Method of Detecting a Repeat Expansion Sequence

The present disclosure relates to the field of molecular biology. In particular, the specification teaches a method of detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample and a method of screening for Multiple Repeat Expansion Disorders in a subject.

BACKGROUND

Expansion of simple sequence repeats that are dispersed throughout the human genome are now known to directly cause over 35 human diseases. The most frequent repeat expansion disorders are caused by trinucleotide repeats, although tetra-, penta-, hexa-, and even dodecanucleotide repeat expansions have also been identified as the underlying mutations in other diseases.

Repeat expansion disorders often manifest with markedly varied phenotypes and are usually difficult to distinguish by signs and symptoms alone due to extensive clinical overlap and other accompanying phenotypes. Molecular genetic testing is therefore necessary to identify the causative mutations to confirm disease status in symptomatic individuals. Several diseases are caused by common repeat expansion mutation located at different gene loci; for example CGG or GCG repeat expansions cause fragile X syndrome (FXS), several types of oculopharyngodistal myopathies, oculopharyngeal muscular dystrophy and developmental and epileptic encephalopathy, CCG repeat expansions cause a type of intellectual development and an ocuopharyngeal myopathy with leukoencephalopathy, CAG repeat expansions cause Huntington disease and several types of spinocerebellar ataxia (SCA), CTG repeat expansions cause myotonic dystrophy type 1, Huntington disease-like 2, and Fuchs endothelial corneal dystrophy, and TTTCA pentanucleotide repeat expansions cause several types of familial adult myoclonic epilepsies and a spinocerebellar ataxia. Multiple rounds of genetic testing may be required to achieve the correct diagnosis in affected individuals, resulting in additional costs and time.

Accordingly, there is a need to overcome, or at least to alleviate, one or more of the above-mentioned problems.

SUMMARY

Disclosed herein is a method of detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject, the method comprising: i) contacting the nucleic acid sample under amplification conditions for each of the two or more genes with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene wherein the gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence, and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; and ii) analyzing the amplification products.

Disclosed herein is a kit for detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject, the kit comprising: a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene.

Disclosed herein is a composition comprising a nucleic acid sample obtained from a subject, the composition comprising a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene;

wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene.

Disclosed herein is a method of screening for one or more Multiple Repeat Expansion Disorders in a subject, the method comprises: i) contacting a nucleic acid sample from the subject under amplification conditions with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; and ii) analyzing the amplification products to screen for the one or more Multiple Repeat Expansion Disorders in the subject.

Disclosed herein is a method of screening for one or more Multiple Repeat Expansion Disorders in a subject and treating the subject, the method comprises: i) contacting a nucleic acid sample from the subject under amplification conditions with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; ii) analyzing the amplification products to screen for the one or more Multiple Repeat Expansion Disorders in the subject; and iii) treating the subject found to have at least one Multiple Repeat Expansion Disorder.

BRIEF DESCRIPTION OF THE DRAWINGS

Some embodiments of the present invention will now be described by way of non-limiting example only, with reference to the accompanying drawings in which:

Figure 1. Schematic illustration of the *FMR1* and *AFF2* duplex TP-PCR reaction. The locus-specific flanking primer annealing positions are indicated (black and grey arrows with its end tagged with a star symbol), as are the TP primer annealing positions within the repeat (black and grey arrows). Grey arrows indicate poorer amplification because the TP primer is further from the locus-specific primer or because of interruption-mediated mismatch. Expected electropherogram patterns of samples carrying only normal *FMR1* and normal *AFF2* alleles (A), an expanded *FMR1* allele (B), and an expanded *AFF2* allele (C) are shown. Black rectangles represent *FMR1* CGG or *AFF2* CCG trinucleotides. Grey rectangles represent non-CGG or non-CCG trinucleotides as indicated above each rectangle.

Figure 2. Electropherograms of *FMR1* (FAM/dark grey) and *AFF2* (HEX/light grey) duplex TP-PCR products from normal, *FMR1* premutation and *FMR1* full mutation male and female DNA samples. Electropherograms are displayed in both FAM and HEX (left), FAM only (center), and HEX only (right) fluorescent channels. dark grey peaks indicate the FAM-labeled *FMR1* TP-PCR product, while light grey peaks indicate the HEX-labeled *AFF2* TP-PCR product. The threshold repeat size separating normal from expanded alleles is indicated by a dotted vertical line.

Figure 3. Electropherograms of *FMR1* (FAM/dark grey) and *AFF2* (HEX/light grey) duplex TP-PCR products from *AFF2* premutation and full mutation DNA samples. Electropherograms are displayed in both FAM and HEX (left), FAM only (center), and HEX only (right) fluorescent channels. Dark grey peaks indicate the FAM-labeled *FMR1* TP-PCR product, while light grey peaks indicate the HEX-labeled *AFF2* TP-PCR product. The threshold repeat size separating normal from expanded alleles is indicated by a dotted vertical line.

Figure 4. *AFF2* CCG repeat size and structure distributions. A, Population distribution of *AFF2* CCG repeat sizes (x-axis) and frequencies (y-axis) in African American (grey-filled), Caucasian (unfilled), Chinese (black-filled), Indian (dark grey bold stripe) and Malay (grey forward slash) populations. B, Comparison of Chinese allele frequency distributions in Zhong et al. (grey) and this study (black). C, Comparison of Caucasian allele frequency distributions in Zhong et al. (light grey) and this study (grey). D, Heatmaps depicting the population distribution of X chromosomes with different *FMR1* CGG and *AFF2* CCG repeat size

combinations (top), and the population repeat size distribution and abundance of the common and variant *AFF2* alleles (bottom).

Figure 5. Spectrum of *AFF2* CCG repeat structures and corresponding TP-PCR electropherogram patterns. A, Common normal allele with rs868914124(T) reference nucleotide. B, Variant normal allele with rs868914124(C) variant nucleotide. C, Variant normal allele with rs868914124(T) reference nucleotide and rs1389911365(T) variant nucleotide. D, Full mutation allele with rs868914124(C) variant nucleotide and three consecutive non-CCG interruptions. "+" indicates a non-CCG interruption, while "+++" indicates three consecutive non-CCG interruptions. The 111 bp leftmost solitary peak is generated from TP primer annealing upstream of the repeat tract. The first repeat-generated TP-PCR product of the rs868914124(T) common allele migrates as a 138 bp fragment. The first repeat-generated TP-PCR product from the rs868914124(C) rare allele migrates as a 132 bp fragment. Intra-repeat interruptions lead to mismatched pairing of the TP primer, resulting in gaps of absent peaks within the peak cluster.

Figure 6. *AFF2* CCG repeat structures and corresponding TP-PCR electropherogram patterns of (A) full mutation male FX0229 and (B) his full mutation maternal uncle FX0230. A, Full mutation allele carries the rs868914124(C) rare variant and three consecutive non-CCG interruptions at repeats 8-10. B, Full mutation allele with the rs868914124(C) rare variant and multiple non-CCG interruptions. "+++" indicates three consecutive non-CCG interruptions.

Figure 7. *AFF2* CCG repeat structures and corresponding TP-PCR electropherogram patterns of (A) full mutation male DNA_25926 and (B) his premutation daughter DNA_3802. A, Full mutation allele carrying the rs868914124(C) rare variant and three consecutive non-CCG interruptions at repeats 6-8. B, Normal allele of 22 CCG repeats carrying the rs868914124(T) common reference nucleotide, and premutation allele of ~137 repeats with the rs868914124(C) rare variant and three consecutive non-CCG interruptions at repeats 6-8. "+++" indicates three consecutive non-CCG interruptions.

Figure 8. Multiplex triplet-primed PCR of seven common SCA repeat loci. Schematic illustrations of SCA1, SCA2, SCA3, SCA6, SCA7, SCA12, and DRPLA repeat structures and primer annealing positions **(A)** and the expected TP-PCR electropherograms for each repeat

locus **(B)**. Asterisk indicates a fluorophore tag. The upper limit of the normal allele repeat size for each locus is indicated by a dotted vertical line.

Figure 9. Heptaplex TP-PCR capillary electropherogram result from a DNA sample negative for expansion at any of the seven SCA repeat loci. Top panel shows electrophoretic peaks representing TP-PCR product from all seven repeat loci visible with all four fluorophore channels (Fam, Vic, Ned, and Pet) turned on. The four panels below show the same capillary electropherogram but with one fluorophore channel turned on at a time. The upper limit of the normal allele repeat size for each locus is indicated by a dotted vertical line.

Figure 10. Heptaplex TP-PCR capillary electropherogram results of SCA-affected DNA samples. For each sample, only the electropherogram with the fluorophore channel showing repeat expansion at the relevant SCA locus is shown. Samples are positive for CAG repeat expansion in *ATXN1* (top row), *ATXN2* (second row), *ATXN3* (third row), *CACNA1A* (fourth row), *ATXN7* (fifth row), *PPP2R2B* (sixth row), and *ATN1* (bottom row). The upper limit of the normal allele repeat size for each locus is indicated by a dotted vertical line.

DETAILED DESCRIPTION

The present specification teaches a method of detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject, the method comprising: i) contacting the nucleic acid sample under amplification conditions for each of the two or more genes with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene wherein the gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence, and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; and ii) analyzing the amplification products.

In one embodiment, the method is a simultaneous method of detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject

Without being bound by theory, the inventors have developed a strategy to simultaneously screen a patient for the presence of repeat expansion mutation in two or more suspected genes, and describe its application in this invention disclosure using two examples. The strategy employs a single-tube assay to screen multiple genetic loci responsible for different repeat expansion disorders that are caused by expansion of the same type of repeat. The specific assays illustrated herein employ multiplex triplet-primed PCR (TP-PCR) to detect expansion mutations involving, for example, trinucleotide repeats that are present in different genes, through the use of differentially labeled locus-specific flanking primers and a common triplet-primed (TP) primer. Amplified products at any locus that show repeat size within the pathogenic size range, or exceed the largest normal repeat size, can be quickly identified and sized by capillary electrophoresis. A single amplification reaction provides a reliable one-step mutation screen of multiple disease genes, therefore considerably shortening the diagnostic odyssey for affected individuals. This strategy can be applied to simultaneous screening of any group of disease genes which share the same repetitive sequence.

