpected, as two foci would be predicted in G1 cells and four foci in G2 cells. The additional foci may be due to rate limiting steps in the processing or transport of the transcripts at sites of accumulation.

The nuclei prepared for cytogenetic studies also showed signals as several discrete foci with antisense probe CAG-30 (SEQ ID NO: 14; Fig. 1D), but did not produce a detectable signal with the sense probe CTG-30 (SEQ ID NO: 15; Fig. 1E), presumably because the DNA was not adequately denatured. No signal was detected when either the sense or antisense probe was hybridized to normal human diploid fibroblast cells (Fig. 1F), confirming that the trinucleotide repeat expansion is not a general phenomenon.

Using the probes described above, we have seen the distribution of poly A^+ RNA containing unstable expanded CTG repeat sequences in the cytoplasm of DM fibroblast cells. Probes CAG-30 (SEQ ID NO: 14) and CTG-30 (SEQ ID NO: 15) were labelled with digoxigenin using terminal transferase and dig-11-dUTP (BMB). The labelled probes were hybridized to the fibroblast cells and detected with anti-digoxigenin-alkaline phosphatase conjugate. The mRNA containing expanded CTG repeats were present in the DM fibroblasts and distributed perinuclearly in the cytoplasm (Fig. 2A). They remained attached to the

cellular structure after non-ionic detergent extraction. Signal was absent with CTG-30 (SEQ ID NO: 15) in DM fibroblasts (Fig. 2B) and CAG-30 (SEQ ID NO: 14) in Detroit 551 cells (Fig. 2C).

The cells of one myotonic dystrophy patient were counted as a function of the number of discrete foci observed in the nucleus. The histogram is shown in Fig. 3. Most nuclei contained less than 6 discrete foci, but some were found with as many as 13. This may indicate that the sample contains cells expressing different lengths of expanded repeat sequence, or that the foci represent pools of nascent and processed RNA. The histogram of the second myotonic dystrophy patient was similar.

Fixed tissue sections (muscle) from a DM patient and a normal control were hybridized with fluoresceinated CAG-30 probe (SEQ ID NO: 14) and counterstained with DAPI. The nuclei of the DM tissues contained a number of discrete foci (Fig. 4A). None were detected in the normal tissue (Fig. 4B). The signal of mRNA containing expanded CTG repeats in the sarcoplasm was predominately at the periphery of the cells in the DM sample (Fig. 4C).

Localization of Normal Mt-PK Transcripts and DM Transcripts

Thirteen oligonucleotide probes (DM1-13; SEQ ID NOs: 1-13) from the 5′-end seven exons of Mt-PK RNA (there are 14 exons total) were labeled with digoxigenin and used as a mixture for in situ hybridization. Probes from the 5′-end of Mt-PK RNA are expected to detect transcripts from both the normal and DM allele, and can not distinguish between transcripts of the normal and expanded alleles. After in situ hybridization, the signal of Mt-PK mRNA was present perinuclearly within the

cytoplasm of normal as well as DM fibroblast cells. No significant difference in the cytoplasmic location of the Mt-PK mRNA was observed in cells between DM patients and normal individuals.

In order to distinguish between the normal allele and the one containing an expanded repeat, we constructed a probe complementary to the CTG repeat. Since the transcript of the allele containing the expansion should have an approximately 400-fold longer trinucleotide repeat target sequence than the transcript from the normal allele, we expected that this probe would give a DM-specific signal. The sites of subcellular localization of transcripts with the expansion could thus be distinguished. Digoxigenin-labeled oligonucleotide probes to the repeat sequence and the sense control (CAG-30 SEQ ID NO:14 and CTG-30 SEQ ID NO:15) were hybridized to the fibroblast cells from normal and DM patients, and the hybridized probe was detected with antidigoxigenin alkaline phosphatase conjugate. The CTG repeat sequence was detectable in the DM fibroblasts and distributed perinuclearly within the cytoplasm. In contrast, detectable signal from the repeat was absent in the cytoplasm of normal fibroblasts. We then confirmed this observation in diseased muscle by investigating the distribution of the CTG expansion in the mRNA from a muscle biopsy from a DM patient and a normal control. The signal of CTG repeat in the sarcoplasm was predominately at the periphery of DM myofibers and was not found in the normal tissue. These results confirmed that the CAG-30 (SEQ ID NO:14) probe detected the presence of cytoplasmic mRNA of the expanded repeat allele of the Mt-PK gene. In the affected fibroblasts, the mRNA with the expansion was localized in the cytoplasm apparently identical to the localization of the mRNA from the normal allele, as determined by using the

5′ probes on normal or affected fibroblasts. Both the normal and the expanded mRNA remained after non-ionic detergent extraction, indicating that both mRNAs were attached to the cytoskeleton. Since there was no detectable abnormality in the spatial distribution or cytoskeletal association of the Mt-PK mRNA observed in affected cells, evidence does not support the hypothesis that DM pathology is caused by mislocalization of the Mt-PK mRNA in the cytoplasm.

