

Published:

- *with international search report (Art. 21(3))*
- *with sequence listing part of description (Rule 5.2(a))*

CENTROMERE ANALYSIS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority to United States provisional patent application serial
5 number 62/422,193, filed November 15, 2016, which is incorporated herein by reference
in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made with government support under grants CA177824 and
10 CA144043 awarded by the National Institutes of Health. The government has certain
rights in the invention.

FIELD

Provided herein is technology relating to the analysis of chromosome centromeres and
15 particularly, but not exclusively, to methods, compositions, kits, and systems for
detecting, identifying, characterizing, and quantifying chromosome centromeres.

BACKGROUND

Centromeres are structures of eukaryotic chromosomes that hold sister chromatids
20 together and ensure proper chromosome segregation during cell division. Functionally,
centromeres are essential for the correct segregation and inheritance of genetic
information by ensuring that each daughter cell receives a copy of each chromosome
during cell division (see, e.g., Jaco (2008) J Cell Bio 181(6): 885-92).

Aneuploidy occurs when an organism has a number of chromosomes that is above
25 or below the normal chromosome number for the species. Errors in meiotic division are
believed to cause of aneuploidy. In particular, it is believed that prematurely separating
centromeres do not attach correctly to spindle fibers and thus cause chromosomal
nondisjunction (Vig (1984) Human Genetics 66(2): 239-243). Common chromosomal
diseases include Down syndrome, Edwards syndrome, and Patau syndrome. Every
30 pregnancy has an associated risk of fetal aneuploidy and associated disease. When
aneuploidy occurs, the resulting chromosomal imbalances have devastating effects such
as death of the fetus or impeded quality of life for the lifetime of those who do not die
prior to birth. Accordingly, aneuploidy causes healthcare burdens.

Prenatal testing is widely recommended and used to screen for chromosomal
35 abnormalities. There are numerous noninvasive and invasive tests available.

Noninvasive prenatal screening includes analysis of maternal serum and nuchal translucency scan. If these screens are positive for abnormalities, subsequently invasive prenatal diagnostic tests are performed. These require invasive procedures such as amniocentesis or chorionic villus sampling, which increase the risk of miscarriage.

5 Diagnostic tests include karyotype, fluorescence in situ hybridization (FISH), and array. Noninvasive tests have a false positive rate of up to 5%. Karyotyping is currently the gold standard for prenatal diagnosis, but carries a risk to the fetus. There is great need for a noninvasive, accurate test to detect, identify, characterize, and quantify chromosome centromeres.

10

SUMMARY

While the importance of centromeres is widely known, centromeres are not well defined at a structural and/or sequence level. For instance, one problem arises due to the presence and abundance of near-identical satellite DNA sequences (termed “alpha-
15 sequences”) at and/or near the centromeres that confound attempts to generate a reliable reference sequence for chromosomes. These problems have limited efforts to understand the relationship between genome sequence and epigenetics that is generally believed to underlie centromere structure, identity, and function (Hayen & Willard (2012) BMC Genomics 13: 324).

20 Accordingly, embodiments of the technology provided herein relate to a method for detecting a human chromosome in a sample. For instance, in some embodiments methods comprise hybridizing a first oligonucleotide and a second oligonucleotide to a nucleic acid comprising a target α -repeat array; producing an amplicon comprising a nucleotide sequence from the target α -repeat array; and detecting the human
25 chromosome in the sample by detecting the amplicon comprising the nucleotide sequence from the target α -repeat array.

In some exemplary embodiments, the human chromosome is chromosome 1 and the target α -repeat array is D1Z5, and/or D1Z7; the human chromosome is chromosome 2 and the target α -repeat array is D2Z1; the human chromosome is chromosome 3 and
30 the target α -repeat array is D3Z1; the human chromosome is chromosome 4 and the target α -repeat array is D4Z1; the human chromosome is chromosome 5 and the target α -repeat array is D5Z1; the human chromosome is chromosome 6 and the target α -repeat array is D6Z1; the human chromosome is chromosome 7 and the target α -repeat array is D7Z1 and/or D7Z2; the human chromosome is chromosome 8 and the target α -
35 repeat array is D8Z2; the human chromosome is chromosome 9 and the target α -repeat

array is D9Z4; the human chromosome is chromosome 10 and the target α -repeat array is D10Z1; the human chromosome is chromosome 11 and the target α -repeat array is D11Z1; the human chromosome is chromosome 12 and the target α -repeat array is D12Z3; the human chromosome is chromosome 13 and the target α -repeat array is D13Z1; the human chromosome is chromosome 14 and the target α -repeat array is D14Z1 and/or D14Z2 and/or D14Z3; the human chromosome is chromosome 15 and the target α -repeat array is D15Z3; the human chromosome is chromosome 16 and the target α -repeat array is D16Z2; the human chromosome is chromosome 17 and the target α -repeat array is D17Z1 and D171b; the human chromosome is chromosome 18 and the target α -repeat array is D18Z1 and/or D18Z2; the human chromosome is chromosome 19 and the target α -repeat array is D19Z4 and/or D19Z5; the human chromosome is chromosome 20 and the target α -repeat array is D20Z2; the human chromosome is chromosome 21 and the target α -repeat array is D21Z1; the human chromosome is chromosome 22 and the target α -repeat array is D22Z4, and/or D22Z5; the human chromosome is chromosome X and the target α -repeat array is DXZ1; and/or the human chromosome is chromosome Y and the target α -repeat array is DYZ3.

The nucleotide sequences for the α -repeat arrays described herein are provided by Figure 9B and/or by SEQ ID NOs: 112 to 151.

Some embodiments provide multiplex technologies (e.g., multiplex methods, compositions for multiplex methods, kits for multiplex methods, e.g., a kit comprising a vessel comprising one or more containers and/or one or more wells comprising one or more compositions for detecting a plurality of chromosomes; a kit comprising a surface comprising an array of probes). For example, in some embodiments the technology relates to a method for detecting two or more human chromosomes in a sample, e.g., a method for detecting two or more human chromosomes in a sample in a single assay. Some embodiments provide methods for detecting 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 chromosomes (e.g., 2 or more autosomes; 2 or more sex chromosomes; 1 or more autosomes and 1 or more sex chromosomes) in a sample, e.g., using a single assay. In some embodiments, multiplex methods comprise use of multiple pairs of amplification oligonucleotides (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 pairs). In some embodiments, multiplex methods comprise use of multiple probes (e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, or 24 probes). In some embodiments, each probe comprises a different detectable label (e.g., each probe comprises a different fluorescent label). In some embodiments, the technology provides a vessel comprising one or more

compositions for detecting a plurality of chromosomes (e.g., a tube, a well of a multiwell plate). In some embodiments, the technology provides a multivessel component (e.g., a multiwell plate, multititre plate, etc., e.g., comprising 6, 24, 96, 384, 1536, 3456, or 9600 wells, e.g., wherein each well has a volume of between approximately one nanoliter to
 5 several milliliters) comprising a plurality of vessels, each of said vessels comprising a composition (e.g., probes and/or primers) for detecting a chromosome or comprising a composition (e.g., a plurality of probes and/or a plurality of primer pairs) for detecting a plurality of chromosomes.

Embodiments provide oligonucleotides (e.g., amplification primers, probes, etc.)
 10 that are specific for target α -repeat arrays. For example, in some embodiments methods comprise use of a first and a second oligonucleotide. In some embodiments, the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID
 15 NO: 7 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 8; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 9 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 10; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 11 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ
 20 ID NO: 12; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 13 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 14; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 15 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 16; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID
 25 NO: 17 or 19 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 18 or 20; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 21 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 22; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 23 and the second oligonucleotide comprises a
 30 nucleotide sequence provided by SEQ ID NO: 24; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 25 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 26; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 27 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 28; the first
 35 oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 29 and the

second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 30; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; the first oligonucleotide
5 comprises a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 54 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 55; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 56
10 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 57; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 58 or 60 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 59 or 61; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 62 or 64 and the second oligonucleotide comprises a nucleotide sequence
15 provided by SEQ ID NO: 63 or 65; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 66 or 68 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 67 or 69; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 70 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 71; the first
20 oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 72 or 74 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 73 or 75; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81; the first oligonucleotide comprises a nucleotide sequence
25 provided by SEQ ID NO: 82 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 83; and/or the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 84 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 85.

In some embodiments, the first oligonucleotide comprises a nucleotide sequence
30 that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 1, 3, or 5 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91,
92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided
35 by SEQ ID NO: 2, 4, or 6; the first oligonucleotide comprises a nucleotide sequence that

35

[illegible]

97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 48, 50, or 52 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 49, 51, or 53; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 54 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 55; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 56 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 57; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 58 or 60 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 59 or 61; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 62 or 64 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 63 or 65; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 66 or 68 and the second oligonucleotide comprises a nucleotide sequence that is at least 80%

identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 67 or 69; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 70 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 71; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 72 or 74 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 73 or 75; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 76, 78, or 80 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 77, 79, or 81; the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 82 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 83; and/or the first oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 84 and the second oligonucleotide comprises a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 85.

The following terms are used to describe the sequence relationships between two or more polynucleotides: “reference sequence”, “identity”, “identical”, “% identity”, “% identical”, “percentage of sequence identity”, and “substantial identity”. A “reference

sequence” is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA sequence given in a sequence listing or may comprise a complete gene sequence. Since two polynucleotides may each (1) comprise a sequence (e.g., a portion of the
5 complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a “comparison window” to identify and compare local regions of sequence similarity. Alignment of sequences for aligning a
10 comparison window may be conducted by the local homology algorithm of Smith and Waterman [Smith and Waterman, Adv. Appl. Math. 2: 482 (1981)] by the homology alignment algorithm of Needleman and Wunsch [Needleman and Wunsch, J. Mol. Biol. 48:443 (1970)], by the search for similarity method of Pearson and Lipman [Pearson and Lipman, Proc. Natl. Acad. Sci. (U.S.A.) 85:2444 (1988)], by computerized
15 implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by inspection, and the best alignment (e.g., resulting in the highest percentage of identity over the comparison window) generated by the various methods is selected. The term “sequence identity” means that two
20 polynucleotide sequences are identical (e.g., on a nucleotide-by-nucleotide basis) over the window of comparison. The term “percentage of sequence identity” or “% identical” and similar terms are calculated by comparing two aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U) occurs in both sequences to yield the number of matched
25 positions, dividing the number of matched positions by the total number of positions in the window of comparison, and multiplying the result by 100 to yield the percentage of sequence identity. The terms “substantial identity” as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, preferably at least 90 to 95
30 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison. The reference sequence may be a subset of a

larger sequence, for example, as a segment of the full-length sequences of the compositions claimed in the present invention.

The term “substantially homologous” when used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone refers to any probe that
5 can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low to high stringency as described above.

The term “substantially homologous” when used in reference to a single-stranded nucleic acid sequence refers to any probe that can hybridize (i.e., it is the complement of) the single-stranded nucleic acid sequence under conditions of low to high stringency as
10 described above.

In some embodiments, the methods comprise technologies for inhibiting (e.g., minimizing, eliminating, etc.) the amplification of one target (e.g., a “non-target”). For example, some embodiments comprise amplifying a target while inhibiting the amplification of a non-target, e.g., to promote, maximize, or otherwise improve the
15 amplification and/or detection of the intended target. In some embodiments, the technology provides methods as described herein and further comprising inhibiting amplification of an amplicon comprising a nucleotide sequence from a non-target α -repeat array. For instance, in some embodiments inhibiting amplification of an amplicon comprising a nucleotide sequence from a non-target α -repeat array comprises
20 hybridizing a clamp oligonucleotide to a nucleic acid comprising the non-target α -repeat array. As used herein, a “clamp oligonucleotide” is an oligonucleotide that binds to a complementary sequence but that does not or cannot act as a primer for polymerase extension (e.g., the clamp does not comprise a 3' OH to provide a substrate for nucleotide addition). In some embodiments, a clamp oligonucleotide comprises a 3' phosphate.

25 In some embodiments, the methods comprise use of modified oligonucleotides, e.g., oligonucleotides modified to modulate (e.g., increase, decrease, alter) the hybridization of the oligonucleotide to its complementary sequence. For example, in some embodiments the first oligonucleotide and/or the second oligonucleotide comprises a locked nucleic acid nucleotide.

30 The technology relates to analysis of both centromeric and pericentromeric regions of chromosomes. Thus, in some embodiments a centromere core comprises the target α -repeat array; in some embodiments, a pericentromeric region comprises the target α -repeat array. In some embodiments, the target α -repeat array is an α -repeat array that is present on a plurality of chromosomes of an organism. In some
35 embodiments, the target α -repeat array is an α -repeat array that is present on all

chromosomes of an organism. For instance, in some embodiments the target α -repeat array is p82H. In some embodiments, the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 86 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 87.

5 Embodiments provide for the evaluation and analysis of the structure of α -repeat arrays. For example in some embodiments, the methods provided herein comprise quantifying amplicons produced from the target α -repeat array to provide a quantity of α -repeat amplicons. Furthermore, some embodiments comprise calculating the relative number of α -repeats in the α -repeat array per human diploid genome by comparing the
10 quantity of α -repeat amplicons to the quantity of a single-copy gene. Some embodiments comprise calculating the ploidy of the human chromosome in the sample by comparing the relative number of α -repeats in the α -repeat array per human diploid genome calculated for the sample to a relative number of α -repeats in the α -repeat array per human diploid genome for a normal sample. In some embodiments, the relative number
15 of α -repeats in the α -repeat array per human diploid genome for a normal sample is a previously determined known value. In some embodiments, the relative number of α -repeats in the α -repeat array per human diploid genome for a normal sample is determined experimentally using a sample from a normal human. The technology finds use, e.g., to evaluate the ploidy of an organism. For example, in some embodiments the
20 ploidy of the human chromosome indicates that the human source of the sample has an aneuploidy. In some embodiments, the ploidy of the human chromosome indicates that the human source of the sample is male. In some embodiments, the ploidy of the human chromosome indicates that the human source of the sample is female. In particular embodiments, the technology detects an aneuploidy that is a trisomy. In particular
25 embodiments, the technology detects an aneuploidy that is trisomy 21, trisomy 13, trisomy 8, trisomy 18, or trisomy X.

 Accordingly, in some embodiments, the technology provides a method for detecting an aneuploidy of a chromosome in a subject. In some embodiments, methods comprise determining a pericentromere size of a chromosome; comparing the
30 pericentromere size of the chromosome to the pericentromere size for the chromosome of a normal subject; and identifying the chromosome as aneuploid by determining that the pericentromere size of the chromosome of the subject is smaller than the pericentromere size for the chromosome of a normal subject. In some embodiments, determining a pericentromere size of a chromosome comprises quantifying copy number of K111 and/or
35 K222. In some embodiments, determining a pericentromere size of a chromosome

comprises amplifying K111 using a first oligonucleotide comprising a sequence provided by SEQ ID NO: 98 or 104 and a second oligonucleotide comprising a sequence provided by SEQ ID NO: 99 or 105. In some embodiments, determining a pericentromere size of a chromosome comprises amplifying K222 using a first oligonucleotide comprising a sequence provided by SEQ ID NO: 101 and a second oligonucleotide comprising a sequence provided by SEQ ID NO: 102.

Some embodiments provide a method for determining the number of α -repeats in an α -repeat array of a chromosome. For example, in some embodiments methods comprise hybridizing a first oligonucleotide and a second oligonucleotide to a nucleic acid comprising a target α -repeat array; producing an amplicon comprising a nucleotide sequence from the target α -repeat array; quantifying the amplicon; and calculating the number of α -repeats in the α -repeat array of the chromosome from the quantity of the amplicon.