In one embodiment, the amplification products are separated according to size. In one embodiment, the change in the sizes of the amplification products for a gene as compared to a reference indicates the presence of a repeat expansion sequence in the gene. In one embodiment, the lack of change in sizes of the amplification products for a gene as compared to a reference indicates the absence of a repeat expansion sequence in the gene.

The terms “detecting”, “determining”, “measuring”, “evaluating”, “assessing” and “assaying” are used interchangeably herein to refer to any form of measurement, and include determining if an element is present or not. These terms include both quantitative and/or qualitative determinations. Assessing may be relative or absolute.

The term "nucleic acid" refers to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides.

As used herein, the term "nucleic acid", and equivalent terms such as “polynucleotide”, refer to a polymeric form of nucleotides of any length, such as ribonucleotides, deoxyribonucleotides

or peptide nucleic acids (PNAs), that comprise purine and pyrimidine bases, or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. The nucleic acid may be double stranded or single stranded. References to single stranded nucleic acids include references to the sense or antisense strands. The backbone of the polynucleotide can comprise sugars and phosphate groups, as may typically be found in RNA or DNA, or modified or substituted sugar or phosphate groups. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. The sequence of nucleotides may be interrupted by non-nucleotide components. The terms nucleoside, nucleotide, deoxynucleoside and deoxynucleotide generally include complements, fragments and variants of the nucleoside, nucleotide, deoxynucleoside and deoxynucleotide, or analogs thereof.

The term "primer" is used herein to mean any single-stranded oligonucleotide sequence capable of being used as a primer in, for example, PCR technology. Thus, a "primer" according to the disclosure refers to a single-stranded oligonucleotide sequence that is capable of acting as a point of initiation for synthesis of a primer extension product that is substantially identical to the nucleic acid strand to be copied (for a forward primer) or substantially the reverse complement of the nucleic acid strand to be copied (for a reverse primer). A primer may be suitable for use in, for example, PCR technology. By single-stranded includes, for example, hairpin structures formed by single-stranded nucleotide sequences.

The design of a primer, for example its length and specific sequence, depends on the nature of the target nucleotide sequence and on the conditions at which the primer is used, for example, temperature and ionic strength.

The primers may consist of the nucleotide sequences described herein, or may be 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100 or more nucleotides which comprise or fall within the sequences described herein, provided they are suitable for specifically binding a target sequence, under stringent conditions. In one embodiment, the primer sequence is less than 35 nucleotides in length, for example the primer sequence is less than 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22 or 21 nucleotides in length.

Slight modifications of the primers or probes, in length or in sequence, can be carried out to maintain the specificity and sensitivity required under the given circumstances. In one embodiment of the present disclosure, probes and/or primers described herein may be extended

in length by 1, 2, 3, 4 or 5 nucleotides or reduced in length by 1, 2, 3, 4 or 5 nucleotides, for example, in either direction.

Primer sequences can be synthesised using any methods well known in the art.

The term “complementary” refers to the base pairing between nucleotides or nucleic acids, such as, for instance, between the two strands of a double stranded DNA molecule or between an oligonucleotide primer and a primer binding site on a single stranded nucleic acid to be sequenced or amplified. Complementary nucleotides are, generally, A and T (or A and U), or C and G. Two single stranded RNA or DNA molecules are said to be complementary when the nucleotides of one strand, optimally aligned and compared and with appropriate nucleotide insertions or deletions, pair with at least about 80% of the nucleotides of the other strand, usually at least about 90% to 95%, and more preferably from about 98 to 100% of the nucleotides of the other strand. Alternatively, complementarity exists when an RNA or DNA strand will hybridize under selective hybridization conditions to its complement. Typically, selective hybridization will occur when there is at least about 65% complementarity over a stretch of at least 14 to 25 nucleotides, preferably at least about 75%, and more preferably at least about 90% complementarity.

As used herein, the term "hybridization" or "hybridizes" refers to the process in which two single-stranded polynucleotides bind non-covalently to form a stable double-stranded polynucleotide. The resulting (usually) double-stranded polynucleotide is a "hybrid". The proportion of the population of polynucleotides that forms stable hybrids is referred to herein as the "degree of hybridization."

Hybridization conditions will typically include salt concentrations of less than about 1M, more usually less than about 500 mM and less than about 200 mM. Hybridization temperatures can be as low as 5ºC, but are typically greater than 22ºC, more typically greater than about 30ºC, and preferably in excess of about 37ºC. Hybridizations are usually performed under stringent conditions, i.e. conditions under which a probe will hybridize to its target subsequence. Stringent conditions are sequence-dependent and are different under different circumstances. Longer fragments may require higher hybridization temperatures for specific hybridization. As other factors may affect the stringency of hybridization, including base composition and length of the complementary strands, presence of organic solvents and extent of base mismatching,

the combination of parameters is more important than the absolute measure of any one alone. Generally, stringent conditions are selected to be about 5ºC lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid composition) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Typically, stringent conditions include salt concentration of at least 0.01 M to no more than 1 M Na ion concentration (or other salts) at a pH 7.0 to 8.3 and a temperature of at least 25ºC. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30ºC are suitable for allele-specific probe hybridizations.

"Specific binding" or "specific hybridisation" of a primer to a target sequence means that the primer forms a duplex (double-stranded nucleotide sequence) with part of the target sequence region or with the entire target sequence as required, under the experimental conditions used, for example under stringent hybridisation conditions, and that under those conditions the primer does not form a duplex with other regions of the nucleotide sequence present in the sample to be analysed.

The term "repeat expansion sequence" may refer to a nucleotide repeat sequence that has undergone an expansion mutation, resulting in the lengthening of the nucleotide repeat sequence beyond the normal repeat size.

In one embodiment, the repeat expansion sequence is a trinucleotide repeat expansion sequence. In other embodiments, the repeat expansion sequence is a tetranucleotide or pentanucleotide repeat expansion sequence or repeat expansion sequences of other lengths, such as 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20 nucleotide repeat sequences.

In one embodiment, the nucleotide repeat sequence is a trinucleotide repeat sequence. In one embodiment, the trinucleotide repeat sequence is selected from $(CGG)_n$, $(CCG)_n$, $(CAG)_n$, $(CTG)_n$, $(GCC)_n$, $(GGC)_n$, $(GAA)_n$ or $(TTC)_n$, wherein n is from 2 to 200 or more. A $(CGG)_n$ repeat sequence is taken to include $(GCG)_n$ and $(GGC)_n$. Similarly a $(CCG)_n$ sequence includes $(CGC)_n$ and $(GCC)_n$, and a $(CTG)_n$ sequence includes $(TGC)_n$ and $(GCT)_n$.

The “common primer” as referred to herein may bind to a “common target sequence” that is shared by the two or more genes, wherein the “common target sequence” is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer

The "gene-specific primer" as referred to herein may refer to a primer that binds specifically to a particular gene, but not other genes.

The term "gene" is used broadly to refer to any nucleic acid associated with a biological function. Genes typically include coding sequences and/or the regulatory sequences required for expression of such coding sequences. The term gene can apply to a specific genomic sequence, as well as to a cDNA or an mRNA encoded by that genomic sequence. Genes also include non-expressed nucleic acid segments that, for example, form recognition sequences for other proteins. Non-expressed regulatory sequences include promoters and enhancers, to which regulatory proteins such as transcription factors bind, resulting in transcription of adjacent or nearby sequences.

In one embodiment, the common primer binds to a common target sequence comprising or consisting of 5, 6, 7, 8, 9, 10 or more consecutive trinucleotide repeats. In one embodiment, the common primer binds to $(CGG)_5$ (SEQ ID NO: 15). In one embodiment, the common primer binds to $(CTG)_5$ (SEQ ID NO: 16).

In one embodiment, the common primer binds to a common target sequence comprising or consisting of 2, 3, 4, 5, 6, 7, 8, 9, 10 or more consecutive tetranucleotide, pentanucleotide or larger nucleotide repeat units. In one embodiment, the pentanucleotide repeat sequence is $(TTTCA)_n$, which is taken to include $(CATTT)_n$ and $(ATTTC)_n$.

In one embodiment, the common primer comprises a unique tail sequence. In one embodiment, the common primer comprises a unique 5’ tail sequence. The term "unique tail sequence" or “unique 5' tail sequence” refers to a sequence that does not hybridise under stringent conditions to any region in any gene or intergenic region in a nucleic acid sample for detecting the presence or absence of a repeat expansion sequence.

In one embodiment, the method comprises providing a tail primer that binds specifically to the unique 5' tail sequence of the common primer. The addition of a unique 5' tail sequence to the common primer and providing a tail primer that binds specifically to the unique 5'tail sequence may improve repeat sizing accuracy.

In one embodiment, the gene-specific primer is labeled. Each gene-specific primer may be, for example, labeled with a different fluorophore to enable amplification products from each gene to be distinguished from one another. For example, a FAM or Hex label can be used.

In one embodiment the fluorescent labels may be active in the blue, yellow, green and far red areas of the spectrum. In a preferred embodiment, non-limiting examples of fluorescent labels that may be used in the method of the disclosure include: fluorescent markers or reporter dyes, for example, 6-carboxyfluorescein (6FAM™), NED™ (Applera Corporation), HEX™ or VIC™ (Applied Biosystems); TAMRA™ markers (Applied Biosystems, CA, USA), ROX. It will be appreciated by those of skill in the art that other alternative fluorescent labels may also be used in the method according to the present disclosure.

In another embodiment chemiluminescent markers may be used, for example Ruthenium probes; and radioactive labels, for example tritium in the form of tritiated thymidine. 32-Phosphorus may also be used as a radiolabel.

Alternatively the label may be selected from the group consisting of electroluminescent tags, magnetic tags, affinity or binding tags, nucleotide sequence tags, position specific tags, and or tags with specific physical properties such as different size, mass, gyration, ionic strength, dielectric properties, polarisation or impedance.

In one embodiment, the detectable label may be directly or indirectly attached to the primer. In one embodiment, the labeled primer is the reverse primer. In one embodiment, the detectable label comprises a fluorescent moiety attached at a $5'$ end of the probe. In a most preferred embodiment, the label is selected from 6-FAM and NED.