In the course of these studies we observed a striking distribution of the in situ hybridization signal in the nuclei of both fibroblasts and muscle cells of DM patients: the Mt-PK transcripts were present as foci of nuclear aggregations. The antisense probe hybridized to a number of foci in the nuclei of intact fibroblasts from DM patients. Foci of hybridization were absent when the sense probe was hybridized to the nuclei of the myotonic dystrophy patient cells or when the probe was hybridized to the nuclei from normal human fibroblasts. Individual cells showed strong signal represented by many discrete foci which were scattered throughout the nucleus in apparently random positions. This punctuate hybridization pattern suggested that the subnuclear localization of the Mt-PK transcripts may have a functional significance.

In order to visualize these foci more clearly, nuclei were isolated to eliminate cytoplasmic contamination, and also showed the signal with antisense probe as a number of discrete foci. No signal was detected with the sense probe in these nuclei. Skin fibroblasts cells from one myotonic dystrophy patient (3132) were analyzed for the number of foci in each nucleus. Nuclei contained a mean of 5 foci, but some were found with up to 13. Cells from another myotonic dystrophy patient (3755) were very similar. In order to

determine that these foci were characteristic of the disease, and were not an artifact of cell culture, tissues with the primary lesion were investigated. Histological preparations of muscle biopsies from DM and normal patients were hybridized to fluoresceinated CAG-30 probe and counterstained with DAPI. It was found that the nuclei of DM tissue contained 1-3 intense foci which were not detected in normal tissues.

Simultaneous Detection of Different Probes on DM mRNA

To unambiguously demonstrate that the foci did not result from repeated sequences unrelated to the myotonic dystrophy allele, and that the repeat probe was detecting Mt-PK transcripts specifically, we co-hybridized with the mixture of the CAG-30 probe (green) and the thirteen oligonucleotide probes from the 5′ end of Mt-PK transcripts (red) to nuclei from myotonic dystrophy patient samples and analyzed the distribution of each probe simultaneously. The 5′-end probes (DM1-13; SEQ ID NOS: 1-13) were labelled such that each probe contained five red fluorochromes; hence, a total of 65 fluorochromes would hybridize to the 5′-end of each transcript. The antisense probe (representing 10 repeats) contained two fluorescein molecules. Since there are approximately 2,000 repeats, the total number of fluorochromes conjugated to the probe which repetitively hybridized to the 3′-end would be as many as 400, generating a greater signal to noise ratio than the 5′-end of the transcript. This was observed: when the 5′-end (DM1-13) and CAG-30 were mixed in equimolar ratios and hybridized to preparations of DM fibroblast nuclei, the intensity of the foci revealed by the 5′ (red) probes contained less signal than the 3′ (green) probe when compared to their respective background levels. Reversal of the fluorochromes on the 5′-end and repeat probes gave

the same results. These results confirmed that the CAG-30 probe hybridized to transcripts arising from the DM allele. To evaluate the extent to which the 5'-end probes co-localized with the CAG repeat probe, we used digital imaging microscopy to provide an assessment of spatial congruence of the two labels. Optical sections at each wavelength were taken on a CCD camera and restored mathematically to remove fluorescent light not contributing to the specific section (Carrington *et al*, 1990, Non-Invasive Techniques in Cell Biology, Wiley-Liss Inc., NY, pp. 53-72). Two images from the same Z plane were superimposed using fiduciary markers. It was found that the green foci always co-localized exactly with red foci indicating that foci containing the expanded CTG repeat sequences were present only in the Mt-PK transcript. However, one, and sometimes two, red (5'-end) foci did not co-localize with any green (i.e. contained no repeat hybridization). These would be expected to represent the transcription sites of the normal allele. Therefore, the large number of foci obtained using the repeat probe were in excess of the number of transcription sites, and must represent released transcripts. Additionally, all of these supernumerary foci contained the repeat expansion, indicating that they have resulted from transcription of the affected, but not the normal allele.

Detection of DM Gene with Trinucleotide Repeat Sense Probe

In order to characterize the nascent transcripts from the released transcripts unequivocally, we hybridized differently labeled probes which were specific for either the DNA or RNA. The sense probe hybridized to only one, and occasionally two, of these foci when DNA in the interphase nucleus was denatured. The antisense

probe hybridized to the DNA and its transcripts. Co-localization of the DNA and RNA signals confirmed that only one of the foci contains nascent transcripts, whereas the other supernumerary foci contained post-transcriptional RNA. Actinomycin D treatment did not change the number of foci significantly (data not shown), further supporting the argument that almost all the foci are post-transcriptional accumulations.

In the nuclei of cells derived from normal individuals, the 5'-end probes (DM1-13) showed only one or two foci of signal (data not shown), consistent with previous observations that these are the sites of transcription of both alleles (Zhang *et al*, 1994, Nature **372**:809-812). These foci were considerably dimmer than the signal seen in the DM cells using the same probe. Therefore, it appeared that the Mt-PK transcript in normal cells, in contrast to the transcript from the affected gene, was efficiently processed and transported to the cytoplasm. In DM cells, the build-up of discrete foci in the nucleus may be the consequence of a rate limiting step in RNA processing such as splicing, polyadenylation or transport to the cytoplasm.