In specific embodiments, the human chromosome is chromosome 1 and the target α -repeat array is D1Z5, and/or D1Z7; the human chromosome is chromosome 2 and the target α -repeat array is D2Z1; the human chromosome is chromosome 3 and the target α -repeat array is D3Z1; the human chromosome is chromosome 4 and the target α -repeat array is D4Z1; the human chromosome is chromosome 5 and the target α -repeat array is D5Z1; the human chromosome is chromosome 6 and the target α -repeat array is D6Z1; the human chromosome is chromosome 7 and the target α -repeat array is D7Z1 and/or D7Z2; the human chromosome is chromosome 8 and the target α -repeat array is D8Z2; the human chromosome is chromosome 9 and the target α -repeat array is D9Z4; the human chromosome is chromosome 10 and the target α -repeat array is D10Z1; the human chromosome is chromosome 11 and the target α -repeat array is D11Z1; the human chromosome is chromosome 12 and the target α -repeat array is D12Z3; the human chromosome is chromosome 13 and the target α -repeat array is D13Z1; the human chromosome is chromosome 14 and the target α -repeat array is D14Z1 and/or D14Z2 and/or D14Z3; the human chromosome is chromosome 15 and the target α -repeat array is D15Z3; the human chromosome is chromosome 16 and the target α -repeat array is D16Z2; the human chromosome is chromosome 17 and the target α -repeat array is D17Z1 and/or D171b; the human chromosome is chromosome 18 and the target α -repeat array is D18Z1 and/or D18Z2; the human chromosome is chromosome 19 and the target α -repeat array is D19Z4 and/or D19Z5; the human chromosome is chromosome 20 and the target α -repeat array is D20Z2; the human chromosome is chromosome 21 and the target α -repeat array is D21Z1; the human chromosome is chromosome 22 and the

target α -repeat array is D22Z4, and/or D22Z5; the human chromosome is chromosome X and the target α -repeat array is DXZ1; and/or the human chromosome is chromosome Y and the target α -repeat array is DYZ3.

Particular oligonucleotides find use in some embodiments. For example, in some
5 embodiments the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 7 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 8; the first oligonucleotide comprises a
10 nucleotide sequence provided by SEQ ID NO: 9 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 10; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 11 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 12; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 13 and the second
15 oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 14; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 15 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 16; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 17 or 19 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO:
20 18 or 20; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 21 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 22; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 23 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 24; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID
25 NO: 25 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 26; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 27 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 28; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 29 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ
30 ID NO: 30; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO:
35 49, 51, or 53; the first oligonucleotide comprises a nucleotide sequence provided by SEQ

ID NO: 54 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 55; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 56 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 57; the first oligonucleotide comprises a nucleotide sequence provided by
 5 SEQ ID NO: 58 or 60 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 59 or 61; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 62 or 64 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 63 or 65; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 66 or 68 and the second
 10 oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 67 or 69; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 70 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 71; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 72 or 74 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID
 15 NO: 73 or 75; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81; the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 82 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 83; and/or the first oligonucleotide
 20 comprises a nucleotide sequence provided by SEQ ID NO: 84 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 85.

Related embodiments provide compositions, e.g., compositions for producing an amplicon from a human chromosome (e.g., from an α -repeat array). Some embodiments provide a composition comprising one or more of a first oligonucleotide comprising a
 25 nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 7 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 8; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 9 and a
 30 second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 10; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 11 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 12; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 13 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 14; a
 35 first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 15 and a

second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 16; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 17 or 19 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 18 or 20; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 21 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 22; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 23 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 24; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 25 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 26; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 27 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 28; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 29 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 30; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 54 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 55; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 56 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 57; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 58 or 60 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 59 or 61; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 62 or 64 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 63 or 65; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 66 or 68 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 67 or 69; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 70 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 71; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 72 or 74 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 73 or 75; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO:

77, 79, or 81; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 82 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 83; and/or a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 84 and a second oligonucleotide comprising a nucleotide sequence
5 provided by SEQ ID NO: 85.

In some embodiments, the technology provides reaction mixtures (e.g., amplification reaction mixtures, probe hybridization reaction mixtures, real-time PCR reaction mixtures, probe hydrolysis ("TAQMAN"-type) reaction mixtures, etc.). For example, some embodiments provide a reaction mixture comprising a nucleic acid target
10 (e.g., a nucleic acid comprising a centromeric or pericentromeric region and/or a centromeric or pericentromeric sequence, e.g., a nucleic acid comprising an α -repeat array) and one or more amplification oligonucleotides (e.g., one or more amplification oligonucleotides complementary to a nucleic acid comprising a target α -repeat array, e.g., complementary to a portion of a nucleic acid flanking an α -repeat array). Related
15 embodiments of reaction mixtures comprise a polymerase. In further related embodiments, reaction mixtures comprise one or more deoxynucleotide monomers (e.g., dNTPs, e.g., dATP, dCTP, dGTP, dTTP, and/or one or more modified deoxynucleotide monomers), for incorporation into an amplicon during an amplification reaction. In some
20 embodiments, reaction mixtures comprise one or more buffers, cations (e.g., polymerase cofactors), salts, or other amplification reaction reagents. In reaction mixtures related to amplifying a centromeric or pericentromeric region, the reaction mixtures comprise, e.g., a target nucleic acid, oligonucleotide primers, a polymerase, dNTPs, and, in some
25 embodiments, an amplicon produced from the target nucleic acid, e.g., an amplicon comprising a centromeric and/or pericentromeric region and/or sequence, e.g., by an amplification reaction using the oligonucleotide primers, polymerase, and dNTPs. In some embodiments, one or more amplification oligonucleotides comprises a label (e.g., a covalently attached label, e.g., a fluorescent label); in related embodiments, a reaction mixture comprises one or more amplification oligonucleotides comprises a label and/or
30 an amplicon that incorporates one or more amplification oligonucleotides comprising a label.

In some embodiments, a reaction mixture comprises a nucleic acid target (e.g., a nucleic acid comprising a centromeric or pericentromeric region and/or a centromeric or pericentromeric sequence, e.g., a nucleic acid comprising an α -repeat array) and one or more oligonucleotide probes (e.g., one or more oligonucleotide probes complementary to a
35 nucleic acid comprising a target α -repeat array, e.g., complementary to a portion of a

nucleic acid flanking an α -repeat array). In some embodiments, reaction mixtures comprise an oligonucleotide probe comprising a label (e.g., an oligonucleotide probe comprising a covalently attached label, e.g., a fluorescent label). In some embodiments, reaction mixtures comprise a plurality of oligonucleotide probes, wherein each
 5 oligonucleotide probe of the plurality of oligonucleotide probes comprises a fluorescent label that is distinguishable from one or more fluorescent labels of one or more other oligonucleotide probes.

Some embodiments of reaction mixtures comprise both oligonucleotide primers (e.g., for producing an amplicon) and one or more oligonucleotide probes (e.g., for
 10 detecting an amplicon).

In some embodiments, a reaction mixture comprises a nucleic acid comprising a sequence from a human chromosome, e.g., a sequence from chromosome 1, e.g., a target α -repeat array that is D1Z5, and/or D1Z7; a sequence from chromosome 2, e.g., a target α -repeat array that is D2Z1; a sequence from chromosome 3, e.g., a target α -repeat array
 15 that is D3Z1; a sequence from chromosome 4, e.g., a target α -repeat array that is D4Z1; a sequence from chromosome 5, e.g., a target α -repeat array that is D5Z1; a sequence from chromosome 6, e.g., a target α -repeat array that is D6Z1; a sequence from chromosome 7, e.g., a target α -repeat array that is D7Z1 and/or D7Z2; a sequence from chromosome 8, e.g., a target α -repeat array that is D8Z2; a sequence from chromosome
 20 9, e.g., a target α -repeat array that is D9Z4; a sequence from chromosome 10, e.g., a target α -repeat array that is D10Z1; a sequence from chromosome 11, e.g., a target α -repeat array that is D11Z1; a sequence from chromosome 12, e.g., a target α -repeat array that is D12Z3; a sequence from chromosome 13, e.g., a target α -repeat array that is D13Z1; a sequence from chromosome 14, e.g., a target α -repeat array that is D14Z1
 25 and/or D14Z2 and/or D14Z3; a sequence from chromosome 15, e.g., a target α -repeat array that is D15Z3; a sequence from chromosome 16, e.g., a target α -repeat array that is D16Z2; a sequence from chromosome 17, e.g., a target α -repeat array that is D17Z1 and/or D171b; a sequence from chromosome 18, e.g., a target α -repeat array that is D18Z1 and/or D18Z2; a sequence from chromosome 19, e.g., a target α -repeat array that
 30 is D19Z4 and/or D19Z5; a sequence from chromosome 20, e.g., a target α -repeat array that is D20Z2; a sequence from chromosome 21, e.g., a α -repeat array that is D21Z1; a sequence from chromosome 22, e.g., a target α -repeat array that is D22Z4-1, and/or D22Z5; a sequence from chromosome X, e.g., a target α -repeat array that is DXZ1; and/or a sequence from chromosome Y, e.g., a target α -repeat array that is DYZ3.

In some embodiments, the compositions and methods further employ control reagents or kit components (e.g., positive controls, negative controls). In some embodiments, the control reagents include a synthetic target nucleic acid. In some embodiments, the control reagents include reagents for detecting a target nucleic acid sequence, e.g., a chromosome sequence expected to be present in a sample (e.g., a centromeric, pericentromeric, α -repeat sequence). In some embodiments, a control target nucleic acid, whether synthetic or endogenous in a sample, is selected such that amplification primers that amplify the target nucleic acid also amplify the control target nucleic acid. In some such embodiments, a probe that detects the target nucleic acid or an amplicon generated therefrom does not detect the control target or an amplicon generated therefrom. In some embodiments, a control probe is provided that detects the control target nucleic acid or an amplicon generated therefrom but does not detect the target nucleic acid or an amplicon generated therefrom. In some embodiments, internal standards are provided for quantification.

In some embodiments, the technology provides a reaction mixture comprising any composition described herein or a composition comprising a target nucleic acid, e.g., a chromosome (e.g., a centromeric, pericentromeric, α -repeat sequence) and any composition described herein.

In related embodiments, the technology provides a composition (e.g., in some embodiments, a reaction mixture, a reaction mixture intermediate (e.g., a composition for producing a reaction mixture), a kit component, etc.) comprising a pair of oligonucleotides specific for one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3. Consequently, in some embodiments compositions comprise a target of the oligonucleotides described, e.g., in some embodiments the compositions comprise a nucleic acid comprising a sequence from an α -repeat array from a human chromosome, e.g., a nucleic acid comprising a sequence from one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3. Still further embodiments further provide reaction mixtures for producing an amplicon from a target α -repeat array, e.g., in some embodiments the compositions provided further comprise a

polymerase. The technology encompasses embodiments in which compositions comprise one or more products of a reaction mixture, e.g., in some embodiments, compositions comprise an amplicon produced from one or more α -repeat arrays.

Some embodiments provide one or more oligonucleotide probes (e.g., one or more
 5 detectably labeled oligonucleotide probes) that hybridize specifically to one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3.

Related embodiments provide kits for producing an amplicon from a human
 10 chromosome (e.g., from an α -repeat array). For example, some embodiments of kits comprise one or more of a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6; a first oligonucleotide comprising a
 15 nucleotide sequence provided by SEQ ID NO: 7 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 8; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 9 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 10; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 11 and a second oligonucleotide
 20 comprising a nucleotide sequence provided by SEQ ID NO: 12; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 13 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 14; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 15 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 16; a
 25 first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 17 or 19 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 18 or 20; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 21 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 22; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID
 30 NO: 23 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 24; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 25 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 26; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 27 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ
 35 ID NO: 28; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID

NO: 29 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 30; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; a first oligonucleotide
 5 comprising a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 54 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 55; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 56 and a
 10 second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 57; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 58 or 60 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 59 or 61; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 62 or 64 and a second oligonucleotide comprising a nucleotide sequence provided by
 15 SEQ ID NO: 63 or 65; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 66 or 68 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 67 or 69; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 70 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 71; a first oligonucleotide comprising a
 20 nucleotide sequence provided by SEQ ID NO: 72 or 74 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 73 or 75; a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81; a first oligonucleotide comprising a nucleotide sequence provided by SEQ
 25 ID NO: 82 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 83; and/or a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 84 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 85.

In some embodiments, a kit comprises a first oligonucleotide comprising a
 30 nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 1, 3, or 5 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the
 35 sequence provided by SEQ ID NO: 2, 4, or 6; a first oligonucleotide comprising a

[illegible]

oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 21 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 22; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 23 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 24; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 25 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 26; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 27 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 28; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 29 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 30; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96,

97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided
5 by SEQ ID NO: 48, 50, or 52 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 49, 51, or 53; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89,
10 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 54 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 55; a first oligonucleotide comprising a nucleotide sequence
15 that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 56 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by
20 SEQ ID NO: 57; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 58 or 60 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96,
25 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 59 or 61; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 62 or 64 and a second oligonucleotide comprising a nucleotide sequence that is at least 80%
30 identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 63 or 65; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 66 or 68 and
35 a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical

(e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 67 or 69; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 70 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 71; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 72 or 74 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 73 or 75; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 76, 78, or 80 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 77, 79, or 81; a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 82 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 83; and/or a first oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 84 and a second oligonucleotide comprising a nucleotide sequence that is at least 80% identical (e.g., at least 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.2, 99.5, 99.9, or 100% identical) to the sequence provided by SEQ ID NO: 85.

Related embodiments provide a kit comprising one or more pairs of oligonucleotides specific for one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1;

D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3.

Some embodiments of kits provide one or more oligonucleotide probes (e.g., one or more detectably labeled oligonucleotide probes) that hybridize specifically to one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3.

Additional embodiments will be apparent to persons skilled in the relevant art based on the teachings contained herein.

BRIEF DESCRIPTION OF THE DRAWINGS

These and other features, aspects, and advantages of the present technology will become better understood with regard to the following drawings:

Figure 1 shows detection of centromere α -repeat arrays in individual human chromosomes. Representative α -repeat arrays in each human chromosome (Y Axis) were detected and the number of repeats quantified by qPCR using specific primers. The image shows the results of gel electrophoresis of qPCR products amplified from DNA of human/rodent hybrid cells, with each hybrid cell containing only one human chromosome (displayed on the X-axis). DNA from rodent parental mouse or hamster cells is included to control for cross-species hybridization of repeats and human DNA isolated from peripheral blood lymphocytes is a positive control. By convention, the nomenclature of these α -repeat arrays begins with the letter D, followed by the chromosome number (1-22, X or Y), then followed by a Z and a number indicating the historical order in which the sequence was discovered. Using the primers and qPCR conditions described herein in the examples and in Tables 1 and 2 (see also Fig. 2), specific centromeric α -repeat arrays were identified for each human chromosome (e.g., D2Z1, D3Z1, D4Z1, etc.). Certain α -repeat arrays were found in two or more chromosomes (e.g., D1Z7/D5Z2 in chromosomes 1 and 5; D13Z1/D15Z1/D21Z1 in chromosomes 13, 15, and 21; D14Z1/D22Z1 in chromosomes 14 and 22; and D19Z4/D21Z2 and D19Z5/D21Z3 in chromosomes 19 and 21). Primers specific for the ubiquitous α -repeat p82H amplified centromeres from all human chromosomes. Assays for the D13Z1 and D21Z1 arrays in Fig. 1 did not use LNA primers as described herein and as shown in used in Figures 10 and 12).