In an alternative embodiment, the nucleic acid is detected with a nucleic acid-intercalating fluorophore. Preferably, the intercalating fluorophore is SYBR Green or EvaGreen and the like. It will be appreciated by those of skill in the art that other intercalating fluorophores may be

used that are active in the blue, yellow, green and far red areas of the spectrum. It will be further appreciated that other intercalating fluorophores may be used in accordance with the present disclosure.

The term "sample" referred to herein may be derived from a biological fluid, cell, tissue, organ, or organism, that includes a nucleic acid or a mixture of nucleic acids having at least one nucleic acid sequence that is to be screened for copy number variation. In certain embodiments the sample has at least one nucleic acid sequence whose copy number is suspected of having undergone variation. Such samples include, but are not limited to sputum/oral fluid, amniotic fluid, blood, a blood fraction, or fine needle biopsy samples, urine, peritoneal fluid, pleural fluid, and the like. Although the sample is often taken from a human subject (e.g., a patient), the sample may be taken from any mammal, including, but not limited to dogs, cats, horses, goats, sheep, cattle, pigs, etc. The sample may be used directly as obtained from the biological source or following a pretreatment to modify the character of the sample. For example, such pretreatment may include preparing plasma from blood, diluting viscous fluids, and so forth. Methods of pretreatment may also involve, but are not limited to, filtration, precipitation, dilution, distillation, mixing, centrifugation, freezing, lyophilization, concentration, amplification, nucleic acid fragmentation, inactivation of interfering components, the addition of reagents, lysing, etc. If such methods of pretreatment are employed with respect to the sample, such pretreatment methods are typically such that the nucleic acid(s) of interest remain in the test sample, sometimes at a concentration proportional to that in an untreated test sample (e.g., namely, a sample that is not subjected to any such pretreatment method(s)).

A control sample may be a negative or positive control sample. A "negative control sample" or "unaffected sample" refers to a sample including nucleic acids that is known or expected to have a repeat sequence having a number of repeats within a range that is not pathogenic. A "positive control sample" or "affected sample" is known or expected to have a repeat sequence having a number of repeats within a range that is pathogenic. Repeats of the repeat sequence in a negative control sample typically have not been expanded beyond a normal range, whereas repeats of a repeat sequence in a positive control sample typically have been expanded beyond a normal range. As such, the nucleic acids in a test sample can be compared to one or more control samples.

The term "biological fluid" herein refers to a liquid taken from a biological source and includes, for example, blood, serum, plasma, sputum, lavage fluid, cerebrospinal fluid, urine, semen, sweat, tears, saliva, and the like. As used herein, the terms "blood," "plasma" and "serum" expressly encompass fractions or processed portions thereof. Similarly, where a sample is taken from a biopsy, swab, smear, etc., the "sample" expressly encompasses a processed fraction or portion derived from the biopsy, swab, smear, etc.

A "label" refers to a reporter molecule (such as a fluorophore) that is capable of generating a measurable signal and is covalently or non-covalently joined to a polynucleotide.

In one embodiment, the method comprises analyzing the amplification products using a size separation technique or a sequencing technique. The size separation technique may be an electrophoresis based technique (such as capillary electrophoresis). For example, capillary electrophoresis through a size separation polymer can be used. Fluorescently labelled PCR product (via the use of a 5' end-labeled primer) is detected via laser excitation as it migrates and resolves through the polymer-filled capillary. Another size separation technique is slab gel PAGE (polyacrylamide gel electrophoresis). The PCR product can be labelled fluorescently using a 5' end-labeled primer and detected via laser excitation as it migrates and resolves through the slab gel. The PCR product can also be labeled radioactively via a 5' end-labeled primer or via incorporation of radio-isotope tagged nucleotides during the PCR process, and detected by exposure of the slab gel to X-ray film. The sequencing technique may be next generation sequencing. By using long-read next-generation sequencing, the PCR product can be sequenced, and all the reads belonging to the target gene can be aligned from shortest to longest or vice versa.

The two or more genes as referred to herein may be adjacent, or in close proximity to one another on the same chromosome. They may alternatively be on different chromosomes.

In one embodiment, the two or more genes consist of or comprises *FMR1* and *AFF2*. *FMR1* and *AFF2* are located on the same chromosome (X) on the q-arm in adjacent chromosome bands. This provides a simple way of simultaneously testing two or more genes that include *FMR1* and *AFF2*, either of which may be responsible for a disorder exhibited by a patient.

In one embodiment, the method comprises contacting the nucleic acid sample with a) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 1; b) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 2; and c) a common primer comprising or consisting of a nucleic acid sequence having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 3 or $(CGG)_5$ (SEQ ID NO 15).

In one embodiment, the two or more genes include or are selected from the group consisting of SCA1, SCA2, SCA3, SCA6, SCA7, SCA12 and DRPLA. In one embodiment, the two or more genes consists of two, three, four, five, six or seven or more genes including or selected from the group consisting of SCA1, SCA2, SCA3, SCA6, SCA7, SCA12 and DRPLA

In one embodiment, the two or more genes is HTT and JPH3.

In one embodiment, the two or more genes consist of *ATXN1, ATXN2, ATXN3, CACNA1A, ATXN7, PPP2R2B,* and *ATN1.*

In one embodiment, the two or more genes include or are selected from the group consisting of *ATXN1, ATXN2, ATXN3, CACNA1A, ATXN7, PPP2R2B, ATN1, TBP, ATXN8, ATXN8OS, AR, HTT, JPH3, TCF4,* and *DMPK,* and consists of two, three, four, five, six, seven, eight, nine, ten or more, or all of these genes.

In one embodiment, the two or more genes include or are selected from the group consisting of *FMR1*, *AFF2, NUTM2BAS1, LRP12, GIPC1, NOTCH2NLC, RILPL1, PABPN1,* and *ARX,* and consists of two, three, four, five, six, seven, eight or all of these genes.

In one embodiment, the two or more genes include or are selected from the group consisting of *DAB1, SAMD12, STARD7, TNRC6A,* and *RAPGEF2,* and consists of two, three, four or all of these genes.

In one embodiment, the method comprises contacting the nucleic acid sample with: a) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 7; b) a gene-specific primer comprising or consisting

of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 8; c) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 9; d) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 10; e) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 11; f) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 12; and/or g) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 13, and h) a common primer comprising or consisting of a nucleic acid sequence having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 14 or $(CTG)_5$ (SEQ ID NO: 16).

Disclosed herein is a kit for detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject, the kit comprising: a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene.

Disclosed herein is a composition comprising a nucleic acid sample obtained from a subject, the composition comprising a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene.

Disclosed herein is a method of screening for one or more Multiple Repeat Expansion Disorders in a subject, the method comprises: i) contacting a nucleic acid sample from the subject under amplification conditions with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; and ii) analyzing the amplification products to screen for the one or more Multiple Repeat Expansion Disorders in the subject.

The term "multiple repeat expansion disorder" refers to a genetic disorder caused by an increase in the number of DNA repeat motifs (such as trinucleotide repeat motifs) in certain genes exceeding a normal, stable, threshold, which differs per gene. This term is intended to include all disorders of this nature, and includes the disorders listed in Table 1.

Table 1. Disorders caused by expansions of different repeat motifs

Disease	Symbol	OMIM	Inheritance	Gene	Cytogenetic Location	Type	Repeat Motif
Spinocerebellar ataxia 1	SCA1	164400	AD	*ATXN1*	6p23	coding	**CAG**
Spinocerebellar ataxia 2	SCA2	183090	AD	*ATXN2*	12q24	coding	**CAG**
Machado-Joseph disease	MJD	109150	AD	*ATXN3*	14q32.1	coding	**CAG**
Spinocerebellar ataxia 6	SCA6	183086	AD	*CACNA1A*	19p13	coding	**CAG**
Spinocerebellar ataxia 7	SCA7	164500	AD	*ATXN7*	3p14.1	coding	**CAG**
Spinocerebellar ataxia 12	SCA12	604326	AD	*PPP2R2B*	5q32	promoter	**CAG**
Spinocerebellar ataxia 17	SCA17	607136	AD	*TBP*	6q27	coding	**CAG**
Dentatorubral-pallidoluysian atrophy	DRPLA	125370	AD	*ATN1*	12p13.31	coding	**CAG**
Spinocerebellar ataxia 8	SCA8	608768	AD	*ATXN8* & *ATXN8OS*	13q21	coding & ncRNA	**CAG** and **CTG**
X-linked spinal and bulbar muscular atrophy 1	SMAX1	313200	XL	*AR*	Xq12	coding	**CAG**
Huntington disease	HD	143100	AD	*HTT*	4p16.3	coding	**CAG**
Huntington disease-like 2	HDL2	606438	AD	*JPH3*	16q24.2	coding	**CTG**