These results demonstrate that focal accumulations of post-transcriptional RNA are a characteristic of the expanded repeat sequences from the affected allele of a gene responsible for a disease such as myotonic dystrophy. No other examples exist which show RNA accumulated in foci subsequent to its transcription; the only foci identified to date are the sites of nascent chain transcription (Lawrence *et al., supra*; Shermoen and O'Farrel, 1991, Cell **67**:303-310). The fact that this observation occurs only in the nuclei of affected cells, and only with transcripts from the affected allele, suggests that these foci represent the primary events of the disease lesion. Furthermore, because the repeat

transcripts appear to "build up" in the nuclei, these results suggest that some aspects of nuclear RNA metabolism may be responsible for the foci we observe.

Other embodiments are within the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Singer, Robert H.
Taneja, Krishan L.

(ii) TITLE OF INVENTION: DETECTION OF TRINUCLEOTIDE REPEATS BY IN SITU HYBRIDIZATION

(iii) NUMBER OF SEQUENCES: 15

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: FISH & RICHARDSON P.C.
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: U.S.A.
(F) ZIP: 02110-2804

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30B

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 08 March 1995
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: 08/214,823
(B) FILING DATE: 17 March 1994

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Creason, Gary L.
(B) REGISTRATION NUMBER: 34,310
(C) REFERENCE/DOCKET NUMBER: 06353/010WO1

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (617) 542-5070
(B) TELEFAX: (617) 542-8906
(C) TELEX: 200154

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGCAGCCCC GTCCAGGCCC GGAGCCCGGC TGCAGGC 37

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGTCCCTGGC GTCCCCCGGG CCTCTCGCCA CTTCTCC 37

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GCCGAGGCCC CTCCCCTCTC CCCACCCCTT GGTCCAC 37

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CCCCTCCTCC AGGGCCCTCA GAACCCTCAG TGCTAGG 37

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGCATGGAGA ACTCAGCACA CTGCACCCCA AAAA 34

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GCGTCCCTCG CAGTCGGACC CCTTAAGCCC ACCACGAG 38

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCATGATCTC ATGGCAACAC CTGGCCCGCT GCTTCATCT 39

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCAGGTAGTC TCATCCGGAA GGCGAAGGCA GCTGCGG 37

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GCAGCAGGCC CCGCCCACGA ATACTCCAGA CCAGGA 36

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GACAATCCCG CCAGGAGAAG CGCGCCATCC GGCCGGAAC 39

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGGCCACAGC GGCCAGCAGG ATGTGTCGGG TTTGATGC 38

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCCACAGCCG CAGGATCCGG GGGACAGGAG TCTGGGGG 38

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGGAATCCG CGAGAAGGGC GCTGCCCAAG AACATTC 37

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG T 31

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTGCTGCTGC TGCTGCGCTG CTGCTGCTGC T 31

Claims

What is claimed is:

1. A method for detecting a trinucleotide repeat expansion by *in situ* hybridization, wherein the method comprises the steps of:

(a) providing a sample of nucleated cells;

(b) contacting said sample with a labelled trinucleotide repeat-specific oligonucleotide probe under conditions that allow said probe to hybridize with said trinucleotide repeat expansion in said sample; and

(c) detecting the hybridized probe by a means whose sensitivity distinguishes between a signal from probes hybridized to an expanded repeat and a signal from probes hybridized to a non-expanded repeat.

2. The method of claim 1 wherein said hybridized probe is labeled with a fluorescent moiety and detected by fluorescent microscopy.

3. The method of claim 1 wherein said hybridized probe is labeled with a fluorescent moiety and detected by image cytometry.

4. The method of claim 1 wherein said probe comprises a trinucleotide sequence selected from the group consisting of CAG, AGC, and GCA.

5. The method of claim 4 wherein said probe comprises between 5 and 15 trinucleotide repeats.

6. The method of claim 5 wherein said probe comprises 10 trinucleotide repeats.

7. The method of claim 4 wherein detection of said hybridized probe is indicative of an expanded trinucleotide repeat sequence.

8. The method of claim 4 wherein detection of said hybridized probe is indicative of myotonic dystrophy.

9. The method of claim 1 wherein said probe comprises a trinucleotide sequence selected from the group consisting of CTG, TGC, and GCT.

10. The method of claim 9 wherein detection of said hybridized probe is indicative of myotonic dystrophy.

11. The method of claim 1, further comprising a DNA denaturation step.

12. The method of claim 11, wherein said probe comprises CTG trinucleotide repeats.

13. The method of claim 11 wherein said cells are from an asymptomatic patient.

14. The method of claim 13 wherein said patient is *in utero*.

15. The method of claim 11 wherein said cells are white blood cells.

16. The method of claim 11 wherein said cells are epithelial cells.

17. An *in situ* hybridization method for detecting a gene containing an expanded trinucleotide repeat and its transcripts by differentially detecting the coding strand of the gene and the transcripts, said method comprising the steps of:

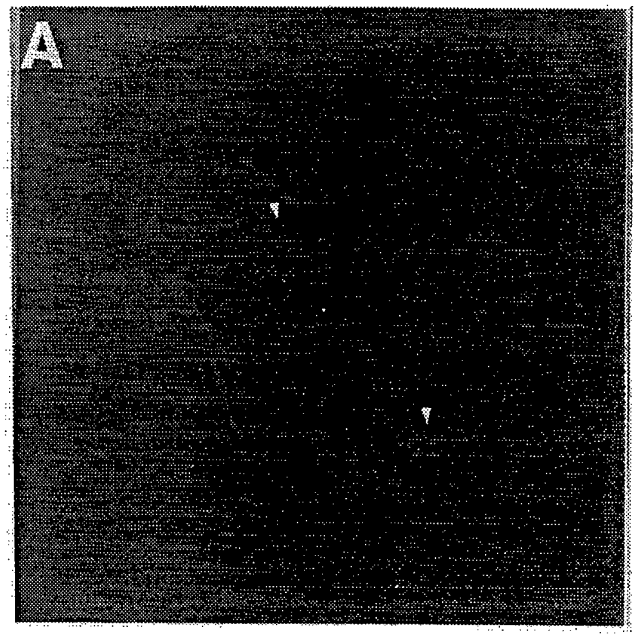
(a) providing a sample of nucleated cells;

(b) denaturing the DNA in said cells;

(c) hybridizing to said DNA a trinucleotide repeat-specific sense probe, said probe being labelled with a first detectable label, and

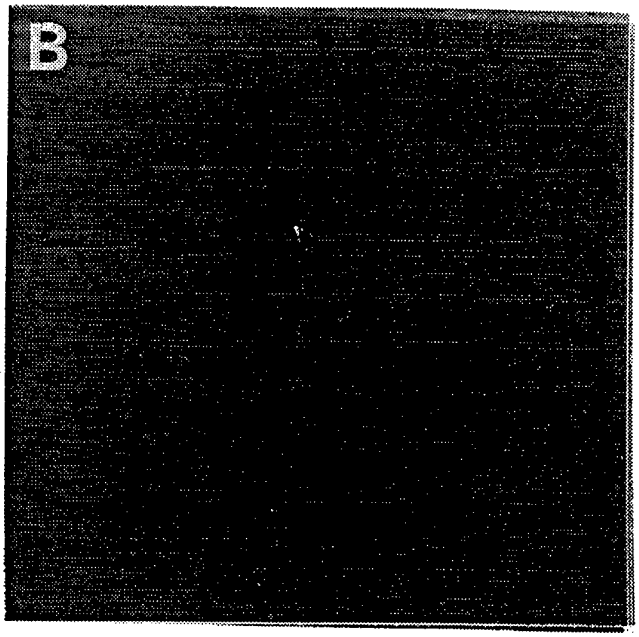
(d) hybridizing to said transcripts a trinucleotide repeat-specific antisense probe, said probe being labelled with a second detectable label.

18. The method of claim 17 wherein said transcripts comprise an expanded trinucleotide repeat consisting of CTG repeats.



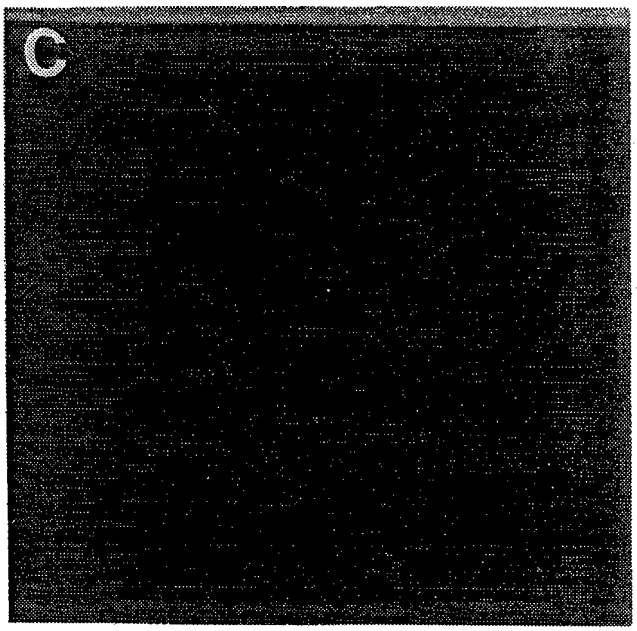
(SEQ ID NO:14)

FIG. 1A



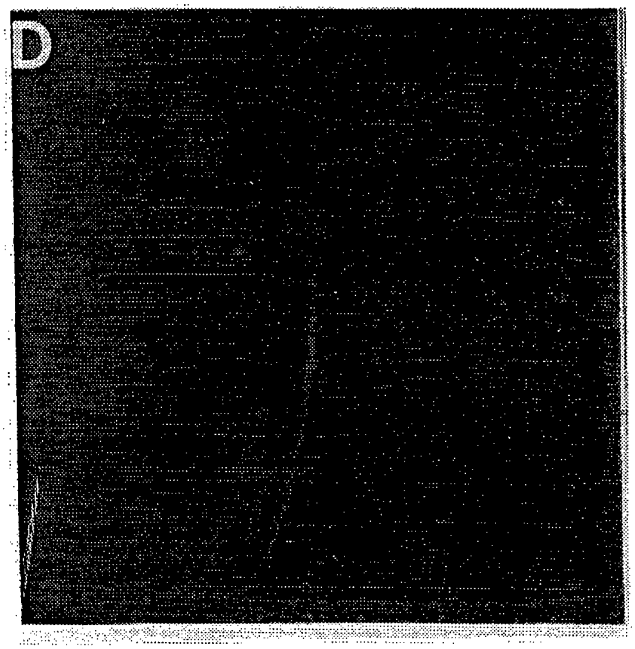
(SEQ ID NO:14)