Figure 2 shows detection of recently identified centromeric arrays that had been previously assigned to chromosomes 13, 14, 21, and 22. Representative arrays identified in the most recent human genome assembly (Hg38) (Miga et al., 2014; Miga, 2015) and temporarily assigned to centromeres 13, 14, 21, and 22 (Y Axis) were assessed by PCR assays to verify chromosomal identity and determine if these arrays provide markers of a given chromosome (see also Fig. 1). The number of repeats of each array was quantified by qPCR using specific primers. The figure shows an image from a gel electrophoresis experiment in which qPCR products were amplified from DNA of human/rodent hybrid cells, with each hybrid cell containing only one human chromosome (displayed on the X-axis). DNA from rodent parental mouse or hamster cells is included to control for cross-species hybridization of repeats and human DNA isolated from peripheral blood lymphocytes is a positive control. These arrays had not been named using the original nomenclature for centromere arrays. Accordingly, these arrays were identified using a designation comprising the letter D, followed by the chromosome number (13, 14, 21 or 22, reassigned after our PCR validation), followed by a Z and a number indicating the order in which the original sequence was discovered. Using the primers and qPCR conditions outlined herein (Tables 1 and 2), it was found that some of these centromeric arrays exist in two or more of the homologue centromeres 13, 14, 21, and 22 (e.g., D13Z1-D13Z9) and sometimes the arrays are found exclusively in only one centromere (e.g., D14Z2 in chromosome 14; and D22Z4 and D22Z5 in chromosome 22.). The accession numbers of these arrays are provided in Table 2.

Figure 3 shows variation in the size of α -repeat arrays in individual humans. Figure 3 is a heat map representing the abundance of α -repeats in each centromeric array and the number of pericentric retroviruses K111 and K222 (right, Y-axis) assessed by qPCR of 50 ng of DNA obtained from the human peripheral blood lymphocytes (PBLs) of five individuals (X-axis). D13Z1 and D21Z1 assays were not performed using conditions for LNA primers. The chromosomal location of each centromeric repeat is indicated on the right. No significant gene variation was seen in the single copy *top3A*, *ccr5*, *dek*, or *β -actin* genes. The α -repeat p82H is present in the centromeres of all human chromosomes. The retroviruses K111 and K222 are present in the pericentromeres of fifteen and nine human chromosomes, respectively. The intensity of the heat map is depicted by the log base 2 Z-score of each α -repeat shown in the grayscale gradient on the upper left left. Hierarchical clustering trees were constructed to represent the content and/or size of every α -repeat array (tree). The tree splits into two main branches, one indicating arrays in centromere “cores” and a second with

multiple-copy K111 and K222 sequences represented in higher abundance along with “pericentromere” arrays. Black indicates higher copy numbers and white indicates lower copy numbers or single copy genes. Note that K111 clusters with the centromeric arrays. Without being limited to a theory or mechanism it is contemplated that this observation results from the high copy number of K111 or that it resides at the proximal pericentromere border.

Figure 4 shows that limited variation in the abundance of centromeric arrays is seen in populations of diverse ethnicity derived from the 1000 genomes project. The figure shows a heat map representing the abundance of α -repeats in centromeric arrays, pericentromeric proviruses K111 and K222, and single copy genes (right Y-axis) determined in silico by BLAST analysis of Illumina sequences from diverse populations (X-axis) (see also Fig. 5). The $\log_2$ Z-score of each array is depicted by a gradient grey scale on the left of the heat map. High and low copy numbers are thus indicated by black and white, respectively. The analysis was performed by obtaining the number of sequence reads in each population that are matched by BLAST to approximately 100 bp of query sequence, allowing no more than 10 base pair mismatches and indels. Hierarchical clustering was performed to differentiate between clades of repeats based on abundance (left tree). The tree split into two main branches, one indicating arrays in centromeric “cores,” present in higher abundance, and “pericentromere” arrays that include K111 and K222 proviruses. The latter branch further splits into a branch representing single copy genes. No significant gene variation was seen in the single copy *top3a*, *ccr5*, *dek*, or *β -actin* genes. Surprisingly, BLAST analysis did not retrieve significant matches for p82H, a repeat that is present in the centromeres of all human chromosomes. Without being limited to theory or mechanism, it is contemplated that this observation results from p82H sequences having more mutations than the number allowed under the screening criteria used for the analysis. Analysis of pericentromere proviruses was only performed on the 5' integration site area specific to K111/K222 to avoid sequence hits from other endogenous proviruses of the same HERV-K family that resemble the K111/K222 proviral genome by more than 95%. ACB: African Caribbean in Barbados; ASW: African Ancestry in Southwest US; CEU: European from Utah; CHS: Southern Han Chinese, China; CLM: Colombian in Medellin, Colombia; GBR: British in England and Scotland; GWD: Gambian in Western Division, The Gambia; GIH: Gujarati Indian in Houston, TX; IBS: Iberian populations in Spain; JPT: Japanese in Tokyo, Japan; KHV: Kinh in Ho Chi Minh City, Vietnam; LWK: Luhya in Webuye, Kenya;

MXL: Mexican Ancestry in Los Angeles, California; PEL: Peruvian in Lima, Peru; PUR: Puerto Rican in Puerto Rico; TSI: Toscani in Italy; and YRI: Yoruba in Ibadan, Nigeria.

Figure 5 shows the positive correlation of copy number in each specific α -repeat centromeric array as determined by qPCR assays, by in silico analysis of the 1000
 5 genomes project, and by Southern blotting hybridization. Figure 5A is a bar diagram representing the average log copy number of α -repeats in each centromeric array, pericentromeric proviruses K111 and K222, and single copy genes per diploid genome determined by either qPCR or by in silico analysis of the 1000 genomes project. The average values are shown in Table 3. Figure B shows the correlation of α -repeat copy
 10 number in each array, proviruses K111 and K222, and single copy genes as determined by qPCR and our bioinformatics analysis (see also Fig. 3 and 4). The Spearman's correlation coefficient and the p value are shown. A statistically significant positive correlation was found between the two detection methodologies. A discordant correlation was found in the number of K111 and p82H copies, indicating that these sequences were
 15 detected using PCR assays but not with bioinformatics analysis. K111 copies were determined by qPCR using a probe that targets a 6 bp non-continuous mutation in the *env* gene present in K111 but not in other HERV-K proviruses. BLAST analysis of K111 *env* has limited utility using the parameters described, as it will retrieve sequence reads from other HERV-Ks. The BLAST analysis was thus performed at the 5' integration site
 20 sequence of K111 proviruses. The centromere repeat p82H, present in all human centromeres, was reliably detected by qPCR assays, and confirmed by Sanger sequencing analysis of the PCR products. In contrast to the PCR assays, in silico analysis was unable to retrieve all p82H sequences, either due to the stringency of the analytic parameters or technical limitations relating to sequencing these loci. Figure C
 25 shows the correlation of the copy number of α -repeats in each array determined by the qPCR assays to the estimated number reported in the literature. The plot shows the correlation of the copy numbers for α -repeats in each centromeric array, proviruses K111 and K222, and single copy genes determined by qPCR with the estimated size reported in the literature as obtained by Southern blotting. The average values are
 30 shown in Table 3. The Spearman's correlation coefficient and the p value are shown. A statistically significant positive correlation was found in the number of α -repeats calculated by qPCR and the number of repeats reported in the literature.

Figure 6 shows that variations in the content of centromere repeats in somatic and sex chromosomes identify ploidy. The figure shows a heat map representing the
 35 abundance of the somatic centromere arrays D8Z2, D18Z1, D18Z2 (in chromosomes 8

and 18, respectively) and the α -repeat arrays in the sex chromosomes, DXZ1 and DYZ3 (X axis), in DNA from individuals of different gender, including individuals with somatic trisomy 8 and trisomy 18 (Y axis). The phenotypic gender of the individuals is represented by medium grey or black bars on the right, with black representing males and medium grey representing females. The gradient bar on the right indicates the color intensity of the heat map, with overrepresented repeats in black and underrepresented repeats in white. An asterisk (*) indicates a significant difference in the content of α -repeats; as expected, all subjects have at least one sex chromosome that is highly represented. Trisomy 18/X Subject B is an individual with trisomies 18 and X. The XY female is a subject who is phenotypically female but genetically male. As expected, DYZ3 from chromosome Y was detected only in the male population (black bands with stars), except for the XY female. The number of α -repeats of the chromosome X DXZ1 was higher in females than in males and much higher in an individual with trisomy 18 and X. Individuals with trisomy 8 have significantly higher content of D8Z2 repeats, whereas individuals with trisomy 18 have significantly higher content of D18Z1 and D18Z2 repeats. Thus, the centromeric PCR assays quickly and accurately detect ploidy for these chromosomes.

Figure 7 is a heat map showing RNA expression of α -repeats of centromere arrays. RNA expression was quantified in the prostate epithelial cell lines RWPE-1, PNT2, and 957E-hTERT, and the prostate cancer cell lines DU145, LnCaP, PC3, and VCaP using the qPCR assays described herein, but including an extra RT step. RNA isolated from these cells was treated with DNase to eliminate DNA contamination. A PCR reaction without the RT step confirmed the elimination of genomic DNA. Quantitative real-time RT-PCR analysis indicated that specific centromeric α -repeat arrays are transcribed at varying levels. Differences in expression are indicated by log₂ values of α -repeat transcript content.

Figure 8 shows the functional capacity of α -repeat arrays in recruiting the centromere proteins CENP-A and CENP-B as assessed by PCR-based assays of centromeric DNA. ChIP was performed on LnCaP prostate cancer cells using CENP-A and CENP-B antibodies or control mouse IgG antibody. The occupancy of each centromeric protein on each array was measured by qPCR with specific primers and compared to the input chromatin that provided a measure of the full-length of each array. No enrichment of *gapdh*, *top3a*, *β -actin*, and certain α -repeat arrays, was observed. Figure 8A shows the occupancy of CENP-A on specific centromere arrays. At least one array in the centromere of each human chromosome recruited CENP-A. In

centromeres with two or more major arrays, CENP-A was found bound to the larger array (e.g., D1Z7 was preferred relative to D1Z5 in chromosome 1; D7Z1 was preferred relative to D7Z2 in chromosome 7; D14Z1 was preferred relative to D14Z2 in chromosome 14; D17Z1 was preferred relative to D17Z1b in chromosome 17; and D18Z1 was preferred relative to D18Z2 in chromosome 18). Little or no significant binding of CENP-A to pericentromeric arrays was seen (e.g., D1Z5, D19Z4, D19Z5, D22Z4, and D22Z5). CENP-A immunoprecipitation in LnCaP using the ChIP conditions indicated that when this protein binds to a given α -repeat array, it occupies between 10% to 40% of the arrays: CENP-A occupied approximately 70% of the array DYZ3 in chromosome Y.

Asterisks indicate dominant arrays in each chromosome that recruit CENP-A. D14Z1/D22Z1 arrays dominate the recruitment of CENP-A to centromeres 14 and 22. The array D19Z3, which resembles D1Z7, is likely recruiting CENP-A to centromere 19. Figure 8B shows the occupancy of CENP-B on centromeric arrays. At least one array in the centromere of each human chromosome (other than Y) recruited CENP-B. Four types of CENP-B boxes (Table 4 and Fig. 9) were found in arrays that recruited CENP-B. However, arrays that did not recruit CENP-B did not have intact CENP-B box sequences. CENP-B did not bind the array DYZ3 in chromosome Y. CENP-B immunoprecipitation in LnCaP indicated that CENP-B occupies approximately 100% of several arrays. Furthermore, CENP-B immunoprecipitation in LnCaP indicated that CENP-B occupancy decreases when some CENP-B boxes are mutated and that CENP-B occupancy is absent in arrays lacking intact CENP-B boxes (e.g., D7Z2). Asterisks indicate arrays that contain CENP-B box sequences.

Figure 9 shows the identification of CENP-B boxes in human centromeric array sequences. Figure 9A shows the sequences of CENP-B boxes. Figure 9B indicates the CENP-B boxes identified in centromere array sequences. In Figure 9B, CENP-B boxes are indicated by the following delimiters that flank the CENP-B box sequences: CENP-B Box 1 (slashes, “/”), CENP-B box 2 (parentheses, “()”), CENP-B box 3 (brackets, “[]”), and CENP-B box 4 (braces, “{ }”) (see also Fig. 9A). The nucleotides in bold black promote (e.g., are necessary for) CENP-B binding. Boxes having mutations in the bolded bases are not able to recruit CENP-B. These sequences are delimited by asterisks, “* *”.

Figure 10 shows that PCR using locked nucleic acid primers discriminates between centromeres 13 and 21, which have similar sequences and specific nucleotide substitutions in α -repeats. The data indicated that nucleotide substitutions present in the homologue arrays D13Z1/D21Z1 from centromere 13 or 21, respectively, provide for accurately detecting the abundance of repeats in each of these nearly identical

centromeres. PCR assays were developed to detect either centromere 13 or 21 specifically. Figure 10A shows that the centromere arrays D13Z1/D21Z1 present in chromosomes 13 and 21 are nearly identical, except for two single nucleotide substitutions present in either centromere 13 or 21 (bases indicated by italic characters).

5 Primers modified with locked nucleic acids (LNAs) were designed to target the underlined sequences, with LNA modification being present at the nucleotide substitution and one base before and one base after it. For the centromere 13 (D13Z1)-specific variant, the PCR reaction comprises a forward primer with LNA modification; the reverse primer used was not modified (bottom sequence in lowercase letters). Figure

10 10B shows the specific detection of chromosome 13 D13Z1 or chromosome 21 D21Z1 using modified LNA primers. DNA isolated from human/rodent hybrids containing a single human chromosome was used to assess whether the LNA-modified primers differentiate between centromeres 13 and 21. The PCR assay for the centromere 13 repeat D13Z1 comprised an LNA forward primer and an unmodified reverse primer that

15 binds both centromere 13 and 21 repeats (see also Fig. 11A). The PCR assay for centromere 21 D21Z1 contains both forward and reverse LNA primers that bind D21Z1 substitutions and a clamp for D13Z1, which is the same forward primer used to detect D13Z1, but which is phosphorylated at its 3' end to inhibit D13Z1 amplification (see also Fig. 11B). The D21Z1 assay detects substitutions in both centromere 13 and centromere

20 21, but substantially reduces the non-specific amplification of D13Z1. The gel displays the amplification products at 16 PCR cycles. The Ct data are shown in Fig. 11. Figure 10C shows the frequency of detection of D13Z1 and D21Z1 variants in humans, hominids, and apes. Nucleotide substitutions specific for D13Z1 and D21Z1 were genotyped in silico on Illumina sequence reads obtained in diverse populations studied

25 in the human genome diversity project (HGDP), the most recent human genome assembly (Hg38), sequences from extinct hominids, and sequence reads from populations of apes. BLAST was used to identify sequence reads that match with 100% similarity to a 38-bp sequence surrounding the D13Z1/D21Z1 substitution (T/C). The numbers of sequence reads per million base pairs (bp) that match D13Z1 and D21Z1

30 variants are shown in the bar plot. D13Z1 and D21Z1 substitutions were almost undetectable in apes and varied between hominid and modern human. The accession numbers of the SRA libraries are: orangutan (SRX015451), gorilla (ERX196778), chimpanzee (ERX007121), neanderthal (ERX232274), denisovan (ERX007980), and human (ERX007976, ERX004003).