Fuchs endothelial corneal dystrophy 3	FECD3	613267	AD	*TCF4*	18q21.2	intron	**CTG**
Myotonic dystrophy 1	DM1	160900	AD	*DMPK*	19q13	3'UTR	**CTG**
Myotonic dystrophy 2	DM2	602668	AD	*CNBP*	3q21.3	intron	**CCTG**
Friedreich ataxia	FRDA	229300	AR	*FXN*	9q21.11	intron	**GAA**
Spinocerebellar ataxia 10	SCA10	603516	AD	*ATXN10*	22q13.31	intron	**ATTCT**
Spinocerebellar ataxia 31	SCA31	117210	AD	*BEAN1*	16q21	intron	**TGGAA**
Spinocerebellar ataxia 36	SCA36	614153	AD	*NOP56*	20p13	intron	**GGCCTG**
Cerebellar ataxia, neuropathy, and vestibular areflexia syndrome	CANVAS	614575	AR	*RFC1*	4p14	intron	**AAAAG**
Spinocerebellar ataxia 37	SCA37	615945	AD	*DAB1*	1p32.2	intron	**TTTCA**
Familial adult myoclonic epilepsy 1	FAME1	601068	AD	*SAMD12*	8q24.11/12	intron	**TTTCA**
Familial adult myoclonic epilepsy 2	FAME2	607876	AD	*STARD7*	2q11.2	intron	**TTTCA**
Familial adult myoclonic epilepsy 6	FAME6	618074	AD	*TNRC6A*	16p12.1	5'UTR	**TTTCA**
Familial adult myoclonic epilepsy 7	FAME7	618075	AD	*RAPGEF2*	4q32.1	intron	**TTTCA**
Myoclonic epilepsy of Unverricht and Lundborg	EPM1	254800	AR	*CSTB*	21q22.3	promoter	**CCCCGCCCCGCG (SEQ ID NO: 17)**
Frontotemporal dementia and/or amyotrophic lateral sclerosis 1	FTDALS1	105550	AD	*C9orf72*	9p21.2	intron	**GGGGCC**
X-linked intellectual development disorder 109	XLID109 (FRAXENSID)	309548	XL	*AFF2*	Xq28	5'UTR	**CCG**
Oculopharyngeal myopathy with leukoencephalopathy	OPML1	618637	AD	*NUTM2BAS1*	10q22.3	5'UTR	**CCG**
Fragile X syndrome	FXS	300624	XL	*FMR1*	Xq27.3	5'UTR	**CGG**
Oculopharyngodistal myopathy 1	OPDM1	164310	AD	*LRP12*	8q22.3	5'UTR	**CGG**
Oculopharyngodistal myopathy 2	OPDM2	618940	AD	*GIPC1*	19p13.12	5'UTR	**CGG**
Neuronal intranuclear inclusion disease	NIID	603472	AD	*NOTCH2NLC*	1q21.2	5'UTR	**CGG**
Oculopharyngodistal myopathy 3	OPDM3	619473					
Hereditary essential tremor 6	ETM6	618866					
Oculopharyngodistal myopathy 4	OPDM4	--	AD	*RILPL1*	12q24.31	5'UTR	**CGG**
Oculopharyngeal muscular dystrophy	OPMD	164300	AD	*PABPN1*	14q11.2	coding	**GCG**
Developmental and epileptic encephalopathy 1	DEE1	308350	XL	*ARX*	Xp21.3	coding	**GCG**

In one embodiment, the one or more Multiple Repeat Expansion Disorders comprise or consist of one or more disorders listed in Table 1.

In one embodiment, the one or more Multiple Repeat Expansion Disorders comprise or consist of a disorder including or selected from the group consisting of fragile X syndrome (FXS), fragile X-related primary ovarian insufficiency (FXPOI), fragile X-associated tremor/ataxia syndrome (FXTAS) and fragile XE non-syndromic intellectual disability (FRAXE NSID).

In one embodiment, the method differentiates between FXS, FXPOI, FXTAS and FRAXE NSID.

In one embodiment, the one or more Multiple Repeat Expansion Disorders is spinocerebellar ataxias (SCA) and/or dentatorubral pallidoluysian atrophy (DRPLA).

In one embodiment, the method differentiates between SCA and DRPLA. In one embodiment, the method differentiates between the different SCA types and DRPLA.

In one embodiment, the one or more Multiple Repeat Expansion Disorders is Huntington disease (HD) and Huntington disease-like 2 (HDL2).

In one embodiment, the method differentiates between Huntington disease (HD) and Huntington disease-like 2 (HDL2).

In one embodiment, the method differentiates between the different oculopharyngodistal myopathy (OPDM) types, oculopharyngeal muscular dystrophy (OPMD), oculopharyngeal myopathy with leukoencephalopathy (OPML) and developmental and epileptic encephalopathy (DEE).

In one embodiment, the method differentiates between the different familial adult myoclonic epilepsy (FAME) types and spinocerebellar ataxia (SCA) type 37.

By “subject” or “patient” is meant any single subject for which therapy is desired, including humans, cattle, horses, pigs, goats, sheep, dogs, cats, guinea pigs, rabbits, chickens, insects and so on. Also intended to be included as a subject are any subjects involved in clinical research trials not showing any clinical sign of disease, or subjects involved in epidemiological studies, or subjects used as controls.

Disclosed herein is a method of screening for one or more Multiple Repeat Expansion Disorders in a subject and treating the subject, the method comprises: i) contacting a nucleic acid sample from the subject under amplification conditions with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein

each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; ii) analyzing the amplification products to screen for the one or more Multiple Repeat Expansion Disorders in the subject; and iii) treating the subject found to have at least one Multiple Repeat Expansion Disorder.

The terms “treating”, “treatment” and the like also include relieving, reducing, alleviating, ameliorating or otherwise inhibiting the effects of the condition for at least a period of time. It is also to be understood that terms “treating”, “treatment” and the like do not imply that the condition, or a symptom thereof, is permanently relieved, reduced, alleviated, ameliorated or otherwise inhibited and therefore also encompasses the temporary relief, reduction, alleviation, amelioration or otherwise inhibition of the condition, or of a symptom thereof.

The method of treating the subject may comprise administering a drug to the subject or may comprise providing early intervention management to the subject.

The term “administering” refers to contacting, applying, injecting, transfusing or providing an inhibitor as referred to herein to a subject.

Throughout this specification and the statements which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

As used in the subject specification, the singular forms "a", "an" and "the" include plural aspects unless the context clearly dictates otherwise. Thus, for example, reference to "a method" includes a single method, as well as two or more methods; reference to "an agent" includes a single agent, as well as two or more agents; reference to "the disclosure" includes a single and multiple aspects taught by the disclosure; and so forth. Aspects taught and enabled herein are encompassed by the term "invention". Any variants and derivatives contemplated

The reference in this specification to any prior publication (or information derived from it), or to any matter which is known, is not, and should not be taken as an acknowledgment or admission or any form of suggestion that that prior publication (or information derived from it) or known matter forms part of the common general knowledge in the field of endeavour to which this specification relates.

EXAMPLES

Certain embodiments of the invention will now be described with reference to the following examples which are intended for the purpose of illustration only and are not intended to limit the scope of the generality hereinbefore described.

Example 1: *FRAXA/FRAXE*

Materials and Methods

Biological samples

Genomic DNA was extracted from 408 unrelated and anonymized cord bloods of 161 Chinese, 158 Malay and 89 Indian male babies born at the National University Hospital, Singapore. Another 44 Caucasian and 14 African American male DNA samples from the Human Variation DNA panel (HD100CAU and HD100AA-2) were purchased from Coriell Cell Repositories (Camden, New Jersey, USA). Archived and previously characterized genomic DNAs consisting of 40 normal, 17 *FMR1* premutation positive, 23 *FMR1* full mutation positive, and six *AFF2* expansion-positive samples were utilized in assay validation. The de-identified normal and *FMR1* CGG repeat expansion-positive samples were obtained from the KK Women's and Children's Hospital. The *AFF2* CCG repeat expansion-positive samples were de-identified archival samples from Baylor College of Medicine (Houston, TX, USA) and The University of Adelaide (Adelaide, South Australia, Australia). Four of the six *AFF2* CCG repeat expansion-positive samples are related. FX0230 and FX0229 are maternal uncle and nephew, respectively, while DNA_25926 and DNA_3802 are father and daughter.

Duplex TP-PCR of the *FMR1* and *AFF2* trinucleotide repeats

The duplex screen for *AFF2* and *FMR1* triplet repeat expansions utilized a triplet-primed (TP) PCR approach, involving four primers, *Fam*-labeled *FMR1-R* (5'-AGCCCCGCACTTCCACCACCAGCTCCTCCA-3' (SEQ ID NO: 1)), *Hex*-labeled *AFF2-F* (5'-CCATGTCGCGGCTTCTAGCTGTCCAGGCTCC-3' (SEQ ID NO: 2)), and shared primers *TP* (5'-TGCTCTGGACCCTGAAGTGTGCCGTTGATACGGCGGCGGCGGCGG-3' (SEQ ID NO: 3)) and *Tail* (5'-TGCTCTGGACCCTGAAGTGTGCCGTTGATA-3' (SEQ ID NO: 4)). Each 15 µl PCR reaction contained 100 ng of genomic DNA, 2.5x Q-Solution (Qiagen, Hilden, Germany), 1x PCR buffer containing 1.5 mM of $MgCl_2$ (Qiagen), 2 mM dNTP at a 5:1 ratio of dGTP/dCTP to dATP/dTTP (Roche Applied Science, Mannheim, Germany), 0.6 µM each of *AFF2*-F, *FMR1*-R and Tail primers, 0.0006 µM of TP primer, and 5 U of HotStarTaq DNA polymerase (Qiagen). An initial 15 min enzyme activation at 95°C was followed by 40 cycles of 99°C for 45 sec, 55°C for 45 sec and 70°C for 8 min with a 15 sec increment per extension cycle, and a final extension of 10 min at 72°C.

Standard PCR across the *FMR1* and *AFF2* trinucleotide repeats

Standard/conventional PCR across the *FMR1* CGG repeat utilized 0.6 µM each of primers *5'-F1* and *Fam*-labeled *FMR1-R*, while PCR across the *AFF2* CCG repeat utilized *Hex*-labeled *AFF2-F* and *AFF2-R* (5'-CGCTGCGGGCTCAGGCGGGCT-3' (SEQ ID NO: 5)). PCR reaction and cycling conditions were similar to the above-mentioned duplex TP-PCR except that 50 ng of genomic DNA was used in each reaction.

Capillary electrophoresis

A 2 µl aliquot of the duplex TP-PCR product was mixed with 9 µl of Hi-Di™ formamide and 0.5 µl of GeneScan™ 500 ROX™ dye size standard (Applied Biosystems, Foster City, California, USA), denatured at 95°C for 5 min, cooled to 4°C, and resolved in a 3130xl Genetic Analyzer using a 36 cm capillary filled with POP-7™ polymer (Applied Biosystems). The mixture was electrokinetically injected at 1 kV for 5 sec and electrophoresed for 40 min at 60°C. Analysis was performed using GeneMapper™ Software v4.0 (Applied Biosystems). If an expanded allele was detected after the initial capillary electrophoresis (CE) run, a second CE run was performed using an electrokinetic injection at 10 kV for 5 sec and electrophoresed for 40 min at 60°C.