FIG. 1B



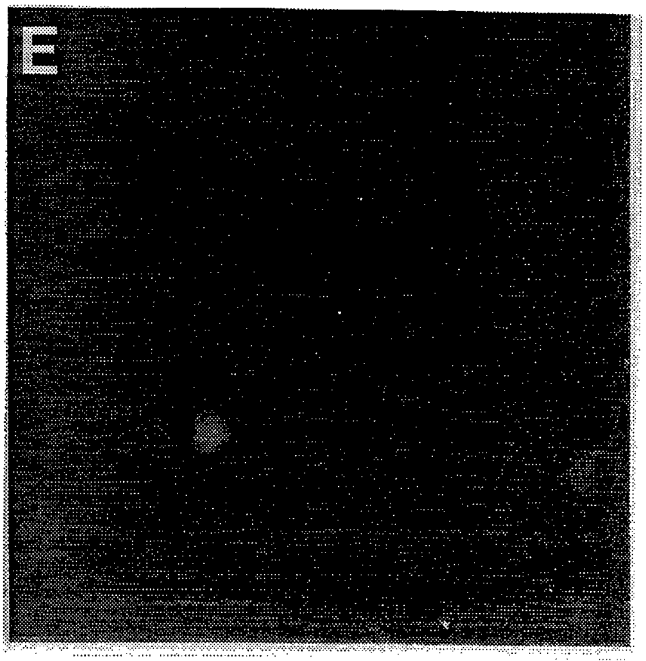
(SEQ ID NO:15)

FIG. 1C



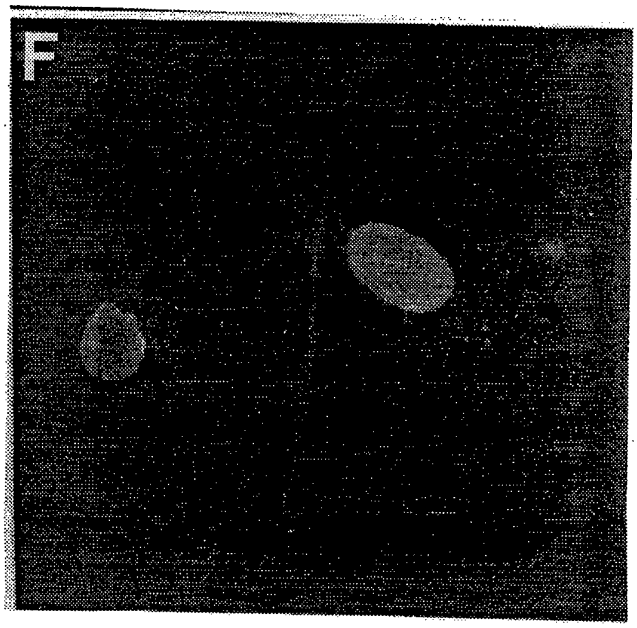
(SEQ ID NO:14)

FIG. 1D



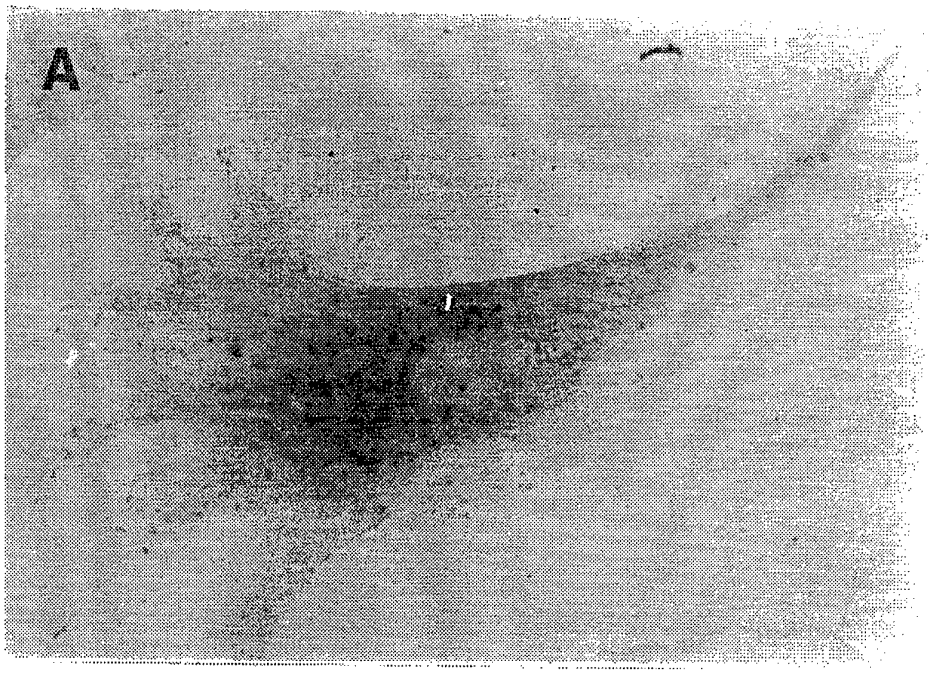
(SEQ ID NO:15)