Figure 11 shows PCR detection of nucleotide substitutions in centromeres 13 and 21. The homologue arrays D13Z1/D21Z1, present in centromere 13 and centromere 21, are nearly identical. An A/T substitution, however, has been shown to exist in centromere 13 in the D13Z1 array (see also Fig. 10). Figure 11A shows results from a qPCR assay that was developed to detect the “T” substitution present in D13Z1 using a primer with locked nucleic acid (LNA) modifications at the nucleotide substitution and one base before and after it. This primer, in combination with a reverse primer that binds to both D13Z1 and D21Z1, was included in a PCR reaction to detect the D13Z1 array specifically. This combination of primers was tested in a PCR reaction with DNA isolated from human/rodent hybrid cells that contain human chromosome 13 or hybrids that contain other chromosomes, using an annealing/extension temperature gradient (60-70°C, see panels A to F showing the data collected at the annealing temperatures tested). Detection of D13Z1 sequences was confirmed by sequencing. Increasing annealing temperatures in a PCR reaction showed that at 68°C (panel E), the D13Z1 LNA primer exclusively detected D13Z1, while the D21Z1 array is not detected. At 70°C (panel F), the PCR assay detected neither array. Figure 11B shows results from a qPCR that was developed to detect these “C and G” substitutions in D21Z1 using primers with locked nucleic acid (LNA) modifications at the substitutions and one base before and after them. These forward and reverse primers, in combination with a D13Z1 primer clamp, which is the same LNA primer used to detect D13Z1 but comprising a phosphate at the 3' end to inhibit the amplification of D13Z1, was included in a PCR reaction to detect the D21Z1 array specifically. The combination of primers and clamps was amplified in a temperature gradient reaction (60-70°C) using DNA isolated from human/rodent cell hybrids that contain chromosome 21, or hybrids that contain other chromosomes. Panel A – amplification of the centromere 13 array D13Z1 is substantially inhibited with the D13Z1 clamp at an annealing temperature of 64°C in a PCR reaction. Panel B – the D21Z1 LNA primers preferentially detect the D21Z1 array. Panel C – assessing the DNA from human/rodent hybrids shows that the PCR conditions favor the amplification of the centromere of chromosome 21 rather than that of chromosomes 13 or Y.

Figure 12 shows centromere and pericentromere instability in individuals with Trisomy 21. Figure 12A shows detection of D13Z1 and Figure 12B shows detection of D21Z1 variants in individuals with trisomy 13 or 21 by qPCR using LNA primers and clamps. The copy number of each repeat variant was determined in 50 ng of DNA. Figure 12C shows K111 and K222 pericentric provirus sequences used as markers to

study human pericentromeres. K111 is present in the pericentromeres of 15 chromosomes and K222 in the pericentromeres of 9 chromosomes, usually as single copies with some exceptions – in particular, hundreds of copies of K111 have accumulated at pericentromere 21 (Contreras-Galindo et al., 2013), and dozens of K222 copies have accumulated in pericentromeres 13, 14 and 15 (Zahn et al., 2015). A PCR assay was developed for K111 plus K222 *env* (Figure 12C) and a PCR assay was developed that is specific for K111 *gag* (Figure 12D) to assess the structural variation (length) of pericentromeres 13 and 21 in DNA from healthy individuals and individuals with trisomy 13 or 21 (see Fig. 13). The K111 + K222 assay (Figure 12C) predicts the lengths of pericentromere 13 and 21, whereas the K111-specific PCR (Figure 12D) predicts the length of pericentromere 21. In contrast to healthy individuals, loss of pericentromeric K111 and K222 sequences was observed in the DNA of individuals with trisomy 21. Sequencing confirms that pericentromeric losses occurred in centromere 21 in the individuals with trisomy 21 (Figure 13). Statistical significance amongst the groups was calculated using the t test. P values < 0.05 (*), <0.01 (**), <0.001 (***), and <0.0001 (****) are shown.

Figure 13 shows that the pericentric K111 marker indicates loss and apparent recombination in trisomy 21. Figure 13 shows a Bayesian inference tree of the K111 5' LTR insertions amplified from each individual human chromosome together with total K111 proviruses amplified in healthy individuals and subjects with trisomy 21. Sequence labels are shaded to indicate from which human chromosome they arise. Each chromosome was assigned a shade depicted on the legend at left. Note that each shade tends to cluster to specific evolutionary branches, indicating that individual K111 sequences often spread within an individual chromosome (black bold numbers or triangle shaded branches). Posterior probability values greater than 70 are shown. K111 sequences from healthy individuals (M382, nl, and 5051) and DNA from individuals with trisomy 21 (legend on the right: Subject A 1258, Subject B 5277, Subject C 4904, and Subject D sp707) distribute along the tree clustering close to the K111 sequences specific to each chromosome. As an example, K111 sequences from a healthy individual (M386) are indicated (dots). Several K111 sequences in trisomy 21 subjects A, B, and C cluster to novel branches (black triangle branches), likely the result of homologous recombination, and K111 sequences are substantially reduced in number in chromosome 21 in these individuals with Down Syndrome. The tree was generated using Bayesian inference with four independent chains run for at least 10,000,000 generations until sufficient trees were sampled to generate >99% credibility. 5' LTRs, 3' LTRs, and

soloLTR lineages are shown along with the chimpanzee LTRs (CERV-K111). Nucleotide sequence substitutions that are specific for the K111 group of sequences found in each chromosome were used to generate the tree (Contreras-Galindo et al., 2013).

Figure 14 is a heat map showing centromere instability in scleroderma (e.g.,
 5 diffuse cutaneous scleroderma). The heat map shows the abundance of α -repeats specific for each centromere array (Y-axis) obtained by PCR in 50 ng of DNA from healthy fibroblasts (N59, N60, and N64) and from diffuse cutaneous scleroderma ("dcSSc") skin fibroblasts (5043, 5060, 5065, 5074) (X-axis). The intensity of each cell in the heat map is depicted by the $\log_{10}$ Z-score gradient on the right. The log numbers for each α -repeat
 10 were normalized to the average copy number of a given repeat in DNA from healthy cells. D1Z5, for example, refers to the array Z5 in centromere 1. Significant differences between the two groups are indicated with stars (t-test). Cells for DYZ3, N64 and DYZ3, 5060 had Z-scores of approximately 4 and the cell D1Z5, 5065 had a Z-score of approximately -6. Cells for D2Z1, 5043; D1Z5, 5060; and D15Z3, 5060 had Z-scores of
 15 approximately -5 and cells D19Z5, 5060 and D19Z5, 5065 had Z-scores of approximately -3.

Figure 15A is an immunoblot and Figure 15B is a heat map. Experiments were conducted in which CHON-002 fibroblasts were treated with different concentration of bleomycin (1.75, 3.5, and 7.0 μ M), paraquat (5 and 50 μ M), or agent orange 2,4,5 T (5
 20 and 50 μ M) for 3 hours. Figure 15A is an image of an immunoblot showing the detection of expression of H2AX, a chromatin modification induced early after DSBs. Actin B was used as a loading control. Figure 15B is a heat map representing the abundance of α -repeats specific for each centromere array (Y-axis) obtained by PCR in 50 ng of DNA from fibroblasts treated with the three different chemicals. The intensity of the heat
 25 map is depicted by the $\log_2$ Z-score color gradient on the right. The $\log_2$ Z-score color bar (right) depicts the color intensity of the heat map. The log numbers for each α -repeat were normalized to the average copy number of a given repeat in vehicle-treated cells. No variation is seen in the genes top3a and gapdh (not shown).

It is to be understood that the figures are not necessarily drawn to scale, nor are
 30 the objects in the figures necessarily drawn to scale in relationship to one another. The figures are depictions that are intended to bring clarity and understanding to various embodiments of apparatuses, systems, and methods disclosed herein. Wherever possible, the same reference numbers will be used throughout the drawings to refer to the same or like parts. Moreover, it should be appreciated that the drawings are not intended to
 35 limit the scope of the present teachings in any way.

DETAILED DESCRIPTION

Chromosomal abnormalities have been associated with over 60 syndromes. They are present in 50% of spontaneous abortions, 6% of stillbirths, and approximately 0.5% of newborns. In women aged 35 or over chromosomal abnormalities are detected in 2% of all pregnancies. Common autosomal numerical abnormalities in infants include trisomy 21 (approximately 1/800 live births), trisomy 18 (approximately 1/8140 live births), and trisomy 13 (approximately 1/19,000). Down syndrome (trisomy 21) is the most common chromosomal abnormality in live births. The prevalence is approximately 1 in 600-800 live births. The prevalence increases with increasing maternal age, so that the risk is approximately 1 in 1,340 at age 25; approximately 1 in 353 at age 35; and approximately 1 in 35 at age 45.

Prenatal testing is widely recommended and used to screen for chromosomal abnormalities. Noninvasive screenings can detect chromosomal abnormalities and often indicate subsequent testing with invasive diagnostic assays. In particular, several million births occur in the U.S. each year – prior to birth, approximately 60-70% of pregnant women receive prenatal care and approximately 2 million mothers and fetuses receive prenatal screens. Then, several hundred thousand of these screens are followed by invasive procedures, including karyotype, array, and FISH.

While useful, the current noninvasive screening technologies carry a high false positive rate. Furthermore, the invasive diagnostic tests have significant associated risks of miscarriage. Thus, there is a significant need for an accurate noninvasive test to reduce this risk to the fetus. Accordingly, provided herein is technology relating to the analysis of chromosome centromeres and particularly, but not exclusively, to methods, compositions, kits, and systems for detecting, identifying, characterizing, and quantifying chromosome centromeres.

In this detailed description of the various embodiments, for purposes of explanation, numerous specific details are set forth to provide a thorough understanding of the embodiments disclosed. One skilled in the art will appreciate, however, that these various embodiments may be practiced with or without these specific details. In other instances, structures and devices are shown in block diagram form. Furthermore, one skilled in the art can readily appreciate that the specific sequences in which methods are presented and performed are illustrative and it is contemplated that the sequences can be varied and still remain within the spirit and scope of the various embodiments disclosed herein.

All literature and similar materials cited in this application, including but not limited to, patents, patent applications, articles, books, treatises, and internet web pages are expressly incorporated by reference in their entirety for any purpose. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of ordinary skill in the art to which the various embodiments described herein belongs. When definitions of terms in incorporated references appear to differ from the definitions provided in the present teachings, the definition provided in the present teachings shall control. The section headings used herein are for organizational purposes only and are not to be construed as limiting the described subject matter in any way.

Definitions

To facilitate an understanding of the present technology, a number of terms and phrases are defined below. Additional definitions are set forth throughout the detailed description.

Throughout the specification and claims, the following terms take the meanings explicitly associated herein, unless the context clearly dictates otherwise. The phrase “in one embodiment” as used herein does not necessarily refer to the same embodiment, though it may. Furthermore, the phrase “in another embodiment” as used herein does not necessarily refer to a different embodiment, although it may. Thus, as described below, various embodiments of the invention may be readily combined, without departing from the scope or spirit of the invention.

In addition, as used herein, the term “or” is an inclusive “or” operator and is equivalent to the term “and/or” unless the context clearly dictates otherwise. The term “based on” is not exclusive and allows for being based on additional factors not described, unless the context clearly dictates otherwise. In addition, throughout the specification, the meaning of “a”, “an”, and “the” include plural references. The meaning of “in” includes “in” and “on.”

As used herein, an “alpha-repeat array” or “ α -repeat array” refers to a form of chromosomal satellite DNA comprising an array of tandemly repeating nucleic acid sequences called “alpha-repeats” or “alpha-repeat sequences”. In some embodiments, each alpha-repeat comprises approximately 171 nucleotides and the alpha-repeats are organized in a head-to-tail arrangement in which the alpha-repeats are adjacent to each other in the alpha-repeat array. That is, in some embodiments, an alpha-repeat array is a long continuous DNA molecule that contains multiple copies of the same DNA

sequences (alpha-repeats) linked in series. By convention, the nomenclature of α -repeat arrays starts with the letter "D", followed by the chromosome number in which they reside (e.g., 1-22, X, or Y), followed by a "Z", and then a number indicating the order in which the sequence was discovered. For example, in the array D2Z1, D2 stands for an α -repeat array in chromosome 2, and Z1 indicates that the array was the first array identified in this particular chromosome.

As used herein, a "nucleic acid" shall mean any nucleic acid molecule, including, without limitation, DNA, RNA, and hybrids thereof. The nucleic acid bases that form nucleic acid molecules can be the bases A, C, G, T and U, as well as derivatives thereof. Derivatives of these bases are well known in the art. The term should be understood to include, as equivalents, analogs of either DNA or RNA made from nucleotide analogs. The term as used herein also encompasses cDNA, that is complementary, or copy, DNA produced from an RNA template, for example by the action of a reverse transcriptase. It is well known that DNA (deoxyribonucleic acid) is a chain of nucleotides consisting of 4 types of nucleotides—A (adenine), T (thymine), C (cytosine), and G (guanine)—and that RNA (ribonucleic acid) is a chain of nucleotides consisting of 4 types of nucleotides—A, U (uracil), G, and C. It is also known that all of these 5 types of nucleotides specifically bind to one another in combinations called complementary base pairing. That is, adenine (A) pairs with thymine (T) (in the case of RNA, however, adenine (A) pairs with uracil (U)), and cytosine (C) pairs with guanine (G), so that each of these base pairs forms a double strand. The term "nucleic acid" encompasses nucleic acids that include any of the known heterocyclic bases and base analogs of DNA and RNA including, but not limited to, adenine, guanine, cytosine, thymine, uracil, inosine, xanthine, hypoxanthine, or a heterocyclic derivative, analog, or tautomer thereof. A nucleobase can be naturally occurring or synthetic. When a nucleic acid such as an oligonucleotide is represented by a sequence of letters, such as "ATGCCTG," it will be understood that the nucleotides are in 5' to 3' order from left to right and that "A" denotes deoxyadenosine, "C" denotes deoxycytidine, "G" denotes deoxyguanosine, "T" denotes thymidine, and "U" denotes uracil, unless otherwise noted. The letters A, C, G, and T may be used to refer to the bases themselves, to nucleosides, or to nucleotides comprising the bases, as is standard in the art.

As used herein, a "nucleotide" comprises a "base" (alternatively, a "nucleobase" or "nitrogenous base"), a "sugar" (in particular, a five-carbon sugar, e.g., ribose or 2-deoxyribose), and a "phosphate moiety" of one or more phosphate groups (e.g., a monophosphate, a diphosphate, or a triphosphate consisting of one, two, or three linked

phosphates, respectively). Without the phosphate moiety, the nucleobase and the sugar compose a “nucleoside”. A nucleotide can thus also be called a nucleoside monophosphate or a nucleoside diphosphate or a nucleoside triphosphate, depending on the number of phosphate groups attached. The phosphate moiety is usually attached to the 5-carbon of the sugar, though some nucleotides comprise phosphate moieties attached to the 2-carbon or the 3-carbon of the sugar. Nucleotides contain either a purine (in the nucleotides adenine and guanine) or a pyrimidine base (in the nucleotides cytosine, thymine, and uracil). Ribonucleotides are nucleotides in which the sugar is ribose. Deoxyribonucleotides are nucleotides in which the sugar is deoxyribose.