A 1 µl aliquot of one-twentieth diluted standard PCR product was mixed with 9 µl of Hi-Di™ formamide (Applied Biosystems) and 0.3 µl of GeneScan™ 500 ROX™ dye size standard

(Applied Biosystems), denatured at 95°C for 5 min, cooled to 4°C, and resolved in a 3130xl Genetic Analyzer (Applied Biosystems) using a 36 cm capillary filled with POP-7™ polymer. The mixture was electrokinetically injected at 1.2 kV for 23 sec and electrophoresed for 20 min at 60°C.

Post-CE analysis was performed using GeneMapper™ Software v4.0 (Applied Biosystems).

Data interpretation

Primer *AFF2*-F anneals to the *AFF2* sequence immediately upstream of its CCG repeat and generates *Hex*-labeled TP-PCR amplicons, while primer *FMR1*-R anneals to the *FMR1* sequence immediately downstream of its CGG repeat and generates *Fam*-labeled TP-PCR products. Both TP-PCR reactions utilize a common triplet-primed (TP) primer and Tail primer and their products can be analyzed separately using different fluorescence detection channels, or together. The TP primer is designed to anneal optimally to any stretch of five CCG trinucleotides on the *AFF2* sense strand or the *FMR1* antisense strand. When there is a continuous uninterrupted stretch of more than five repeats, the electropherogram will display a continuous series of peaks that are set 3 bp apart from each other. If a non-CCG interruption is present, the electropherogram will display a gap of ~18 bp between the fluorescent peaks, equivalent to the absence of 5 fluorescent peaks (Figure 1). The repeat size of an uninterrupted stretch is the total number of continuous fluorescent peaks plus four (the first fluorescent peak is generated from the TP primer annealing to the first 5 repeats). For an allele with non-CCG interruption(s), the repeat size is the total number of continuous fluorescent peaks added to the total number of missing peaks, plus four. In heterozygous females, one will see a drop in fluorescence intensity for peaks beyond a certain size; the repeat size of the smaller allele can be derived from the number of peaks before the drop while the repeat size of the larger allele can be derived from the total number of peaks (data not shown). For the *AFF2* TP-PCR reaction, the leftmost solitary peak is not generated from within the repeat but instead arises from annealing of the TP primer to a $(CCG)_4CCT$ (SEQ ID NO: 6) sequence immediately upstream of the CCG repeat, which is defined as starting after the CTG trinucleotide (Figure 1). *AFF2* allele sizing and structure are further elaborated in the Results.

Sequencing of the *AFF2* CCG repeat

AFF2 TP-PCR and standard/conventional PCR were performed as described above, except that all primers were unlabeled. TP-PCR and standard/conventional PCR products were purified

using AMPure® beads (Agencourt Bioscience, Beverly, Massachusetts, USA) according to manufacturer's instruction, and quantitated using the Nanodrop™ 1000 Spectrophotometer (Thermo Scientific™, Waltham, Massachusetts, USA). Each 20 µl sequencing reaction contained 10-50 ng of purified standard PCR or TP-PCR product, 1x BigDye® Terminator Ready Reaction Mix (Applied Biosystems), 2.5x Q-Solution (Qiagen), and 3.2 pmol of *AFF2*-F primer. An initial denaturation at 96°C for 1 min was followed by 25 cycles of 98°C for 10 sec, 60°C for 5 sec, and 60°C for 4 min. Extension products were purified using the Oligo Clean & Concentrator™ column (Zymo Research, Irvine, California, USA) according to manufacturer's instruction. Eluted purified extension products were vacuum dried in a Savant SpeedVac® Concentrator (Thermo Scientific) for 5 min, resuspended in 12 µl of Hi-Di™ formamide (Applied Biosystems) and resolved in a 3130xl Genetic Analyzer (Applied Biosystems) using a 36 cm capillary filled with POP-7™ polymer. The mixture was electrokinetically injected at 1.2 kV for 16 sec and electrophoresed for 20 min at 60°C. Post-CE analysis was performed using Sequencing Analysis Software v6.0 (Applied Biosystems).

RESULTS

Evaluation of duplex TP-PCR on normal and *FMR1* and *AFF2* triplet repeat expansion-positive samples

The FRAXE folate-sensitive fragile site on the X chromosome is associated with several medical conditions, including intellectual disability, obsessive-compulsive disorder and premature ovarian insufficiency. FRAXE fragility is caused by hyper-expansion of CCG trinucleotide repeats within the 5' untranslated region (UTR) in exon 1 of the *AFF2* (previously *FMR2*) gene (ClinVar: VCV000010526.1). Hyper-expansion from a normal allele range of 6-30 repeats to >200 repeats is accompanied by CpG methylation of the repeat. This in turn silences expression of AFF2, a subunit of SEC-L2 that regulates the transcription of several genes. *AFF2* CCG repeat hyper-expansion is the genetic mutation which gives fragile XE non-syndromic intellectual disability (FRAXE NSID; OMIM 309548), a mild (IQ 50-70) to borderline (IQ 70-85) intellectual disability affecting an estimated one in 50000 to 100000 males, and also other cognitive/behavioral abnormalities including obsessive compulsive disorder. Paradoxically, alleles with fewer than 11 repeats or with microdeletions within or near the repeats are enriched in premature ovarian insufficiency patients. Studies associating intermediate (31-60) repeat sizes and Parkinson's disease, a neurodegenerative motor system disorder, have been inconclusive.

FRAXE shares similar genetic features with the well-studied fragile site on the X chromosome, FRAXA, which has a clearer demonstrated clinical involvement. The FRAXA site contains CGG repeats within the 5' UTR in exon 1 of the *FMR1* gene. Hyper-expansion of *FMR1* CGG repeats to >200 is accompanied by CpG methylation and *FMR1* gene silencing, which results in fragile X syndrome (FXS; OMIM 309550) (ClinVar: VCV000009972.1). FXS is the most common inherited monogenic cause of intellectual disability, affecting approximately one in 5000 males and one in 4000 to 8000 females. Premutation alleles are associated with many behavioral features, including anxiety, obsessive-compulsive traits and depression. In addition, approximately one-fifth of females who carry a premutation allele (55-200 CGG repeats) are affected with fragile X-related primary ovarian insufficiency (FXPOI). *FMR1* premutations have been identified in 2% of sporadic POI and 14% familial POI patients, making FXPOI the most common genetic cause of POI in euploid women. Furthermore, a subset of *FMR1* premutation carriers will eventually develop fragile X-associated tremor/ataxia syndrome (FXTAS), a neurodegenerative disorder with features of ataxia, tremor and parkinsonism that affects approximately 1 in 4000 men and 1 in 7800 women over the age of 55. The *FMR1* gene has also been associated with Parkinson's disease.

Although both FRAXA and FRAXE fragile sites are strongly associated with intellectual disabilities, the mild-to-borderline phenotype of FRAXE NSID could lead to under-ascertainment and under-diagnosis compared to FXS. Under-diagnosis may be partly attributed to the lack of a rapid, simple and inexpensive assay to screen for repeat expansions at the FRAXE site. Although standard PCR methods have been described for repeat expansions at the FRAXE site, either as a standalone or multiplexed assay, they cannot detect large premutation and full mutation alleles, whose detection continues to rely on Southern blot analysis. Molecular diagnosis is crucial to confirm FRAXE due to its non-syndromic nature, that is, a lack ofdistinctive and consistent clinical presentations, in contrast to e.g. facial dysmorphism (prominent ears, jaw, forehead and long face) and macro-orchidism present in FXS patients.

Twelve DNA samples were initially utilized to evaluate the duplex TP-PCR assay. Figures 2 and 3 present the TP-PCR electropherogram patterns generated from a normal male and female, as well as *FMR1* and *AFF2* premutation (PM) or full mutation (FM) individuals. The leftmost peak generated from the *AFF2* TP-PCR reaction arises not from annealing of the TP primer within the *AFF2* CCG repeat, which starts after the last CTG trinucleotide of the 5' flanking

sequence. Instead this solitary peak, which migrates as an apparently 111 bp fragment, arises from annealing of the TP primer to the $(CCG)_4CCT$ (SEQ ID NO: 6) sequence immediately upstream of the repeat tract (Figure 1). The first true peak arising from annealing of the TP primer to CCGs 1-5 of the *AFF2* repeat appears on the electropherogram as an apparently 138 bp fragment. TP primer annealing occurs at all other positions in the repeat tract containing five consecutive CCGs (CCGs 2-6, 3-7, etc.), with the final fluorescent peak arising from annealing of the TP primer to the last five CCGs of the repeat tract. An *AFF2* allele's repeat size can therefore be rapidly and easily determined by counting the number of peaks generated from within the repeat tract (i.e. excluding the leftmost solitary peak) and adding a four to that number.

Absence of fluorescent peaks beyond 55 repeats indicates absence of expansion, while continuous fluorescent peaks extending beyond 55 repeats and up to 200 repeats indicates presence of a premutation (PM) allele, and fluorescent peaks extending beyond 200 repeats indicates presence of a full mutation (FM). These results indicate that the duplex TP-PCR assay successfully detected and accurately identified the *FMR1* CGG and *AFF2* CCG repeat premutations and full mutations in both male and female DNA samples (Figures 2 and 3).

Population allele distributions and identification of repeat structures

A total of 466 male DNA samples were genotyped using the duplex TP-PCR, comprising 161 Chinese, 158 Malay, 89 Indian, 44 Caucasian and 14 African American DNAs to determine allele size distributions (Figure 4 and Table 2). Different allele distributions were observed among ethnic groups, with modal *AFF2* repeat size of 18 in the Chinese and Malays (32.3% and 27.8% respectively), but 15 in Caucasians (40.9%), African Americans (50.0%) and Indians (32.6%). Allele size ranges were wider in Chinese (5–31) and Malays (6–37) compared to Indians (10–27), Caucasians (9–26) and African Americans (14–28), which could be partially attributed to a larger sample size for Chinese and Malays. Using the conventional categorization, 24 samples have minimal alleles (<11 repeats), 440 samples have normal alleles (11 – 30 repeats) and the remaining 2 samples have intermediate alleles (31 – 60 repeats). No premutation and full mutation alleles were observed. These results are similar to an earlier study on the *AFF2* allele distributions in the mainland Han Chinese (mode 18, range 9–26) and New York City Caucasians (mode 16, range 8–34) (Figure 4A-C).