FIG. 1E



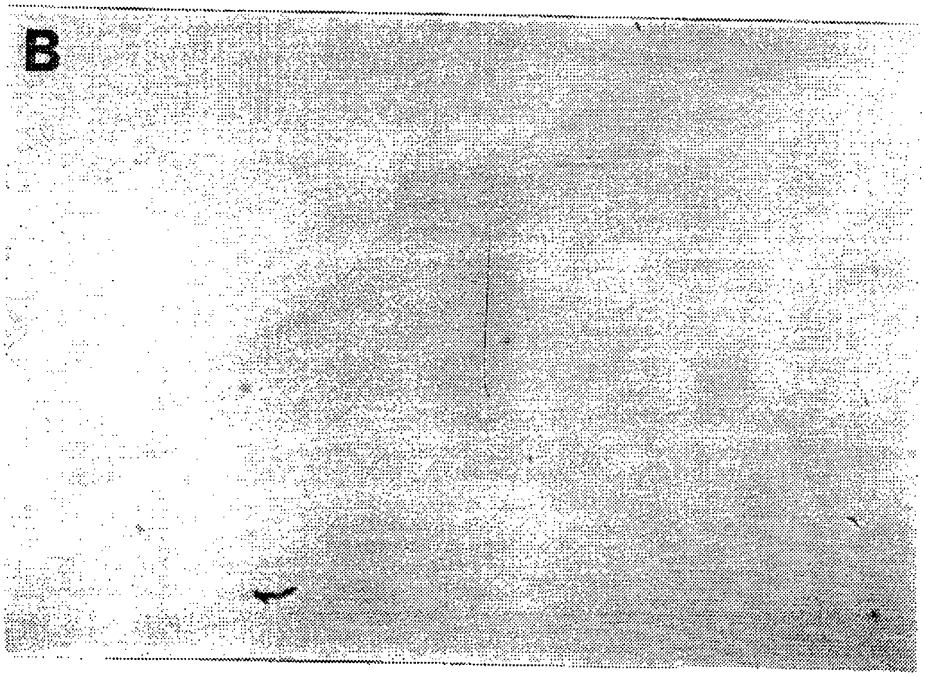
(SEQ ID NO:14)

FIG. 1F



(SEQ ID NO:14)

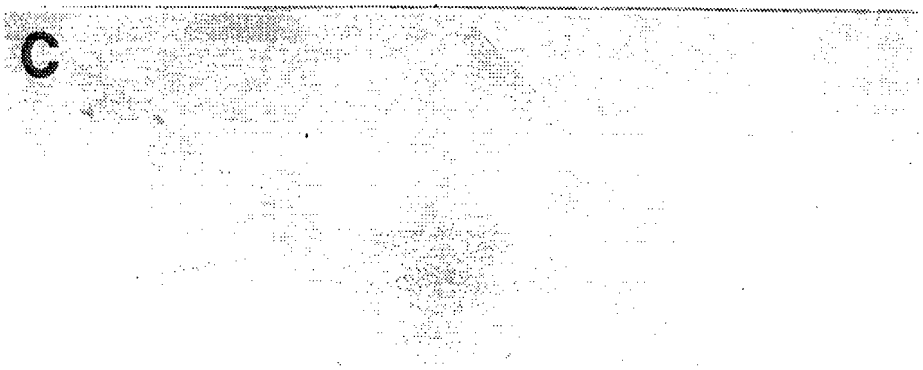
FIG. 2A



(SEQ ID NO:14)

(SEQ ID NO:15)

FIG. 2B



(SEQ ID NO:14)