In some embodiments, a nucleotide comprises a heterocyclic base (e.g., nucleobase) such as adenine, guanine, cytosine, thymine, uracil, inosine, xanthine, hypoxanthine, or a heterocyclic derivative, analog, or tautomer thereof. A nucleobase can be naturally occurring or synthetic. Non-limiting examples of nucleobases are adenine, guanine, thymine, cytosine, uracil, xanthine, hypoxanthine, 8-azapurine, purines substituted at the 8 position with methyl or bromine, 9-oxo-N⁶-methyladenine, 2-aminoadenine, 7-deazaxanthine, 7-deazaguanine, 7-deaza-adenine, N⁴-ethanocytosine, 2,6-diaminopurine, N⁶-ethano-2,6-diaminopurine, 5-methylcytosine, 5-(C3-C6)-alkynylcytosine, 5-fluorouracil, 5-bromouracil, thiouracil, pseudoisocytosine, 2-hydroxy-5-methyl-4-triazolopyridine, isocytosine, isoguanine, inosine, 7,8-dimethylalloxazine, 6-dihydrothymine, 5,6-dihydrouracil, 4-methyl-indole, ethenoadenine, 4-acetylcytosine, 8-hydroxy-N⁶-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxyl-methyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethyl-aminomethyluracil, dihydrouracil, inosine, N⁶-isopentenyladenine, 1-methyladenine, 1-methylpseudo-uracil, 1-methylguanine, 1-methylinosine, 2,2-dimethyl-guanine, 2-methyladenine, 2-methylguanine, 3-methyl-cytosine, 5-methylcytosine, N⁶-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxy-amino-methyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N⁶-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, 2,6-diaminopurine, and the non-naturally occurring nucleobases described in U.S. Pat. Nos. 5,432,272 and 6,150,510 and PCT applications WO 92/002258, WO 93/10820, WO 94/22892, and WO 94/24144, and Fasman (“Practical Handbook of Biochemistry and

Molecular Biology”, pp. 385-394, 1989, CRC Press, Boca Raton, La.), all herein incorporated by reference in their entireties.

Reference to a base, a nucleotide, or to another molecule may be in the singular or plural. That is, “a base” may refer to a single molecule of that base or to a plurality of the base, e.g., in a solution.

As used herein, the term “oligonucleotide” refers to a nucleic acid that includes at least two nucleic acid monomer units (e.g., nucleotides), typically more than three monomer units, and more typically greater than ten monomer units. The exact size of an oligonucleotide generally depends on various factors, including the ultimate function or use of the oligonucleotide. To illustrate, oligonucleotides are typically less than 200 residues long (e.g., between 15 and 100), however, as used herein, the term is also intended to encompass longer polynucleotide chains. Oligonucleotides are often referred to by their length. For example a 24-residue oligonucleotide is referred to as a “24-mer”. Typically, the nucleoside monomers are linked by phosphodiester bonds or analogs thereof, including phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phosphoranilidate, phosphoramidate, and the like, including associated counterions, e.g., H^+ , NH_4^+ , Na^+ , and the like, if such counterions are present. Further, oligonucleotides are typically single-stranded. Oligonucleotides are optionally prepared by any suitable method, including, but not limited to, isolation of an existing or natural sequence, DNA replication or amplification, reverse transcription, cloning and restriction digestion of appropriate sequences, or direct chemical synthesis by a method such as the phosphotriester method.

As used herein, the term “primer” refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, that is capable of acting as a point of initiation of synthesis when placed under conditions in which synthesis of a primer extension product that is complementary to a nucleic acid strand is induced (e.g., in the presence of nucleotides and an inducing agent such as a biocatalyst (e.g., a DNA polymerase or the like) and at a suitable temperature and pH). The primer is typically single stranded for maximum efficiency in amplification, but may alternatively be double stranded. If double stranded, the primer is generally first treated to separate its strands before being used to prepare extension products. In some embodiments, the primer is an oligodeoxyribonucleotide. The primer is sufficiently long to prime the synthesis of extension products in the presence of the inducing agent. The exact lengths of the primers will depend on many factors, including temperature, source

of primer and the use of the method. In certain embodiments, the primer is a capture primer.

As used herein, the term “probe” refers to an oligonucleotide, whether occurring naturally as in a purified restriction digest or produced synthetically, recombinantly or by PCR amplification, that is capable of hybridizing to at least a portion of another oligonucleotide of interest. A probe may be single-stranded or double-stranded. Probes are useful in the detection, identification, and isolation of particular nucleic acids, gene sequences, etc.

As used herein, the terms “complementary” or “complementarity” are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence “5'-A-G-T-3',” is complementary to the sequence “3'-T-C-A-5'.” Complementarity may be “partial,” in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be “complete” or “total” complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

As used herein, the term “amplifying” or “amplification” in the context of nucleic acids refers to the production of multiple copies of a polynucleotide, or a portion of the polynucleotide, typically starting from a small amount of the polynucleotide (e.g., a single polynucleotide molecule), where the amplification products or amplicons are generally detectable. Amplification of polynucleotides encompasses a variety of chemical and enzymatic processes, including but not limited to polymerase chain reaction (PCR), ligase chain reaction (LCR), helicase-dependent amplification, multiplex ligation-dependent probe amplification, real time PCR, reverse transcription PCR, nucleic acid sequence-based amplification (NASBA), and transcription-mediated amplification (TMA).

As used herein, the term “amplicon” refers to a nucleic acid generated in a nucleic acid amplification reaction, e.g., PCR and the like. As used herein, the terms “PCR product” or amplicon or “PCR fragment” generally refer to the resultant mixture of amplified DNA after two or more cycles of the PCR steps of denaturation, annealing, and extension are complete. The sequence of an amplicon includes the amplified segment of the target DNA as well as the sequence of the primers flanking the amplified region that were employed to carry out the PCR. These terms are also meant to

encompass the case where there has been amplification of one or more segments of one or more target sequences.

A “polymerase” is an enzyme generally for joining 3'-OH, 5'-triphosphate nucleotides, oligomers, and their analogs. Polymerases include, but are not limited to, DNA-dependent DNA polymerases, DNA-dependent RNA polymerases, RNA-dependent DNA polymerases, RNA-dependent RNA polymerases, T7 DNA polymerase, T3 DNA polymerase, T4 DNA polymerase, T7 RNA polymerase, T3 RNA polymerase, SP6 RNA polymerase, DNA polymerase 1, Klenow fragment, *Thermophilus aquaticus* DNA polymerase, Tth DNA polymerase, Vent DNA polymerase (New England Biolabs), Deep Vent DNA polymerase (New England Biolabs), Bst DNA Polymerase Large Fragment, Stoeffel Fragment, 9° N DNA Polymerase, Pfu DNA Polymerase, Tfi DNA Polymerase, RepliPhi Phi29 Polymerase, Tli DNA polymerase, eukaryotic DNA polymerase beta, telomerase, Therminator polymerase (New England Biolabs), KOD HiFi DNA polymerase (Novagen), KOD1 DNA polymerase, Q-beta replicase, terminal transferase, AMV reverse transcriptase, M-MLV reverse transcriptase, Phi6 reverse transcriptase, HIV-1 reverse transcriptase, novel polymerases discovered by bioprospecting, and polymerases described in US 2007/0048748, U.S. Pat. Nos. 6,329,178, 6,602,695, and 6,395,524, each herein incorporated by reference. These polymerases include wild-type, mutant isoforms, and genetically engineered variants.

As used herein, a “sequence” of a biopolymer refers to the order and identity of monomer units (e.g., nucleotides, amino acids, sugars, etc.) in the biopolymer. The sequence (e.g., base sequence) of a nucleic acid is typically read in the 5' to 3' direction.

As used herein, “nucleic acid sequence”, “nucleotide sequence”, and the like denotes any information or data that is indicative of the order of the nucleotide bases (e.g., adenine, guanine, cytosine, and thymine/uracil) in a molecule (e.g., a whole genome, a chromosome, a whole transcriptome, an exome, oligonucleotide, polynucleotide, fragment, etc.) of DNA or RNA.

As used herein, “moiety” refers to one of two or more parts into which something may be divided, such as, for example, the various parts of a tether, a molecule or a probe.

As used herein, a “sample” refers to anything capable of being processed or analyzed by the technology provided herein. In some embodiments, the sample comprises or is suspected to comprise one or more nucleic acids capable of amplification or comprises an amplicon after amplification. In certain embodiments, for example, the samples comprise nucleic acids (e.g., DNA, RNA, cDNAs, etc.) from one or more subjects,

tissues, or cells. Samples can include, for example, blood, semen, saliva, urine, feces, rectal swabs, cells, mucous, and the like. The sample can be obtained by a variety of manners such as by biopsy, swabbing, and the like. The samples may be obtained by a physician in a hospital or other health care environment.

5 As used herein, the term “subject” refers to any animal (e.g., a mammal), including, but not limited to, humans, non-human primates, rodents, and the like. Typically, the terms “subject” and “patient” are used interchangeably herein in reference to a human subject. Thus, subjects include, but are not limited to, a mammal, a bird, or a reptile. The subject may be a cow, horse, dog, cat, or a primate. The subject may be
10 living or dead.

 As used herein, the term “sample template” refers to nucleic acid originating from a sample that is analyzed for the presence of “target” (defined below). In contrast, “background” is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It
15 may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be removed (e.g., reduced, minimized, substantially or effectively eliminated) from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

 As used herein, the term “target” refers to a nucleic acid sequence or structure to
20 be detected or characterized.

 As used herein, the term “amplification reagents” refers to those reagents (deoxyribonucleotide triphosphates, buffer, etc.), needed for amplification except for primers, nucleic acid template, and the amplification enzyme. Typically, amplification reagents along with other reaction components are placed and contained in a reaction
25 vessel (test tube, microwell, modular random access vessel, etc.).

 As used herein, the term “detector” refers to a system or component of a system, e.g., an instrument (e.g. a camera, fluorimeter, charge-coupled device, scintillation counter, etc) or a reactive medium (X-ray or camera film, pH indicator, etc.), that can convey to a user or to another component of a system (e.g., a computer or controller) the
30 presence of a signal or effect. A detector can be a photometric or spectrophotometric system, which can detect ultraviolet, visible or infrared light, including fluorescence or chemiluminescence; a radiation detection system; a spectroscopic system such as nuclear magnetic resonance spectroscopy, mass spectrometry or surface enhanced Raman spectrometry; a system such as gel or capillary electrophoresis or gel exclusion
35 chromatography; or other detection system known in the art, or combinations thereof.

As used herein, the term “recombinant,” when used in reference to a polynucleotide or a protein, generally refers to a polynucleotide or a polypeptide molecule that is produced using genetic engineering techniques and that is distinct from a naturally occurring nucleic acid or polypeptide molecule. Recombinant DNA (sometimes represented as “rDNA”) is an artificial DNA sequence resulting from the combining of two other DNA sequences in a plasmid/vector. The term recombinant DNA refers to a new combination of DNA molecules that are not found together naturally. Although processes such as crossing over (genetic recombination) technically produce recombinant DNA, the term is generally reserved for DNA produced by joining molecules derived from different biological sources.

As used herein, the term “hybridization” refers to any process by which a strand of nucleic acid binds with a complementary strand through nucleotide base pairing, preferably Watson-Crick base pairing.

As used herein, the term “specific hybridization” or “selective hybridization” of a probe to a target site of a template nucleic acid refers to hybridization of the probe predominantly to the target, such that the hybridization signal may be clearly interpreted. As further described herein, such conditions resulting in specific hybridization vary depending on the length of the region of homology, the GC content of the region, and/or the melting temperature (“T_m”) of the hybrid. Hybridization conditions will thus vary in the salt content, acidity, and temperature of the hybridization solution and the washes.

As used herein, the terms “label” and “detectable label” refer to a molecule capable of detection including, but not limited to radioactive isotopes, fluorophores, chemiluminescent moieties, enzymes, enzyme substrates, enzyme cofactors, enzyme inhibitors, dyes, metal ions, ligands (e.g., biotin or haptens), and the like. In some embodiments, a detectable label comprises a fluorescent moiety (e.g., a fluorogenic dye, also referred to as a “fluorophore” or a “fluor”). “Fluorophore” refers to a substance or a portion thereof which is capable of exhibiting fluorescence in the detectable range. A wide variety of fluorescent moieties is known in the art and methods are known for linking a fluorescent moiety to a nucleotide prior to incorporation of the nucleotide into an oligonucleotide and for adding a fluorescent moiety to an oligonucleotide after synthesis of the oligonucleotide.

Examples of compounds that may be used to modify a nucleotide and/or a nucleic acid (e.g., to provide a detectable label) include but are not limited to xanthene, anthracene, cyanine, porphyrin, and coumarin dyes, e.g., xanthene derivatives such as

fluorescein, rhodamine, Oregon green, eosin, and TEXAS RED dye; cyanine derivatives such as cyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, and merocyanine; naphthalene derivatives (dansyl and prodan derivatives); coumarin derivatives; oxadiazole derivatives such as pyridyloxazole, nitrobenzoxadiazole, and benzoxadiazole; pyrene derivatives such as cascade blue; oxazine derivatives such as Nile red, Nile blue, cresyl violet, and oxazine 170; acridine derivatives such as proflavin, acridine orange, and acridine yellow; arylmethine derivatives such as auramine, crystal violet, and malachite green; and tetrapyrrole derivatives such as porphin, phtalocyanine, bilirubin.

Examples of xanthene dyes that find use with the present technology include but are not limited to fluorescein, 6-carboxyfluorescein (6-FAM dye), 5-carboxyfluorescein (5-FAM dye), 5- or 6-carboxy-4, 7, 2', 7'- tetrachlorofluorescein (TET dye), 5- or 6-carboxy-4'5'2'4'5'7' hexachlorofluorescein (HEX dye), 5' or 6'-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein (JOE dye), 5-carboxy-2',4',5',7'-tetrachlorofluorescein (ZOE dye), rhodol, rhodamine, tetramethylrhodamine (TAMRA dye), 4,7-dichlorotetramethyl rhodamine (DTAMRA dye), rhodamine X (ROX dye), and TEXAS RED dye. Examples of cyanine dyes that may find use with the present invention include but are not limited to CY3 dye, CY3B dye, CY3.5 dye, CY5 dye, CY5.5 dye, CY7 dye, and CY7.5 dye. Other fluorescent moieties and/or dyes that find use with the present technology include but are not limited to energy transfer dyes, composite dyes, and other aromatic compounds that give fluorescent signals. In some embodiments, the fluorescent moiety comprises a quantum dot.