Chromosomes carrying the combination of 29 *FMR1* CGG repeats and 18 *AFF2* CCG repeats were the most common (Figure 4D, top panel). Of the 466 male population DNA samples that were screened, three *AFF2* TP-PCR electropherogram patterns were observed. The most common pattern, as documented in the Genome Reference Consortium Human Build 38 (GRCh38), was present in 459 samples (98.1%) representing most normal and all intermediate alleles, while the other two patterns were identified by Sanger sequencing to be caused by the presence of two single nucleotide polymorphism (SNP) variants (Figure 4D, bottom panel).

One of the SNP variants, a T>C substitution at chrX:148,500,637 (rs868914124, GRCh38), converts a CTG trinucleotide to a CCG at the 5' start of the *AFF2* CCG repeat. Whereas the *AFF2* trinucleotide repeat in the rs868914124(T) common allele starts after the last CTG trinucleotide of the 5' flanking sequence (Figure 5A), this CTG trinucleotide becomes a CCG trinucleotide in the rs868914124(C) variant allele, thus extending the *AFF2* trinucleotide repeat tract 2 CCGs or 6 bp upstream to start after the last CAG trinucleotide of the 5' flanking sequence (Figure 5B). As a consequence, the first TP-PCR product arising from within the repeat tract of the rs868914124(C) variant allele appears as an apparently 132 bp fluorescent peak (Figure 5B), in contrast to a 138 bp first fluorescent peak from the rs868914124(T) common allele (Figure 5A). In the TP-PCR electropherogram, this size difference shows up as a narrower gap of 20 bp between the leftmost solitary peak and the first peak of the *AFF2* CCG repeat of the rs868914124(C) variant allele (Figure 5B), compared to a wider 26 bp gap for the rs868914124(T) common allele (Figure 5A,C).

The rs868914124(C) variant was observed in eight out of 466 *AFF2* alleles (1.72%) (Figure 4D, bottom panel, row B), of which six were Malay alleles. Three of the eight samples with the rs868914124(C) variant contained minimal alleles (<11 repeats), compared to 21 out of 458 samples carrying the rs868914124(T). Using Fisher's exact test (two-tailed), this rare variant was observed to be enriched in the Malay group (odds ratio = 6.01; 95% confidence interval 1.06–61.7; P = .021) and with minimal alleles (odds ratio = 12.3; confidence interval 1.79–68.4; P = .006). Interestingly, all six *AFF2* expanded alleles (5 full mutations and 1 premutation) carry the rare rs868914124(C) variant, although the expanded alleles of an affected male and his maternal uncle, as well as the expanded alleles from a father-daughter pair, are assumed to be identical by descent (Figure 5D and Figures 6 and 7).

Unlike AGG interruptions within the CGG repeats commonly observed in normal *FMR1* alleles, we observed only one normal *AFF2* allele containing non-CCG interruptions (0.21%) within its repeat stretch, namely a CTG interruption at the fifth repeat position (Figure 4D, bottom panel, row C, and Figure 5C). This interruption is caused by a C>T substitution at chrX:148,500,652 (rs1389911365, GRCh38). Annealing of the TP primer at positions that include this interruption will result in mismatched pairing and failure to generate a PCR product from that position. In the TP-PCR electropherogram, failed PCR at primer mismatch positions appears as gaps of absent peaks (Figure 5C).

Interestingly, novel non-CCG interruptions were also observed within all six *AFF2* expansion-positive samples. Interruptions within the repeat were initially suspected based on TP-PCR electropherogram patterns of four *AFF2* FM males, which showed gaps of absent/missing peaks between peak clusters (Figure 3). Sanger sequencing revealed the presence of one or more non-CCG interruptions within the 5' end of the repeat, bearing the sequence CCTGTGCAG that is identical to a 9 nucleotide stretch immediately 5' upstream of the repeat. The interruptions vary in number from one (Figure 6A) to more than four (Figure 6B). They also vary in location, such as from the 8^{th} to 10^{th} repeat positions (Figure 6A) or from the 6^{th} to 8^{th} positions (Figure 7).

We observed that in the father-daughter pair (Figure 7), the *AFF2* CCG repeat expansion was a full mutation in the father (DNA_25926) and a premutation in the daughter (DNA_3802), indicative of a previously documented contraction upon transmission. Although *AFF2* full mutation alleles of the nephew (FX0229) and maternal uncle (FX0230) (Figure 6) are assumed to be identical by descent, their repeat structures differ by the presence of additional non-CCG interruptions in the maternal uncle. The exact cause or mechanism for this difference has not been investigated.

Blinded validation of duplex TP-PCR assay on previously characterized samples

Eighty-two archived and previously characterized genomic DNAs were included in a blinded validation of the duplex TP-PCR assay. The *AFF2* CCG repeats were sized as described in the Materials and Methods. The duplex TP-PCR assay accurately classified all 40 normal, 17 *FMR1* PM, 23 *FMR1* FM, and 2 *AFF2* FM samples that were included in the test (Table 3).

Table 2. *AFF2* CCG repeat allele frequencies in five populations.

Number of repeats	CH	ML	IN	CAU	AA	Total

5	1	0	0	0	0	1
6	0	1	0	0	0	1
7	2	2	0	0	0	4
8	0	1	0	0	0	1
9	0	0	0	1	0	1
10	10	5	1	0	0	16
11	4	1	0	1	0	6
12	1	0	2	1	0	4
13	3	1	3	1	0	8
14	1	2	1	1	1	6
15	3	10	29	18	7	67
16	6	5	4	4	2	21
17	32	20	5	1	0	58
18	52	44	25	6	1	128
19	19	16	11	5	0	51
20	8	9	4	1	0	22
21	4	14	1	0	0	19
22	3	7	1	2	1	14
23	3	5	1	0	0	9
24	3	3	0	0	1	7
25	1	2	0	1	0	4
26	2	4	0	1	0	7
27	1	2	1	0	0	4
28	0	0	0	0	1	1
29	1	0	0	0	0	1
30	0	3	0	0	0	3
31	1	0	0	0	0	1
32	0	0	0	0	0	0
33	0	0	0	0	0	0
34	0	0	0	0	0	0
35	0	0	0	0	0	0
36	0	0	0	0	0	0
37	0	1	0	0	0	1
Total	161	158	89	44	14	466

CH, Chinese; ML, Malay; IN, Indian; CAU, Caucasian; AA, African American

Table 3. Concordance analysis of samples with or without *FMR1* or *AFF2* repeat expansion.

Sample ID	**Previous determination**	**Duplex TP-PCR of *FMR1* & *AFF2***	**Concordance Y/N**
FX0002	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0006	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0008	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0009	*FMR1* Full Mutation	*FMR1* Full Mutation	Y

FX0010	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0011	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0013	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0014	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0019	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0020	*FMR1* Premutation	*FMR1* Premutation	Y
FX0021	*FMR1* Premutation	*FMR1* Premutation	Y
FX0022	*FMR1* Premutation	*FMR1* Premutation	Y
FX0023	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0054	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0056	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0058	*FMR1* Premutation	*FMR1* Premutation	Y
FX0060	*FMR1* Premutation	*FMR1* Premutation	Y
FX0062	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0063	*FMR1* Premutation	*FMR1* Premutation	Y
FX0111	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0112	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0132	*FMR1* Premutation	*FMR1* Premutation	Y
FX0135	*FMR1* Premutation	*FMR1* Premutation	Y
FX0136	*FMR1* Premutation	*FMR1* Premutation	Y
FX0137	*FMR1* Premutation	*FMR1* Premutation	Y
FX0138	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0139	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0140	*FMR1* Premutation	*FMR1* Premutation	Y
FX0141	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0143	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0144	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0145	*FMR1* Premutation	*FMR1* Premutation	Y
FX0146	*FMR1* Premutation	*FMR1* Premutation	Y
FX0148	*FMR1* Premutation	*FMR1* Premutation	Y
FX0150	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0151	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0152	*FMR1* Premutation	*FMR1* Premutation	Y
FX0153	*FMR1* Full Mutation	*FMR1* Full Mutation	Y
FX0156	*FMR1* Premutation	*FMR1* Premutation	Y
FX0157	*FMR1* Premutation	*FMR1* Premutation	Y
FX0158	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0160	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0162	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0163	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0164	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0165	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y

FX0166	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0167	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0168	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0169	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0170	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0171	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0172	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0173	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0174	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0175	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0192	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0193	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0194	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0196	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0197	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0198	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0199	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0200	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0204	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0206	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0207	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0208	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0211	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0213	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0214	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0215	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0216	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y

FX0219	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0220	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0221	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0224	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0225	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0227	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y
FX0229	*AFF2* Full Mutation	*AFF2* Full Mutation	Y
FX0230	*AFF2* Full Mutation	*AFF2* Full Mutation	Y
FX0231	*FMR1* Expansion Negative	*FMR1 & AFF2* Expansion Negative	Y

Example 2: *Spinocerebellar ataxias (SCA)*

Materials and Methods

Biological samples

Genomic DNA and cell lines were obtained from Coriell Cell Repositories (CCR; Coriell Institute for Medical Research, Camden, NJ, USA). NA06926, NA13536, and NA13537 each carry an expanded *ATXN1* CAG repeat, NA14982 carries an expanded *ATXN2* CAG repeat, GM06151 carries an expanded *ATXN3* CAG repeat, NA03562 carries an expanded *ATXN7* CAG repeat, and NA13716 and NA13717 each carry an expanded *ATN1* CAG repeat. GM16243 carries an expanded *FXN* GAA repeat, GM06907 carries an expanded *FMR1* CGG repeat, and GM05164 and GM06075 each carry an expanded *DMPK* CTG repeat. Two clinical DNA samples, 160920 and 183254, which carry an expanded *CACNA1A* CAG repeat and *PPP2R2B* CAG repeat, respectively, were obtained from KK Women's and Children's Hospital. Sixty genotype-known archival DNA samples were obtained from Siriraj Hospital - Mahidol University and included in a blinded evaluation of assay accuracy.