FIG. 2C

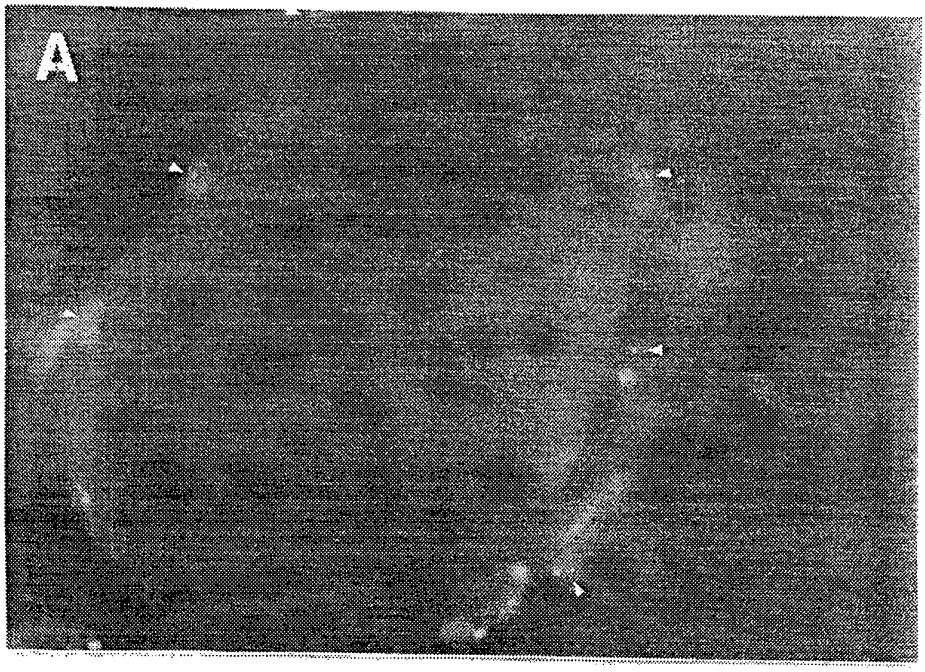


FIG. 4A

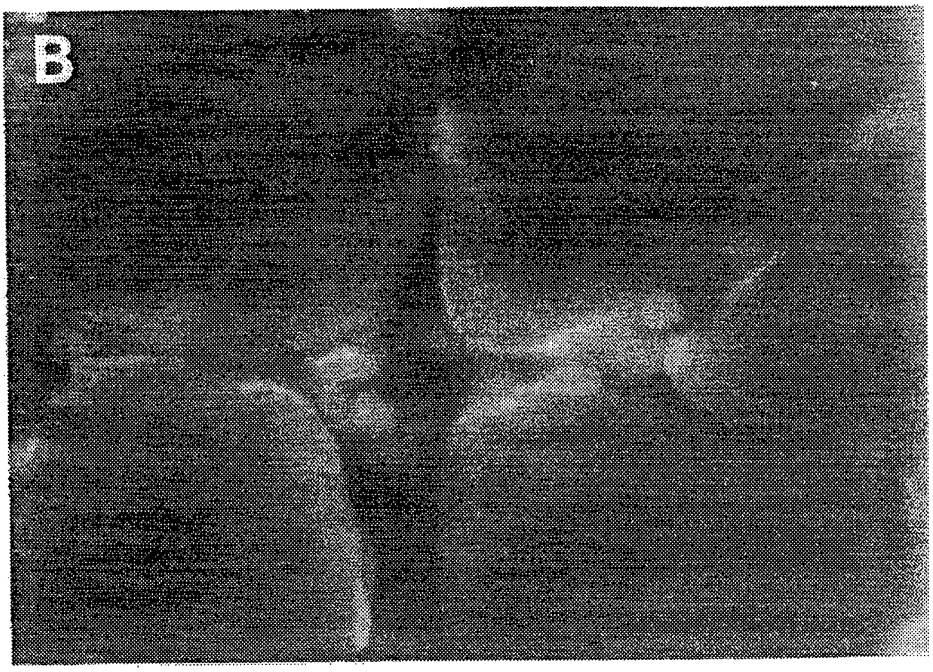
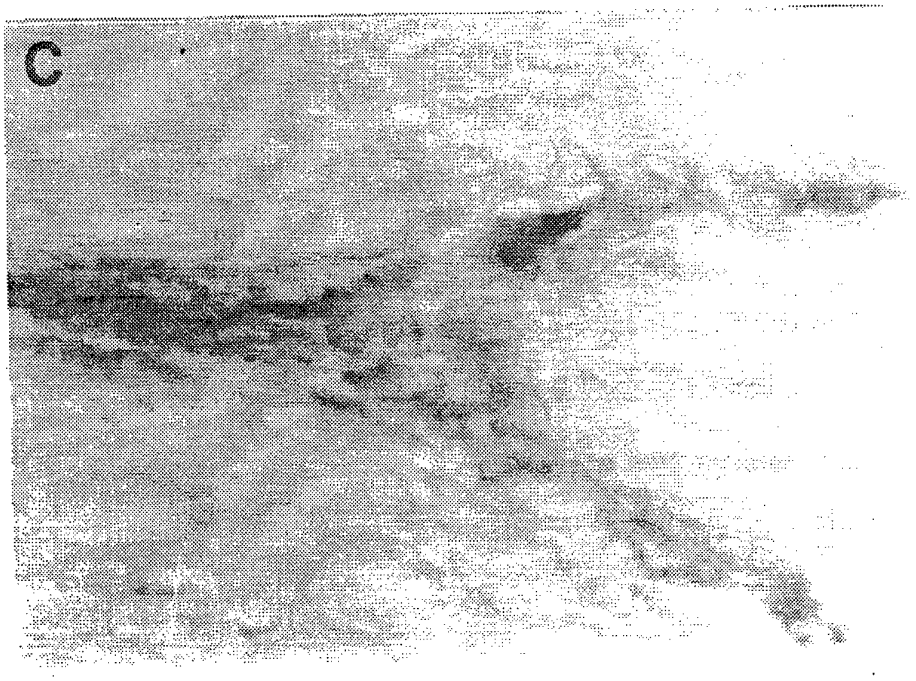
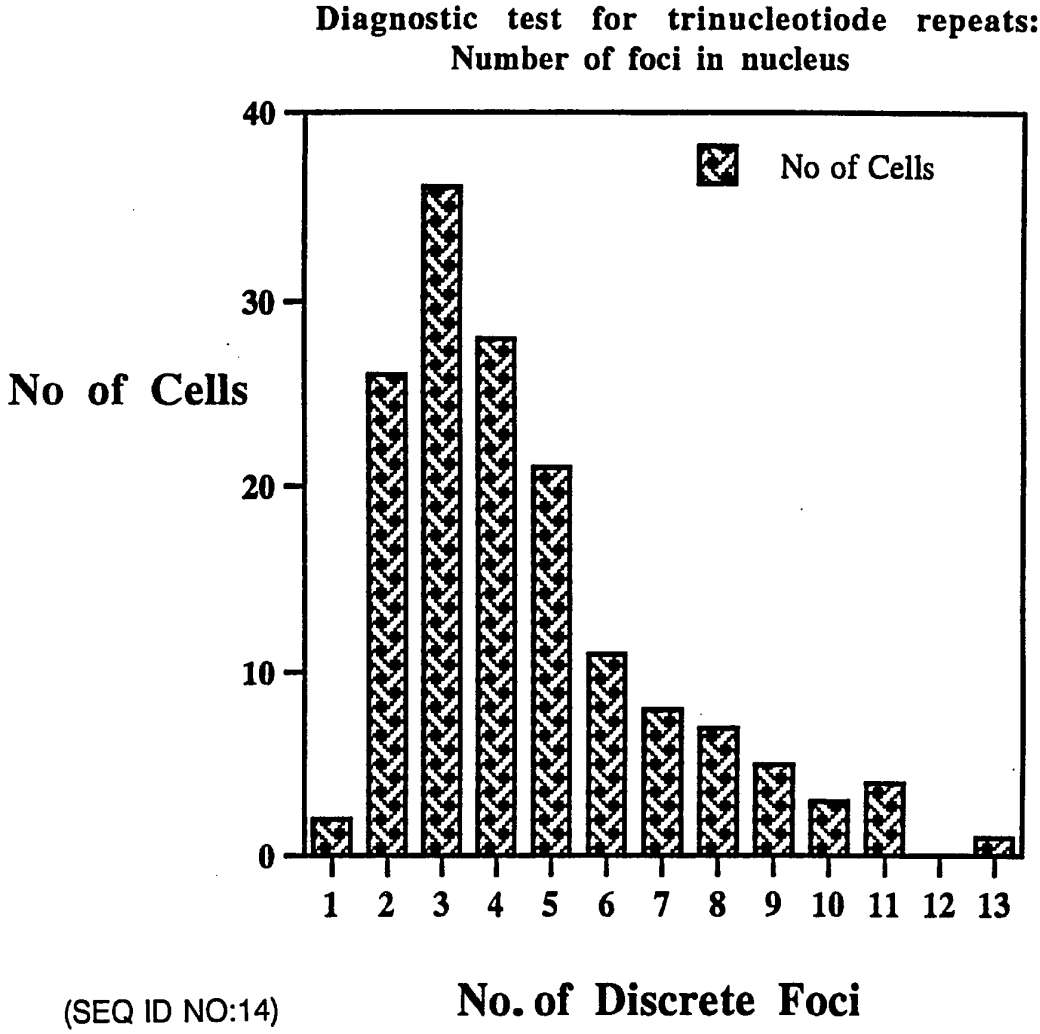