Additional examples of compounds that may be used to modify a nucleotide and/or a nucleic acid include but are not limited to, d-Rhodamine acceptor dyes including CY5 dye, dichloro[R110], dichloro[R6G], dichloro[TAMRA], dichloro[ROX] or the like, fluorescein donor dyes including fluorescein, 6-FAM, 5-FAM, or the like; Acridine including Acridine orange, Acridine yellow, Proflavin, pH 7, or the like; Aromatic Hydrocarbons including 2-Methylbenzoxazole, Ethyl p-dimethylaminobenzoate, Phenol, Pyrrole, benzene, toluene, or the like; Arylmethine Dyes including Auramine O, Crystal violet, Crystal violet, glycerol, Malachite Green or the like; Coumarin dyes including 7-Methoxycoumarin-4-acetic acid, Coumarin 1, Coumarin 30, Coumarin 314, Coumarin 343, Coumarin 6 or the like; Cyanine Dyes including 1,1'-diethyl-2,2'-cyanine iodide, Cryptocyanine, Indocarbocyanine (C3) dye, Indodicarbocyanine (C5) dye, Indotricarbocyanine (C7) dye, Oxacarbocyanine (C3) dye, Oxadicarbocyanine (C5) dye, Oxatricarbocyanine (C7) dye, Pinacyanol iodide, Stains all,

Thiacarbocyanine (C3) dye, ethanol, Thiacarbocyanine (C3) dye, n-propanol, Thiadicarbocyanine (C5) dye, Thiatricarbocyanine (C7) dye, or the like; Dipyrin dyes including N,N'-Difluoroboryl-1,9-dimethyl-5-(4-iodophenyl)-dipyrin, N,N'-Difluoroboryl-1,9-dimethyl-5-[(4-(2-trimethylsilylethynyl)]-, N,N'-Difluoroboryl-1,9-dimethyl-5-phenyldipyrin, or the like; Merocyanines including 4-(dicyanomethylene)-2-methyl-6-(p-dimethylaminostyryl)-4H-pyran (DCM), acetonitrile, 4-(dicyanomethylene)-2-methyl-6-(p-dimethylaminostyryl)-4H-pyran (DCM), methanol, 4-Dimethylamino-4'-nitrostilbene, Merocyanine 540, or the like; Miscellaneous Dyes including 4',6'-Diamidino-2-phenylindole (DAPI), dimethylsulfoxide, 7-Benzylamino-4-nitrobenz-2-oxa-1,3-diazole, Dansyl glycine, Dansyl glycine, dioxane, Hoechst 33258, DMF, Hoechst 33258, Lucifer yellow CH, Piroxicam, Quinine sulfate, Quinine sulfate, Squarylium dye III, or the like; Oligophenylenes including 2,5-Diphenyloxazole (PPO), Biphenyl, POPOP, p-Quaterphenyl, p-Terphenyl, or the like; Oxazines including Cresyl violet perchlorate, Nile Blue, methanol, Nile Red, ethanol, Oxazine 1, Oxazine 170, or the like; Polycyclic Aromatic Hydrocarbons including 9,10-Bis(phenylethynyl)anthracene, 9,10-Diphenylanthracene, Anthracene, Naphthalene, Perylene, Pyrene, or the like; polyene/polyyne including 1,2-diphenylacetylene, 1,4-diphenylbutadiene, 1,4-diphenylbutadiyne, 1,6-Diphenylhexatriene, Beta-carotene, Stilbene, or the like; Redox-active Chromophores including Anthraquinone, Azobenzene, Benzoquinone, Ferrocene, Riboflavin, Tris(2,2'-bipyridyl)pruthenium(II), Tetrapyrrole, Bilirubin, Chlorophyll a, diethyl ether, Chlorophyll a, methanol, Chlorophyll b, Diprotonated-tetraphenylporphyrin, Hematin, Magnesium octaethylporphyrin, Magnesium octaethylporphyrin (MgOEP), Magnesium phthalocyanine (MgPc), PrOH, Magnesium phthalocyanine (MgPc), pyridine, Magnesium tetramesitylporphyrin (MgTMP), Magnesium tetraphenylporphyrin (MgTPP), Octaethylporphyrin, Phthalocyanine (Pc), Porphin, ROX dye, TAMRA dye, Tetra-t-butylazaporphine, Tetra-t-butyl-naphthalocyanine, Tetrakis(2,6-dichlorophenyl)porphyrin, Tetrakis(o-aminophenyl)porphyrin, Tetramesitylporphyrin (TMP), Tetraphenylporphyrin (TPP), Vitamin B12, Zinc octaethylporphyrin (ZnOEP), Zinc phthalocyanine (ZnPc), pyridine, Zinc tetramesitylporphyrin (ZnTMP), Zinc tetramesitylporphyrin radical cation, Zinc tetraphenylporphyrin (ZnTPP), or the like; Xanthenes including Eosin Y, Fluorescein, basic ethanol, Fluorescein, ethanol, Rhodamine 123, Rhodamine 6G, Rhodamine B, Rose bengal, Sulforhodamine 101, or the like; PACIFIC BLUE dye, PACIFIC ORANGE dye, PACIFIC GREEN dye, or the like; or mixtures or combination thereof or synthetic derivatives thereof.

Further examples of compounds that may be used to modify a nucleotide and/or a nucleic acid (e.g., to provide a detectable label) include but are not limited to a fluorescent moiety that is xanthene, fluorescein, rhodamine, BODIPY, cyanine, coumarin, pyrene, phthalocyanine, phycobiliprotein, ALEXA FLUOR® 350, ALEXA FLUOR® 405, ALEXA FLUOR® 430, ALEXA FLUOR® 488, ALEXA FLUOR® 514, ALEXA FLUOR® 532, ALEXA FLUOR® 546, ALEXA FLUOR® 555, ALEXA FLUOR® 568, ALEXA FLUOR® 594, ALEXA FLUOR® 610, ALEXA FLUOR® 633, ALEXA FLUOR® 647, ALEXA FLUOR® 660, ALEXA FLUOR® 680, ALEXA FLUOR® 700, ALEXA FLUOR® 750, or a squaraine dye. In some embodiments, a nucleotide and/or a nucleic acid is modified with a fluorescently detectable moiety as described in, e.g., Haugland (September 2005) MOLECULAR PROBES HANDBOOK OF FLUORESCENT PROBES AND RESEARCH CHEMICALS (10th ed.), which is herein incorporated by reference in its entirety.

In some embodiments a nucleic acid and/or nucleotide is modified with a moiety available from ATTO-TEC GmbH (Am Eichenhang 50, 57076 Siegen, Germany), e.g., as described in U.S. Pat. Appl. Pub. Nos. 20110223677, 20110190486, 20110172420, 20060179585, and 20030003486; and in U.S. Pat. No. 7,935,822, each of which is incorporated herein by reference (e.g., ATTO 390, ATTO 425, ATTO 465, ATTO 488, ATTO 495, ATTO 514, ATTO 520, ATTO 532, ATTO Rho6G, ATTO 542, ATTO 550, ATTO 565, ATTO Rho3B, ATTO Rho11, ATTO Rho12, ATTO Thio12, ATTO Rho101, ATTO 590, ATTO 594, ATTO Rho13, ATTO 610, ATTO 620, ATTO Rho14, ATTO 633, ATTO 647, ATTO 647N, ATTO 655, ATTO Oxa12, ATTO 665, ATTO 680, ATTO 700, ATTO 725, ATTO740).

Non-limiting examples of fluorescent moieties include dyes that can be synthesized or obtained commercially (e.g., Operon Biotechnologies, Huntsville, Alabama). A large number of dyes (greater than 50) is available for application in fluorescence excitation applications. In some embodiments, dyes with emission maxima from 410 nm (e.g., Cascade Blue) to 775 nm (e.g., Alexa Fluor 750) are available and can be used. Of course, one of ordinary skill in the art will recognize that dyes having emission maxima outside these ranges may be used as well. In some cases, dyes ranging between 500 nm to 700 nm have the advantage of being in the visible spectrum and can be detected, e.g., using existing photomultiplier tubes. In some embodiments, the broad range of available dyes allows selection of dye sets that have emission wavelengths that are spread across the detection range and, in some embodiments have emission spectra that do not overlap, minimally overlap, or have distinguishable emission spectra, e.g., to

provide for the detection of a plurality of probes, amplicons, nucleic acids, etc., e.g., in multiplex detection technologies described herein. Detection systems capable of distinguishing many dyes are known in the art.

In some embodiments, the tag or label comprises a radioisotope, a spin label, a quantum dot, or a bioluminescent moiety. In some embodiments, the label is a fluorescently detectable moiety as described in, e.g., Haugland (September 2005) MOLECULAR PROBES HANDBOOK OF FLUORESCENT PROBES AND RESEARCH CHEMICALS (10th ed.), which is herein incorporated by reference in its entirety.

As used herein, “inhibit” means to diminish, decrease, render less effective, impede, minimize, stop, eliminate, and the like.

As used herein, the term “relative” refers to a comparison of one measurement (e.g., a first quantity) compared to another measurement (e.g., a second quantity). The result of comparison may be that one measurement is greater than, less than, or equal to the other measurement. The result of comparison may be a difference of one measurement from the other or the quotient of one measurement divided by the other.

As used herein, to “determine” refers to measuring, characterizing, identifying, detecting (e.g., detecting the presence and/or absence).

Description

The centromere is the structural unit responsible for the faithful segregation of chromosomes. Although some epigenetic factors are known to regulate centromeric function, the structure and function of centromeric DNA sequences are not well defined. Furthermore, existing methodologies for studying centromere genomics in biology are laborious. During the development of embodiments of the technology described herein, experiments were conducted that identified specific markers in the centromeres of human chromosomes that allow for rapid molecular genetic assays. These assays provide information about the genomic landscape of human centromeres at the time of assay. Embodiments of the technology provide a measurement of cellular ploidy and indicate the specific centromere arrays in each chromosome that drive the recruitment of epigenetic modulators. Data collected during the development of the technology show that, surprisingly, loss and rearrangement of DNA in the centromere of chromosome 21 is associated with trisomy 21. This new technology provides a rapid assessment of the genetics and epigenetics of each specific human centromere, which finds use in the diagnosis and treatment of nondisjunction disorders and other biological settings.

The centromere is a structural unit vital for faithful segregation of chromosomes during cell division. Failure of the cell division machinery to bind to centromeres and correctly separate chromosomes results in genomic instability and aneuploidy (gain or loss of chromosomes), hallmarks of all non-disjunction birth disorders and late-stage malignancies. The genomic landscape of the human centromere is primarily composed of thousands of α -satellite repeats, which are repeat units of approximately 171 bp organized in a head-to-tail arrangement (Verdaasdonk et al., 2011; Hayden, 2012; Aldrup-MacDonald and Sullivan, 2014). The sequence of α -repeat units is approximately 75% similar among all centromeres. In each human centromere, homogenous α -repeats arrange into distinct “high order repeat” (HOR) units that organize contiguously and repeatedly to expand megabases of genomic sequence (Jørgensen, 1997; Hayden, 2012). Interestingly, the α -repeats in each centromeric array have undergone homogenization during human evolution, and today the α -repeat content within each array of an individual chromosome is approximately 98-100% similar (Choo et al., 1989; Jørgensen, 1997; Vissel et al., 1992; Roizès, 2006). Such extensive homogeneity in a locus dominated by repetitive sequences represents a major challenge when devising sequencing methodologies to assemble genomes of the human centromeres. Human centromeres thus remain largely absent from the results of genome sequencing projects, which consequently leads to significant gaps in our knowledge of human genomics (Zeitlin, 2010; Hood et al., 2013).

The centromeres of human chromosomes, however, become more diverse towards their peripheries (the pericentromere). Pericentric domains are composed of more divergent repetitive sequences and retroelements that invaded these areas of the genome during evolution (Bersani et al., 2015; Miga, 2015). Examinations of centromeres have estimated that approximately 30% of pericentric sequences originated from segmental duplications in centromeric regions between chromosomes, many of which then subsequently duplicated interchromosomally.

Two of these sequences are retroelements derived from the human endogenous retroviruses K111 and K222, which were identified at centromeric regions during experiments conducted during the developments of embodiments of the technology described herein (Contreras-Galindo, et al., 2011, 2013; Zahn et al., 2015). Examination of these sequences indicated that they arose from single ancestral infections of K111 and K222 viruses which, after integrating into the genome, copied to several other centromeric loci by recombination within and between the pericentromeres of fifteen (K111) or nine (K222) human chromosomes. During the development of embodiments of

the technology provided herein, data were collected that indicated that this mechanism is consistent with the observed presence of K111 and K222 sequences in centromeric regions (Horvath et al., 2003; Kirsch et al., 2005; Vos et al., 2006; Contreras-Galindo et al., 2013; Zahn et al., 2015). Consequently, K111 and K222 elements have provided
5 insights for understanding dynamic changes of centromeric sequences as a function of time, thus providing information about recombination and segmental duplication that occurred during the course of human evolution between different centromeres, which are areas of the genome previously thought to be relatively recalcitrant to the exchange of genetic material (Talbert and Henikoff, 2010; Contreras-Galindo et al., 2011, 2013; Zahn
10 et al., 2015).

Previous studies of human centromeres primarily focused on the epigenetic components that orchestrate centromere function (Carroll et al., 2006; Vos et al., 2006; Stimpson and Sullivan, 2010). For instance, centromere protein A (CENP-A), a histone-3 variant, loads the centromere core sequence and recruits other proteins to form the
15 centromere structure. CENP-A nucleosomes serve as the structural framework for the assembly of kinetochores, to which spindle fibers bind and pull sister chromatids apart during chromosome segregation. Centromere protein B (CENP-B) also assists in kinetochore assembly, as well as facilitating the cohesion of sister chromatids at α -repeat loci through its specific binding to CENP-B DNA boxes, which are 17-nucleotide
20 sequences present in the majority of centromeric α -repeats (Ohzeki et al., 2002; Rosandić et al., 2006; Fachinetti et al., 2015). Numerous additional centromeric proteins have also been identified as important for kinetochore assembly and stabilization (Fachinetti et al., 2015; Falk et al., 2015). Defects in these centromeric functions generate chromosome instability and missegregation of chromosomes.

25 The lack of conserved α -repeat sequences across species, and the observation that centromere proteins form centromere structures in other regions of chromosomes (neocentromeres) when they lack centromere DNA sequences, led to the prevailing thought in the art that centromere identity is dictated solely by epigenetic factors, and that no genomic component is required for this process (Amor and Choo, 2002; Stimpson
30 and Sullivan, 2010; Scott and Sullivan, 2015). The lack of information on human centromere sequences has hindered studying their potential roles in driving centromere function. However, recent studies revealed that CENP-B binds to CENP-B boxes approximately every 340 bp along the centromere sequence. Within this domain, CENP-A binds two approximately 100-bp segments (Fachinetti et al., 2015; Henikoff et al.,
35 2015), demonstrating a link between the genetic and epigenetic processes that directly

affect centromere biology. Centromeric proteins have evolved to maintain their binding affinity for the constantly changing centromere α -repeats (Malik and Henikoff, 2009). These studies indicate that centromere sequences are necessary for centromere formation and mediate the correct segregation of chromosomes, in contrast to the
5 paradigm that only epigenetics dictate centromere assembly.

Methods for studying the sequence and variation of centromere arrays have previously relied on Southern blotting and restriction analysis. These methods have indicated that arrays exist in different lengths in the human population (Maloney et al., 2012; Aldrup-Macdonald and Sullivan, 2014). However, these methods are laborious,
10 time-consuming, and require large amounts of DNA; they are thus not well-suited to studying centromere sequences efficiently. These studies also usually focus on a single array and, therefore, methods aimed at comprehensively analyzing all human centromere sequences in a given biological setting remain unavailable.

To provide a solution to this problem, literature and DNA databases were
15 analyzed during the development of the technology to identify centromere markers specific for every human chromosome (Jørgensen, 1997; Alexandrov et al., 2001; Hayden, 2012; Liehr, 2013) if any existed. Based on this analysis, embodiments of the technology were developed to provide a rapid and powerful set of assays that specifically detect and quantify the number of α -repeat units in each distinct centromeric array.
20 Moreover, embodiments of the technology specifically detect and quantify the number of α -repeat units in each distinct centromeric array in multiple human chromosomes simultaneously. These approximately 30-minute assays target sequences specific for each centromere array and provide for analyzing the dynamics and evolution of human centromere sequences within a time frame that enables widespread analysis.