Heptaplex TP-PCR

Heptaplex TP-PCR was performed in a 25-µl reaction containing 10 ng of genomic DNA, 1.5x Q-Solution (Qiagen, Hilden, Germany), 1x PCR buffer containing 1.5 mmol/L of $MgCl_2$ (Qiagen), deoxyribonucleic triphosphate (dNTP) mix consisting of 0.2 mmol/L each of dATP, dTTP, dCTP, and dGTP (Roche Applied Science, Penzberg, Germany), and 2 units of HotStar Taq DNA polymerase (Qiagen). Eight primers (seven fluorescently labeled locus-specific

primers and a common TP primer, TP-R) were included at their respective optimal working concentrations. Primer sequences, fluorophore tags, and primer concentrations are shown in Tables 4 and 5. Thermal cycling consisted of initial polymerase activation at 95 °C for 15 minutes, followed by 35 cycles of 98 °C for 45 seconds, 60 °C for 1 minute, and 72 °C for 2 minutes, and a final extension at 72 °C for 5 minutes, on the SimpliAmp thermal cycler (Applied Biosystems-Thermo Fisher Scientific, Foster City, CA, USA).

Capillary electrophoresis (CE)

A 1-μL aliquot of the fluorescently labeled TP-PCR product was mixed with 9 μL of Hi-Di™ formamide (Applied Biosystems) and 0.5 μL of GeneScan™ 500 LIZ™ dye size standard (Applied Biosystems) before it was denatured at 95 °C for 5 minutes, rapidly cooled to 4 °C, and resolved in a 3130xl Genetic Analyzer (Applied Biosystems) using a 36-cm capillary filled with POP7™ polymer. Samples were electrokinetically injected at 1 kV for 15 seconds and electrophoresed for 40 minutes at 60 °C. GeneScan analysis was performed using GeneMapper 5.0 software (Applied Biosystems). The amplified product from each locus can be identified by its product size range and peak color. The heptaplex TP-PCR products display electrophoretic peaks of four different colors and can be analyzed together by switching on all four fluorescence detection channels, or separately using one fluorescence detection channel while other channels are switched off.

Results

Detection of expanded CAG repeats by single-tube heptaplex TP-PCR

The spinocerebellar ataxias (SCAs) are neurodegenerative diseases that cause degeneration of the cerebellum and sometimes spinal cord, and are characterized predominantly by ataxic gait, poor hand-eye coordination, and slurred speech. These autosomal dominantly inherited disorders are genetically diverse and can be caused by conventional mutations as well as repeat expansion mutations. As a group, the prevalence of SCA averages 2.7 cases per 100,000 individuals worldwide, with SCA3 being the most common. There are more than 40 genetically distinct SCAs, of which at least 12 are caused by repeat expansion. Among them, SCA1, SCA2, SCA3, SCA6, SCA7, SCA12, and SCA17 are caused by abnormal CAG trinucleotide repeat expansion. Another disorder, dentatorubral pallidoluysian atrophy (DRPLA), which is also caused by CAG repeat expansion, is classified as an SCA due to its heterogeneous and overlapping clinical phenotypes with other SCAs.

SCA1 is caused by CAG repeat expansion in exon 8 of the *ATXN1* gene on chromosome 6p22.3. Non-expanded (normal) *ATXN1* alleles contain 6 to 44 CAT-interrupted CAGs. When uninterrupted, alleles containing 36 to 38 CAGs are mutable normal, which can expand into the pathogenic size range on transmission to the next generation but have not been associated with clinical symptoms themselves, and those containing ≥39 CAGs are pathogenic. SCA2 is caused by CAG repeat expansion in exon 1 of the *ATXN2* gene on chromosome 12q24.12. Normal *ATXN2* alleles contain 14 to 31 pure or CAA-interrupted CAGs, whereas those containing 32 CAGs are intermediate alleles with uncertain clinical significance. Alleles containing 33 to 500 CAGs are pathogenic. SCA3, or Machado-Joseph disease, is caused by CAG repeat expansion in exon 10 of the *ATXN3* gene on chromosome 14q32.12. Normal *ATXN3* alleles contain 12 to 44 CAGs, while those containing 45 to 59 CAGs are expansion-prone intermediate alleles, and those with 60 to 87 CAGs are full-penetrance alleles. SCA6 is caused by CAG repeat expansion in exon 47 of the *CACNA1A* gene on chromosome 19p13.13. Normal *CACNA1A* alleles contain ≤18 CAGs, while full-penetrance alleles contain 20 to 33 CAGs. *CACNA1A* alleles containing 19 CAGs have uncertain clinical significance. SCA7 is caused by CAG repeat expansion in exon 3 of the *ATXN7* gene on chromosome 3p14.1. Normal *ATXN7* alleles contain 7 to 27 CAGs, whereas those containing 28 to 33 CAGs are mutable normal and alleles with ≥34 CAGs are pathogenic. SCA12 is caused by CAG repeat expansion in the 5' region of the *PPP2R2B* gene on chromosome 5q32. Normal *PPP2R2B* alleles contain 7 to 32 CAGs, whereas those containing 51 to 78 CAGs are full-penetrance alleles. DRPLA is caused by CAG repeat expansion in exon 5 of the *ATN1* gene on chromosome 12p13.31. Normal *ATN1* alleles contain 6 to 35 CAGs, whereas those containing ≥48 CAGs are full-penetrance alleles.

The various SCA types often manifest with markedly varied phenotypes and are generally difficult to distinguish by signs and symptoms alone due to extensive clinical overlap and other accompanying non-ataxic phenotypes. Molecular genetic testing is therefore necessary and is the only method to identify the causative mutations to confirm disease status in symptomatic individuals. As it is difficult to target certain genes for testing without known family history or disease-specific symptoms, identification of the causative gene can be time consuming and expensive. The European Molecular Genetics Quality Network has recommended that testing of the five most common SCA types (SCA1, SCA2, SCA3, SCA6, and SCA7) should be minimally offered by all laboratories, while testing of other types will depend on local prevalence.

Detection and sizing of CAG repeats in different SCAs have relied on standard PCR or triplet-primed PCR (TP-PCR), followed by capillary electrophoresis. Repeat sizing by standard PCR often has to be accompanied by low-throughput and labor-intensive Southern blot analysis to confirm the presence and estimate the repeat size of very large alleles when only a single fragment from the normal allele is detected, or by enzymatic digestion to determine the presence of CAT interruptions in expanded *ATXN1* alleles. TP-PCR, first described by Warner et al. [24], has been widely used in the detection of repeat expansions responsible for many repeat expansion disorders regardless of the expansion size. It employs a locus-specific flanking primer, a triplet-primed (TP) primer, and a Tail primer, and is able to detect very large expansions, accurately size all normal alleles and modest expansions, and identify interruptions within a repeat stretch through distinct TP-PCR electropherogram patterns.

As multiple rounds of genetic testing may be required before the causative gene is identified in an SCA patient, simultaneous screening of the trinucleotide repeats of different disease genes can facilitate faster identification of the causative gene, in addition to saving on the cost of multiple tests. We report on the development of a novel single-tube multiplex TP-PCR assay capable of simultaneously screening for expansion mutation at seven gene loci responsible for some of the most common SCAs (SCA1, SCA2, SCA3, SCA6, SCA7, SCA12, and DRPLA). The heptaplex assay employs a differentially labeled locus-specific flanking primer for each repeat locus and a common primer annealing within the CAG

The SCA heptaplex TP-PCR assay utilizes a fluorescently labeled locus-specific flanking primer located upstream of the CAG repeat of each disease gene and a common TP primer that anneals within all the CAG repeats, to enable co-amplification of the CAG repeats from seven different SCA loci in a single reaction tube. Each flanking primer is tagged by one of four fluorophores (Ned, Vic, Fam, or Pet). In addition, the distance of each flanking primer from its respective repeat tract differs such that the electrophoretic peaks generated from normal alleles of one repeat locus do not overlap with the electrophoretic peaks from normal alleles of another repeat locus that is tagged with the same fluorophore (Figure 8). The combined effect of differential fluorophore labeling and positioning of the seven flanking primers is that the TP-PCR products generated from the normal alleles of each repeat locus do not overlap and can be distinguished after capillary electrophoresis. Repeat number is determined by counting the

electrophoretic peaks starting from the left of the electropherogram, with the first peak representing the first five pure CAGs of the repeat tract (Figure 8).

The heptaplex TP-PCR assay was first tested on genotype-known samples to confirm locus-specific annealing of each of the flanking primers and the detection of expanded CAG repeats at each expanded repeat locus. As the TP primer anneals to five continuous CAG repeats, the TP-PCR generates a mixture of fragments each differing by one triplet. The smallest amplified products of all loci contain five CAG triplets, except for *ATXN3* which has 11 triplets due to the presence of CAA in the 3rd and 6th triplet positions and AAG in the 4th triplet position in its wild-type allele (Figure 8A). The electropherogram displays a series of continuous ladder-like peaks (Figure 8B) because the TP primer anneals to multiple positions within the uninterrupted CAG repeat (Figure 8A). Each successive electrophoretic peak represents TP-PCR product that is larger by one repeat or three base pairs, and the repeat size in an allele can be derived from counting the number of peaks. The presence of non-CAG interruptions in the middle of a repeat tract, which may occur in *ATXN1* and *ATXN2* alleles, prevents the TP-primer from annealing efficiently at those positions, resulting in the absence of fluorescent peaks in between peak clusters. Expansion-negative samples produce TP-PCR products generated from the two non-expanded (normal) alleles that lie within the normal repeat size range (Figure 9), whereas expansion-positive samples additionally produce longer TP-PCR products generated from the expanded allele (Figure 10). Samples with repeat sizes extending beyond the upper limit of the normal allele size range at a specific repeat locus would be indicative of an expansion at that repeat locus (Figure 10).