FIG. 4B



(SEQ ID NO:14)

FIG. 4C



(SEQ ID NO:14)

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US95/02861

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68
US CL :435/6
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; 536/24.3-24.32; 935/76-78

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.
Y	Grossman et al., METHODS IN ENZYMOLOGY" Vol. XXI, "Nucleic Acids", Part D, published 1971 by Academic Press (New York), pages 470-480, see pages 470-80.	1-18
Y	CELL, Vol. 68, issued February 1992, Brook et al., "Molecular Basis of Myotonic Dystrophy: Expansion of a Trinucleotide (CTG) Repeat at the 3' End of a Transcript Encoding a Protein Kinase Family Member", pages 799-808, see pages 799-807.	1-18
Y	US, A, 5,102,996 (LITT ET AL.) 07 April 1992, see entire document.	2, 3, and 12
Y	US, A, 5,139,031 (GUIRGUIS) 18 August 1992, see entire document.	2, 3, and 12

☐ 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
"E"	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
18 APRIL 1995

Date of mailing of the international search report
17MAY1995

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
BRADLEY L. SISSON
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS

search terms:myotonic dystrophy, hybridization, in situ, trinucleotide, trinucleotide(w)repeat, huntington?, huntington?(w)disease, in situ(w)hybridization(w)fluorescent, image cytometry