Embodiments of the technology provide nucleic acid amplification assays (e.g., PCR assays) that target and quantify the abundance of pericentromere-specific K111 and K222 sequences. These assays find use in studying the evolution of human pericentromeres (Contreras-Galindo et al., 2011, 2013; Zahn et al., 2015). In addition, the assays find use in determining the abundance of α -repeats specific for each
30 centromere array, thus providing a measure of their length and variation in the human population and in different biological settings. Furthermore, the technology finds use in detecting chromosome ploidy alterations that occur in congenital defects that result from nondisjunction. In some embodiments, the technology finds use in assessing the functional capacity of each individual centromeric array to recruit the centromere
35 epigenetic marks CENP-A and CENP-B, which are proteins necessary for centromere

formation and function. During the development of embodiments of the technology provided herein, data were collected that indicated that centromeric genetic sequence instability is associated with trisomy 21. The technology provided herein thus provides a rapid assay to study centromere variation and evolution in normal biology and in
5 disease.

Uses

In some embodiments, the centromere-specific assays described herein are used to study the centromeres of human chromosomes comprehensively, e.g., in rapid and
10 simultaneous reactions. In some embodiments, the technology finds use in determining the content of α -repeats in one or more centromere arrays, thus providing a technology to determine the size and variation of these arrays, their variation in human populations, and the changes they have undergone over an evolutionary timescale. In contrast to previous techniques used to estimate the size of single centromere arrays,
15 embodiments of the technology described herein provide an improved technology for studying human centromere sequences quickly and more efficiently in a given biological setting.

Human centromere markers find use in understanding the role centromere sequences and their interactions with epigenetic factors play in centromere formation,
20 kinetochore assembly, chromosome segregation, and cell division. Data collected from experiments conducted during the development of embodiments of the technology provided herein indicated that CENP-A binds (e.g., almost exclusively) to the larger α -repeat array in each centromere; and, on those arrays where CENP-A binds, CENP-A loads onto only a portion of the centromere array, confirming and expanding previous
25 observations (Maloney et al., 2012; Aldrup-MacDonald and Sullivan, 2014; Ross et al., 2016). Interestingly, in contrast to centromeric arrays in somatic chromosomes, CENP-A occupies a vast extent of DYZ3 in chromosome Y, an array lacking CENP-B boxes and that thus does not recruit CENP-B protein. Embodiments of the technology described herein comprise ChIP-PCR assays that find use in assessing the functionality of specific
30 human centromere arrays, e.g., to study the recruitment of epigenetic modulators that facilitate centromeric assembly in different biological settings and to study instability in centromere sequences that alters epigenetic regulation and centromere function.

The technology described herein comprising use of centromere markers for genetic studies finds use in clinical settings. Clinical uses provide improved technologies
35 relative to extant methods that rely on detecting single copy genes (Koumbaris et al.,

2016) because the present technology detects contiguous markers present in thousands of copy numbers, thus improving the sensitivity of detection of specific chromosomes. These markers find use in studying ploidy in human samples without the need for karyotyping or requiring live cells. For example, embodiments of the technology use
 5 centromere markers described herein to test samples to determine gender or to identify the presence of nondisjunction chromosome defects. The centromere markers described herein find use in understanding developmental defects associated with centromere instability, e.g., developmental defects such as trisomy 21. Further, these markers find use in studying the pathogenesis of congenital nondisjunction defects, tracking where
 10 and when defects in centromeres appear, such as meiosis of germ cells or after fertilization. Embodiments of the technology find use in detecting centromeric instability, which is causative of birth defects and cancer.

In conclusion, embodiments of the technology described herein provide rapid and user-friendly methods to study the genetics and epigenetics of human centromeres. The
 15 data collected from the experiments described herein indicated that embodiments of the technology find use in sensitively detecting and/or quantifying chromosomal ploidy and/or detecting centromeric and pericentromeric instability that are associated with and/or are causative of some trisomies, especially trisomy 21. The data also indicated that embodiments of the technology find use in detecting the transcription of α -repeats
 20 under specific conditions and in identifying the chromosomes from which the α -repeats are transcribed. Further, embodiments of the technology find use in specific and genome-wide approaches to study centromere epigenetics, e.g., embodiments of PCR-based techniques described herein characterize specific centromeres in “real-time” and thus provide data characterizing specific centromeric genomic elements and their
 25 association in development and the evolution of malignancy. In addition, the technology finds use in assessing, characterizing, or evaluating the nature or quality of stem cells, e.g., induced pluripotent stem cells.

Kits

30 Embodiments provide kits comprising a composition described herein, e.g., a kit for producing an amplicon from a human chromosome (e.g., from an α -repeat array). For example, some embodiments comprise one or more pairs of oligonucleotides specific for one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1

and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3.

Although the disclosure herein refers to certain illustrated embodiments, it is to be understood that these embodiments are presented by way of example and not by way of limitation.

Examples

Experimental Procedures

DNA samples

10 DNA samples from human/rodent somatic hybrid cell lines (each containing a single human chromosome) and their parental rodent cells were obtained from a human chromosomal DNA mapping panel (NIGMS Human/Rodent Somatic Cell Hybrid Mini Mapping Panel # 2 DNA, Coriell Cell Repositories). DNA samples from people of different genders and ethnic origins were obtained from the HRC2 Human Random Control DNA panel 2 (Sigma Aldrich, 96 Caucasian people) and from Human variation panels HD03 (Indo Pakistani), HD05 (Middle Eastern), HD07 (Japanese), HD12 (Africans South of the Sahara), HD20 (Russian Krasnodar), HD21 (Italian), HD22 (Ashkenazi Jewish), HD32 (Chinese), and samples from Mbuti Pygmy (NA10492, NA10493, NA10494, NA10495, NA10496) from Coriell Cell Repositories. DNA samples from individuals with trisomy 8 (Catalog numbers: NA00496, NA00425, NA02030, GM02596, and GM04610), trisomy 13 (Catalog numbers: AG12070, NA00526, NA02948, NA03330, and NA00503), trisomy 18 (Catalog numbers: GM03538, GM03769, NA02732, NA03623 [also trisomy X]), trisomy 21 (AGPDOWN Aging DNA panel-down syndrome), and an XY who is phenotypically female (Catalog number: NA02598) were obtained from the Coriell Cell Repositories. Human tissue from trisomy 21 individuals was also obtained from the NICHD Brain and Tissue Bank for Developmental Disorders at the University of Maryland, Baltimore, MD (Catalog numbers: UMB4904, UMB1258, UMB0707, and UMB5277).

RNA samples

RNA was isolated from the prostate epithelial cell lines RWPE-1, PNT2, 957E-hTERT, and the prostate cancer cell lines PC3, VCaP, LNCaP, and DU145 using Trizol reagent (ThermoFischer Scientific) as recommended by the manufacturer. The authenticity of these cells lines was verified by genotyping. Isolated RNA was treated with the Turbo

DNA-free kit (Ambion) to remove contaminant traces of DNA, which was subsequently verified by PCR.

Search for centromere specific markers

Systematic analyses of literature and sequence databases were performed to identify nucleotide mutations in each centromere array that serve as markers specific for the centromere arrays of each human chromosome. The markers identified herein supplement pericentromeric sequence markers previously identified (Contreras-Galindo et al, 2011, 2013; Zahn et al, 2015, incorporated herein by reference).

In particular, a systematic and comprehensive analysis of the literature and sequence databases was performed to annotate centromere sequences and to identify unique centromere markers for human chromosomes. Several sequences, identified in previous studies by Southern blotting analysis, have been recently annotated and expanded to the most recent assembly of the human genome project (Hg38) (see, e.g., Miga et al., 2014; Miga, 2015). Markers for pericentromeric sequences were identified previously. These pericentromeric markers, proviruses K111 and K222, exist in 15 and 9 human centromeres, respectively (Contreras-Galindo et al., 2011; Contreras-Galindo et al., 2013; Zahn et al., 2015). Hundreds to thousands of these proviruses exist in human pericentromeres, yet K111 accumulated in a particularly high copy number in the centromere of chromosome 21, whereas K222 accumulated in the centromere of chromosomes 13 and 14. The accession numbers of the sequences analyzed in this study are indicated in the Table 2. Due to the sequence similarity between α -repeats, primers were designed and validated to detect each array specifically. Primers and probes to detect proviruses K111 and K222 at pericentromeric loci specifically are as previously described (Contreras-Galindo et al., 2011; Contreras-Galindo et al., 2013; Zahn et al., 2015, each of which is incorporated by reference in its entirety).

Real-time PCR for human centromere arrays

The numbers of α -repeats in centromere arrays of all human chromosomes, pericentric proviruses K111 and K222, and single copy genes in human DNA were measured by qPCR using the primers and conditions described herein. Centromere quantification was performed in DNA isolated from healthy individuals of different gender, diverse ethnicities, and individuals with trisomy 8, trisomy 13, trisomy 18, and trisomy 21.

In particular, copy numbers for each centromeric array (e.g., the number of α -repeats in each array), proviruses K111/K222, and single copy genes were measured by

qPCR using specific primers and PCR conditions as described in Tables 1 and 2. PCR amplification products were confirmed by sequencing. The qPCR was carried out using the FastStart Universal SYBR Master mix (Rox) (Roche) with an initial enzyme activation step of 95°C for 10 minutes, followed by 16-25 cycles of 15 seconds of denaturation at 95°C and 30 seconds of annealing/extension at the temperatures reported in Table 2. The copy number was estimated using serial dilutions of plasmids containing the PCR amplicon. The DNA copy number of each plasmid was estimated by reading the DNA concentration of the plasmid at a wavelength of 260 nm using UV spectrophotometry and estimating the number of copies using this information and the size of each plasmid. The specificity of the qPCR assay detecting the centromere of unique chromosomes was assessed using DNA samples from human/rodent cell hybrids, each one containing a single human chromosome. PCR quantification of K111 and K222 pericentromeric proviruses was performed as described previously (Contreras-Galindo et al., 2011; Contreras-Galindo et al., 2013; Zahn et al., 2015; each of which is incorporated herein by reference in its entirety). Quantification of the single copy genes *top3a*, *dek*, and *ccr5* was performed by qPCR using 40 cycles of 15 seconds of denaturation at 95°C and annealing/extension of 30 seconds at 60°C. The copy number of single copy genes was calculated using serial dilutions of purified plasmid containing the target PCR amplicon as described above. The relative copy number of α -repeats in each array per human diploid genome was estimated in reference to the quantification of the gene *top3A*, which exists as a single copy in the human genome (Hanai et al., 1996). The relative copy number was calculated by dividing the number of copies obtained by qPCR by the number of copies detected using Top3A in equal amounts of cellular DNA.

Real-time qPCR for centromeres 13 and 21 using LNA primers and clamps

The homolog arrays D13Z1/D21Z1 present in the centromere of chromosomes 13 and 21 are almost identical, except for the presence of two nucleotide substitutions present in the α -repeats of the D13Z1 and D21Z1 array (Pellestor et al., 1994; Nilsson et al., 1997). The D13Z1/D21Z1 arrays have the same sequences except that a T in the sequence of D13Z1 is a C in the sequence of D21Z1 and an A in the sequence of D13Z1 is a G in the sequence of D21Z1 (see Fig. 10).

In particular, a qPCR assay was developed to detect these substitutions using primers that have a locked nucleic acid (LNA) modification targeting these nucleotide variations (Ballantyne et al., 2008). The primers also contained LNA modification at the bases right before and after the substitution. The description of the primers is provided

in Table 1. LNA primers offer markedly increased affinity for the complementary strand relative to traditional, unmodified DNA primers. These modified primers were used in a PCR reaction to detect either D13Z1 or D21Z1 specifically. DNA isolated from human/rodent somatic cell hybrids that contain either chromosome 13 or 21 was used to develop conditions for the assay. The PCR assay was carried out as described above, e.g., comprising an enzyme activation step at 95°C for 10 minutes and 20 cycles of 15 seconds of denaturation at 95°C and 30 seconds of annealing/extension at the temperatures reported in Table 2. A temperature gradient showed that at increasing annealing temperatures, the LNA primers uniquely amplified the centromere repeat containing the nucleotide variation. For centromere 13 D13Z1, a forward LNA primer and a regular reverse oligonucleotide (that binds to either D13Z1 or D21Z1; see Figure 10) exclusively detected the centromere of chromosome 13 but not 21 at an annealing/extension temperature of 68°C (see Figure 11A). For the detection and quantification of the centromere 21 D21Z1, forward and reverse LNA primers were used that recognize the D21Z1 nucleotide variations together with an LNA primer clamp that recognizes the substitution of D13Z1. This clamp comprised the same forward LNA primer that detects D13Z1 but is phosphorylated at the 3' end to inhibit the amplification of D13Z1 sequences (Skronski et al., 2013). The LNA PCR reaction specifically amplified D21Z1 and substantially reduced amplification of D13Z1 when an annealing/extension temperature of 64°C was used (See Supplementary Figure 11B). The specificity of the LNA primers was evaluated in human/rodent hybrid cell lines that contain a single human chromosome as shown in Figure 10. An LNA primer was also designed to detect specific mutations in the pericentromeric K111 *gag* and not other HERV-Ks. The specific detection of K111 was verified by sequencing analysis of the PCR product.

Real-Time qRT-PCR

Real-time RT-PCR was performed using the same qPCR conditions described above to quantify human centromere arrays with the addition of an RT step. The PCR reaction contained 0.2 microliter of the Murine Leukemia Virus Reverse Transcriptase (MLV-RT) and was preceded by an RT step of 30 minutes at 50°C. DNA contamination was ruled out by conducting PCR on samples that were not primed with MLV-RT.

Chromatin immunoprecipitation

ChIP assays to assess the association of centromeric proteins CENP-A and CENP-B with centromere core and pericentric sequences were performed using antibodies to the

centromere proteins CENP-A and CENP-B, or with nonspecific IgG antibodies, on chromatin extract from the prostate cancer cell line LnCaP.

In particular, ChIP assays were performed using the iDeal ChIP-seq kit for Histones (Diagenode) following the procedures described by the manufacturer to assess the association of centromeric proteins CENP-A and CENP-B with centromere repeats. Briefly, approximately 70% to 80% confluent LnCaP cells grown in 75 cm² flasks were fixed with 0.1% formaldehyde to cross-link protein to DNA. Cells were lysed and the chromatin was sheared to 100 to 600 bp with a sonicator. Chromatin was immunoprecipitated overnight using a specific monoclonal antibody to CENP-A (Abcam, ab13939), a polyclonal antibody to CENP-B (Abcam, ab134144), or with non-specific IgG antibodies. Centromere protein occupancy on target arrays, indicating the number of α -repeats in each centromere array sequence bound to the centromeric proteins CENP-A or CENP-B, was measured by qPCR using the qPCR assays for centromeric repeats described herein. The relative amount of immunoprecipitated DNA compared to input DNA was calculated using the equation $\% \text{recovery} = 2^{(C_{\text{input}} - C_{\text{sample}})}$, assuming that the efficiency of the PCR reaction was 100%. The relative percentage of each array occupied by CENP-A and CENP-B was calculated by dividing the number of α -repeats in each centromere array precipitated with centromeric proteins by the total number of α -repeats in the array determined in the input DNA. The increase in enrichment (e.g., expressed as a “fold” or “ $\times$ ” enrichment) was also determined based on the cycle differences (ΔC_t) between the sample relative to control (IgG). The genes *top3a*, *dek*, and *β -actin* served as negative controls in the centromere ChIP studies, as these genes localize in chromosomal arms.