Blinded clinical sample validation

In order to assess the ability of the heptaplex TP-PCR assay to accurately identify expansion at the seven repeat loci in affected patient samples, a blinded analysis was performed on 60 genotype-known archival clinical DNA samples. After TP-PCR and capillary electrophoresis, one fluorescence detection channel was switched on at a time to facilitate separate analysis of amplification products labeled with different fluorophores. This allowed samples showing electrophoretic peaks that extend beyond the upper limit of the normal allele size range at any of the seven SCA loci to be unambiguously identified. The heptaplex TP-PCR correctly detected repeat expansion in the expected SCA repeat locus for all 31 DNA samples that were positive for one of the seven SCA repeat expansions (five SCA1-positive, seven SCA2-positive, twelve SCA3-positive, five SCA6-positive, one SCA7-positive, and one DRPLA-positive)

(Table 6). Singleplex PCR involving flanking primer for the relevant repeat locus and the common TP primer was performed on all screen-positive samples to confirm results (data not shown). No expansion was detected in any of the 29 expansion-negative DNA samples.

Table 4. SCA heptaplex TP-PCR primers and expected TP-PCR product sizes

Gene	Primer	Sequence (5'→3')	Fluorophore
ATXN1	SCA1-F	CAACCCCGTCACCAGTGC (SEQ ID NO: 7)	Ned
ATXN2	SCA2-F	GTGCGAGCCGGTGTATGG (SEQ ID NO: 8)	Vic
ATXN3	SCA3-F	ATGTCTAGATTTCCTAAGATCAGCACTTCC (SEQ ID NO: 9)	Fam
CACNA1A	SCA6-F	TGTAGGGCAGTAGTTCCGTAAGTG (SEQ ID NO: 10)	Vic
ATXN7	SCA7-F	TGTAGGAGCGGAAAGAATGTCGG (SEQ ID NO: 11)	Fam
PPP2R2B	SCA12-F	CGTGGCCCTTAGCTGAGTG (SEQ ID NO: 12)	Pet
ATN1	DRPLA-F	GCCAACAGCAATGCCCATCC (SEQ ID NO: 13)	Pet
-	TP-R	GTTTCGGCGTTACGAGTGGA$(CTG)_5$ (SEQ ID NO: 14)	-

Table 5. SCA heptaplex expected TP-PCR product sizes

Gene	Primer	Genbank ID: Nucleotides	Concentration (μM)	Expected TP-PCR product size
ATXN1	SCA1-F	NG_011571: 438636-438653	0.25	153 bp + (CAG)n
ATXN2	SCA2-F	NG_011572: 5615-5632	0.20	63 bp + (CAG)n
ATXN3	SCA3-F	NG_008198.2: 40371-40400	0.70	237 bp + (CAG)n
CACNA1A	SCA6-F	NG_011569.1: 303403-303426	0.25	181 bp + (CAG)n
ATXN7	SCA7-F	NG_008227.1: 53027-53049	0.30	123 bp + (CAG)n
PPP2R2B	SCA12-F	NG_011570.1: 202465-202483	0.40	268 bp + (CAG)n
ATN1	DRPLA-F	NG_008047.1: 12162-12181	0.30	125 bp + (CAG)n
-	TP-R	-	0.20	-

Table 6: Disease status and CAG repeat size of the 60 genotype known archival DNA samples

Sample ID	Disease status	CAG repeat size
SP001	SCA1-affected	*ATXN1*: 26/55

SP002	SCA3-affected	*ATXN3*: 31/75
SP003		
SP004		
SP005		
SP006	SCA3-affected	*ATXN3*: 28/77
SP007		
SP008		
SP009	DRPLA-affected	*ATN1*: 11/56
SP010	SCA6-affected	*CACNA1A*: 14/25
SP011		
SP012		
SP013	SCA1-affected	*ATXN1*: 32/43
SP014		
SP015		
SP016	SCA1-affected	*ATXN1*: 26/47
SP017	SCA2-affected	*ATXN2*: 22/41
SP018		
SP019	SCA2-affected	*ATXN2*: 22/41
SP020	SCA3-affected	*ATXN3*: 14/73
SP021	SCA3-affected	*ATXN3*: 14/77
SP022		
SP023	SCA7-affected	*ATXN7*: 10/77
SP024		
SP025	SCA2-affected	*ATXN2*: 22/41
SP026	SCA2-affected	*ATXN2*: 22/42
SP027		
SP028		
SP029	SCA6-affected	*CACNA1A*: 11/23
SP030	SCA2-affected	*ATXN2*: 22/43
SP031	SCA3-affected	*ATXN3*: 27/76
SP032	SCA3-affected	*ATXN3*: 14/71
SP033	SCA2-affected	*ATXN2*: 22/39
SP034		
SP035	SCA3-affected	*ATXN3*: 28/79
SP036	SCA3-affected	*ATXN3*: 20/68
SP037		
SP038		
SP039	SCA1-affected	*ATXN1*: 30/49/60
SP040		
SP041		
SP042	SCA3-affected	*ATXN3*: 29/76
SP043		
SP044	SCA6-affected	*CACNA1A*: 13/23
SP045		
SP046	SCA3-affected	*ATXN3*: 14/73
SP047		
SP048		
SP049		
SP050		
SP051	SCA1-affected	*ATXN1*: 26/54
SP052	SCA2-affected	*ATXN2*: 22/39
SP053		
SP054	SCA3-affected	*ATXN3*: 14/78

SP055	SCA6-affected	*CACNA1A*: 13/23
SP056		
SP057	SCA6-affected	*CACNA1A*: 13/23
SP058		
SP059	SCA3-affected	*ATXN3*: 28/79
SP060		

References

Zhong, N., et al., *A survey of FRAXE allele sizes in three populations.* Am J Med Genet, 1996. **64**(2): p. 415-9.

CLAIMS

1. A method of detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject, the method comprising:
i) contacting the nucleic acid sample under amplification conditions for each of the two or more genes with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene wherein the gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence, and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; and
ii) analyzing the amplification products.

2. The method of claim 1, wherein the method comprises analyzing the amplification products using a size separation technique or a sequencing technique.

3. The method of claim 2, wherein the size separation technique is an electrophoresis based technique.

4. The method of any one of claims 1 to 3, wherein the amplification products are separated according to size.

5. The method of any one of claims 1 to 4, wherein the change in the sizes of the amplification products for a gene as compared to a reference indicates the presence of a repeat expansion sequence in the gene.

6. The method of any one of claims 1 to 5, wherein the repeat expansion sequence is a trinucleotide repeat expansion sequence.

7. The method of any one of claims 1 to 6, wherein the trinucleotide repeat sequence is selected from $(CGG)_n$, $(CCG)_n$, $(CAG)_n$, $(CTG)_n$, $(GCC)_n$, $(GGC)_n$, $(GAA)_n$ or $(TTC)_n$, wherein n is from 2 to 200 or more.

8. The method of any one of claims 1 to 7, wherein the common primer comprises a unique 5' tail sequence.

9. The method of claim 8, wherein the method comprises providing a tail primer that binds specifically to the unique 5' tail sequence of the common primer.

10. The method of any one of claims 1 to 9, wherein the common primer binds to a common target sequence comprising or consisting of 5, 6, 7, 8, 9, 10 or more consecutive trinucleotide repeats.

11. The method of any one of claims 1 to 10, wherein the gene-specific primer is labeled.

12. The method of any one of claims 1 to 11, wherein the two or more genes consists of or comprises FMR1 and AFF2.

13. The method of claim 12, wherein the method comprises contacting the nucleic acid sample with:

a) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 1;

b) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 2; and

c) a common primer comprising or consisting of a nucleic acid sequence having at least 90% sequence identity to a nucleic acid sequence of $(CGG)_5$ (SEQ ID NO: 15) or SEQ ID NO: 3.

14. The method of any one of claims 1 to 11, wherein the two or more genes is selected from the group consisting of SCA1, SCA2, SCA3, SCA6, SCA7, SCA12 and DRPLA.

15. The method of claim 14, wherein the method comprises contacting the nucleic acid sample with:

a) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 7;

b) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 8;

c) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 9;

d) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 10;

e) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 11;

f) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 12; and/or

g) a gene-specific primer comprising or consisting of a nucleic acid having at least 90% sequence identity to a nucleic acid sequence of SEQ ID NO: 13, and

h) a common primer comprising or consisting of a nucleic acid sequence having at least 90% sequence identity to a nucleic acid sequence of $(CTG)_5$ (SEQ ID NO: 16) or SEQ ID NO: 14.

16. A kit for detecting the presence or absence of a repeat expansion sequence in two or more genes in a nucleic acid sample obtained from a subject, the kit comprising:

a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and

b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene;

wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene.

17. A composition comprising a nucleic acid sample obtained from a subject, the composition comprising a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises

a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene;

wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene.

18. A method of screening for one or more Multiple Repeat Expansion Disorders in a subject, the method comprises:

i) contacting a nucleic acid sample from the subject under amplification conditions with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene; and

ii) analyzing the amplification products to screen for the one or more Multiple Repeat Expansion Disorders in the subject.

19. The method of claim 18, wherein the one or more Multiple Repeat Expansion Disorders comprise or consist of fragile X syndrome (FXS), fragile X-related primary ovarian insufficiency (FXPOI), fragile X-associated tremor/ataxia syndrome (FXTAS), and fragile XE non-syndromic intellectual disability (FRAXE NSID).

20. The method of claim 18, wherein the one or more Multiple Repeat Expansion Disorders is Spinocerebellar ataxias (SCA) and/or dentatorubral pallidoluysian atrophy (DRPLA).

21. A method of screening for one or more Multiple Repeat Expansion Disorders in a subject and treating the subject, the method comprises:

i) contacting a nucleic acid sample from the subject under amplification conditions with a) a gene-specific primer that binds specifically to a distinct target sequence of a respective gene in two or more genes, wherein each gene comprises a nucleotide repeat sequence and wherein the distinct target sequence is upstream or downstream of the nucleotide repeat sequence in each gene; and b) a common primer that binds to a common target sequence that is shared by the two or more genes, wherein the common target sequence is located within the nucleotide repeat sequence and on the opposite strand that is bound by the gene-specific primer of each gene; wherein the gene-specific primer and the common primer are capable of generating one or more amplification products from each gene

ii) analyzing the amplification products to screen for the one or more Multiple Repeat Expansion Disorders in the subject; and

iii) treating the subject found to have at least one Multiple Repeat Expansion Disorder.