Amplification and Sequencing of K111 LTR insertions

K111 insertions were amplified by PCR using the Expand Long Range dNTPack PCR kit (Roche Applied Science, Indianapolis, IN) as previously described (Contreras-Galindo et al., 2013). The amplification products were cloned into the topo TA vector (Invitrogen, Carlsbad, CA) and sequenced. Sequences of K111-related insertions amplified from DNA of human/rodent cell hybrids containing a single human chromosome shown in the Figure 10 are deposited in the NCBI database with Accession Numbers (JQ790790-JQ790967). The primers P1/P4 amplify K111 insertions in several human chromosomes.

Bioinformatics analysis

The number of copies of α -repeats in each centromeric array, the numbers of pericentromeric K111 and K222, and the number of single copy genes *top3a*, *ccr5*, *dek*, *renin*, *gapdh*, and *β -actin*, were analyzed in sequence libraries from diverse populations generated in the 1000 genomes project. The 101-bp read libraries were screened for sequences that matched each query sequence, allowing no more than 10-bp mismatches and/or indels. For pericentromere K111 proviruses, sequences were identified that match to the 5' and 3' integration sites and adjacent flanking sequence. Using these criteria provided for the specific identification of K111 as opposed to other HERV-K proviruses that exhibit alternative integration sites. Detection criteria were constrained to identify reads containing at least 20 bp of flanking sequence (CER: centromere repeat elements), 20 bp of the K111 LTR, and the GAATTC target site duplication as previously described (Contreras-Galindo et al., 2013). For the K222 proviruses, reads were identified that have 20 bp of the CER flanking sequence and 20 of the adjacent *prt* unique to K222 proviruses (Zahn et al., 2015). Additionally, libraries were screened to detect single copy genes using the same criteria used to identify centromeric arrays. Values were normalized to the total number of sequence reads in each group by dividing the number of sequence reads that hit the query sequences by the total number of reads, then multiplying these values by one million of sequence. The calculated data were compared to the number of reads that match to the genes *top3a*, *dek*, *ccr5*, *renin*, *gapdh*, and *β -actin* using the same exercise to contrast the abundance of centromere arrays to that of single copy or small-number-of-copy genes. To screen for sequence variants specifically present in the centromere 13 repeat D13Z1 or centromere 21 repeat D21Z1, arrays that are almost 100% similar, Illumina sequencing libraries were screened from populations of diverse ethnicities obtained by the Human Genome Variation Project (HGVP), the most recent human genome assembly (Hg38), sequencing reads obtained from Denisovan and Neanderthal sequencing projects, and sequence read archives (SRAs) from populations of gorillas, chimpanzees, and orangutans. Reads were identified that were 100% identical to a 38-bp query sequence that has the substitution T/C (T or C) present in either D13Z1 or D21Z1, respectively.

In silico sequence analysis

The K111-related LTR sequences obtained in the DNA of human cells, trisomy 21 samples, and DNA from human/rodent chromosomal cell hybrids were BLASTed against the NCBI database. The sequences were aligned in BioEdit and exported to the MEGA 7

matrix. The K111 tree was generated using Bayesian inference (MrBayes v 3.2; Huelsenbeck et al. 2001; Ronquist and Huelsenbeck, 2003) with four independent chains run for at least 10,000,000 generations until sufficient trees were sampled to generate more than 99% credibility.

5

Statistical analysis

Statistically significant differences between the numbers of α -repeats in each centromere array and K111 and K222 pericentromeric proviruses between samples were calculated using Student's t-test. The relative enrichment of centromeric array DNA associated with the centromere proteins CENP-A and CENP-B in ChIP experiments (using IgG as a control antibody) was compared using Student's t-test. Correlations calculated by qPCR, qRT-PCR, *in silico* analysis of the 1000 genomes project, and by the estimated average number reported in the literature were measured using Spearman's correlation test. Two-tailed *P* values were considered significant at *P* < 0.05.

15

Example 1 – Assays of human centromeres

To develop specific markers for centromeric arrays and to study their structure, function, and evolution, experiments were conducted in which a systematic analysis identified specific DNA variations in every α -repeat array. Based on this analysis, amplification primers were designed, e.g., for PCR assays. For example, embodiments of the technology comprise rapid qPCR assays for these unique markers, which provide for assessing the abundance of α -repeats in each array of the centromeres of human chromosomes (See, e.g., Figure 1). During the development of the technology described herein, experiments were conducted that used DNA from somatic rodent/human hybrids. In these experiments, each hybrid contained one single human chromosome to verify the specificity of the PCR detection. The description of the primers and probes used in this study is provided in Tables 1 and 2.

25

Table 1 – primers

Primer name	Primer nucleotide sequence (5' to 3')	SEQ ID NO:
D1Z5Fb	GAGAATTTTCGTTGGAAACGGATAAAACC	1
D1Z5Rb	ATCCACTTGCAGATACTACGAAA	2
D1Z5Fc	GGCCTATCGTCGTAAAGGAAATA	3
D1Z5Rc	ATGCTCAGCTCTGTGAGTTAAA	4
D1Z7/D5Z2	GTTCCCTTAGACAGAGCAGATTT	5
D1Z7/D5Z2	CAACGCAGTTTGTGGGAATG	6

D2Z1F	TCGTTGGAAACGGGATTGT	7
D2Z1R	CTGCTCTATGAAAGGGACTGTT	8
D3Z1F	GCTTTGAGGCCAATGGTAGA	9
D3Z1R	GTTGAACACACACGTACCAAAG	10
D4Z1F	CTGTAGTATCTGGAAGTGGACATT	11
D4Z1R	GGTTCAACTGTGTTCGTTTAGG	12
D5Z1	AGTCTGCACGTGGATAAGTTG	13
D5Z1	AAAGAGTGTTAGAGTCTGCTCTG	14
D6Z1F	ACGGAAGCAATCTCAGAACTAC	15
D6Z1R	CCTCAAGGCGGTCCAATTAT	16
D7Z1F	GTGGATATATGGACCGCATTGA	17
D7Z1R	CACACAGCACAAAGAAGTTACTG	18
D7Z2F	CGACTTTGTGATGTGTGCATTC	19
D7Z2R	CCTTATCCGCAATGGTCCTAAA	20
D8Z2F	ACGTACACAGCAGCATACTC	21
D8Z2R	GTCCACTTCCAGATACTCCAAA	22
D9Z4F	GGAGAAGCATTCTCAGGAAGTT	23
D9Z4R	GTCCGCTTGCAGATACTACAG	24
D10Z1F	TTGGAAACGGGATTTCCCTCATA	25
D10Z1R	GCTCTCTCTAAAGGAAGGTTCAA	26
D11Z1F	CTTCCTTCGAAACGGGTATATCT	27
D11Z1R	GCTCCATCAGCAGGATTGT	28
D12Z3F	GATGAAGGAGTTTGGAGACACT	29
D12Z3R	CTGTGCAACATTACAGGAAGAAATC	30
D13Z1F LNA	GTGATGTGTGTACCCAG+C+T+AAA	31
D13Z1/D21Z1R	AACGAAATCCTCCAAGCTATCC	32
D13Z1F LNA Clamp	GTGATGTGTGTACCCAG+C+T+AAA/3Phos/	33
D13Z1F	TGATGTGTGTACCCAGCT	34
D13Z1R	GCTATCCAAATATCCACT	35
D13Z2F	CAGAGGGCTTTGTGGAGTATAG	36
D13Z2R	ACGTTGAATGCACACATCAC	37
D13Z3F	GGAATATTTGTGAGCCCATTGAG	38
D13Z3R	TTCAACCCTGTGAGTAGAAGTC	39
D13Z6F	CTTCGTGCGAAACGGGAATA	40
D13Z6R	CAAACCTGCTCTATCAACAGAAAGG	41
D13Z7F	ACTTCTTTGTGATGTGTGCATTC	42
D13Z7R	AGGGCTCCAAATATCCAGTTG	43
D13Z8F	TCTCACAGAGTTGAACCTATCTTATG	44
D13Z8R	TTTCCACAGTAGGCCTCAAAG	45
D13Z9F	TGTGATGCGTGACTCATCTTAC	46
D13Z9R	GCCTCAAGGCGCTCAAA	47
D14Z1/D22Z1F	CAATCTCAGAATCTTCTTTGGGATA	48
D14Z1/D22Z1R	CCAAGCTATCCAAATATCCACTT	49
D14Z2F	GATTTGTTGGAAACGGGATTAC	50

D14Z2R	AGAAAGATCCACGCCTGTTA	51
D14Z3/D22Z6F	GTGCGTGCATTTCATGTCATAG	52
D14Z3/D22Z6R	GTCCTCAAAGGGCTCCAATTA	53
D15Z3F	GACATTTGGATAGCTTTGAGGATTT	54
D15Z3R	GGGCATACATCACAAGGAAGA	55
D16Z2F	CCTTCCTTTAGACAGAGCAGATT	56
D16Z2R	AGTGTCTGGGAACGAGTTTG	57
D17Z1F	GTGGAGATATGGACCGCTTTAG	58
D17Z1R	CTCAGTCGTCACCAAGAGTTT	59
D17Z1bF	TTTCGTAGGGTCTGCAAGTG	60
D17Z1bR	CCGACAATGCTTCTCTCTAGTT	61
D18Z1F	TGGGAAACGGGATTGTCTTC	62
D18Z1R	CTGCTCTACCAAAGGGAATGT	63
D18Z2F	TTCGATCGATTTTCAGGCCTATG	64
D18Z2R	TTGAGGACACACATCACAAAGA	65
D19Z4F	CTAAAGACCTCAATGGGCTCAG	66
D19Z4R	GTGGATTCATCTCACAGACGTTA	67
D19Z5F	GCCTCAATGGGTTTCAGAAATG	68
D19Z5R	TGGATCCATCTCACAGATTTC	69
D20Z2F	TGCTTGGAACGGGAATGT	70
D20Z2R	CCTGCTCTATGAAAGGGAATGT	71
D21Z1F	TGATGTGTGTACCCAGCC	72
D21Z1R	GCTATCCAAATATCCACC	73
D21Z1F LNA	GTGATGTGTGTACCCAG+C+C+AAA	74
D21Z1R LNA	CCTCCAAGCTATCCAAATATCC+A+C+TTGCA	75
D22Z4F1	CATTCTGACGTGGGCATTAAAC	76
D22Z4R1	GTAGGCCTCAAAGGCTAGAAAT	77
D22Z4F2	TGACCAGTTTGGGAACATTCT	78
D22Z4R2	AGGAGTCTGCCTGTCTAGTT	79
D22Z5F	TCATCGCACAGAGTGAAACC	80
D22Z5R	GCCATAAAGGGCTCACAAATATC	81
DXZ1F	CGGGATCACCTTCCCATAAC	82
DXZ1R	GGTGTGCAAACCTGAACTATC	83
DYZ3F	CTGCAAGTGGAATTGGGAAATA	84
DYZ3R	GAATGCACACAGCACAAAGAA	85
p82HF	ATGTTTGCATTCAACTCACAGAG	86
p82HR	CAACACAGTCCAAATATCCAGTTG	87
CCR5F	CAATGTGTCAACTCTTGACAGG	88
CCR5R	ACCTGCATAGCTTGGTCCAACC	89
DEK exon 3F	TCCTATCCACGTACTTCAGGCTG	90
DEK exon 3R	GCCTGGCCTGTAGTAAAGCAGTTT	91
b-actinF	CCCTGAGGCACTCTTC	92
b-actinR	GTGCACGTCACACTT	93
GAPDHF	CCACTCCTCCACCTTTGAC	94

GAPDHR	ACCCTGTTGCTGTAGCCA	95
Top3AF	ACTAGGTCAGAGACCCCTTACTG	96
Top3AR	CAAGGAGAGGCAGTGACAAA	97
K111F	AAGAGCACCAGGATGCTTAATGCC	98
K111R	AGTGACATCCCGCTTACCATGTGA	99
K111P	FAM-TGCCGGTCCTAACAGTAGACTCAC-BHQ1	100
K222F	CAGCGTTCTGGAATCCTATGT	101
K222R	TGTATTGTGGTAACTGGGTATATGT	102
K222P	FAM-ACCCACATGGCAGTGTCTGGATT-BHQ1	103
K111 1584F	TCCTTAAGGTCATAGTGGAGTTGTTGGTATAC	104
K111 1707R LNA	CATAAGCATAGCTTTATG+C+A+AAC	105
P1	ACATTTCAGACCATGGTAGCCGTGT	106
P4	GTACCTTCACCCTAGAGAAAAGCCT	107

In Table 1, a “+” indicates that the previous base is a locked nucleic acid analog base in which the ribose ring is “locked”, which increases the hydrogen bond strength of the LNA base and consequently provides a higher melting temperature upon binding.

- 5 “FAM” indicates a “FAM” fluorogenic dye linked to the primer; “BHQ1” indicates a BLACK HOLE QUENCHER 1 quencher moiety linked to the primer. “/3Phos/” indicates that the primer comprises a 3’ phosphate.

In some cases, the amplification assays detected arrays previously reported to exist in a unique chromosome on another chromosome, albeit with low levels of detection (see, e.g., Table 2). This was particularly interesting in chromosome Y, where the assays detected α -repeat arrays from chromosomes 1, 4, 5, 7, 9, 13, 17, 18, and 21. The existence of these α -repeat arrays in chromosome Y was confirmed by sequencing the PCR products, indicating that the centromere of chromosome Y exchanged material with the centromeres of somatic chromosomes during human evolution.

Table 2 – Assays for detecting arrays specific for each human chromosome

α -repeat array	Accession number	Chr. reported	Chr. detected	Chr. detected (optimized)	Conditions
D1Z5-1	BX248407.26	1	1	1	25, 60°C
D1Z5-2	BX248407.26	1	1	1	25, 60°C
D1Z7/D5Z2/D19Z3	AJ295044.1	1,5,19	1,5,19,22,Y	1,5	16, 62°C
D2Z1	J04773.1	2	2	2	20, 60°C
D3Z1	Z12006.1	3	3	3	20, 60°C
D4Z1	Z12011.1	4	4,Y	4	16, 62°C
D5Z1/D19Z2	M26920.1	5,19	5,Y	5	16, 60°C
D6Z1	GJ211907.1	6	6	6	20, 60°C
D7Z1	AC142529.3	7	7,Y	7	18, 62°C
D7Z2	M16037.1	7	7,Y	7	20, 61°C
D8Z2	M64779.1	8	8	8	25, 60°C

D9Z4	M64320.1	9,4	9,4,Y	9	16, 62°C
D10Z1	X63622.1	10	10,12,16,22	10	18, 62°C
D11Z1	M21452.1	11	11	11	25, 60°C
D12Z3	M28221.1	12	12	12	16, 56°C
D13Z1/D21Z1	D29750.1	13,21	13,15,21,Y	13	20, 61°C
D13Z1 LNA	D29750.1	13,21	13	13	20, 68°C*
D13Z2	GJ211955.2	13,14,21,22	13,14,21,22	13,14,21,22	25, 60°C
D13Z3	GJ211961.2	13,14,21,22	13,14,21,22	13,14,21,22	25, 60°C
D13Z6	GJ211965.2	13,14,21,22	13,14,21	13,14,21	25, 60°C
D13Z7	GJ211967.2	13,14,21,22	13,14,21,22	13,14,21,22	25, 60°C
D13Z8	GJ211968.2	13,14,21,22	13,14,21,22	13,14,21,22	25, 60°C
D13Z9	GJ211969.2	13,14,21,22	13,14,21,22	13,14,21,22	25, 60°C
D14Z1/D22Z1	M22273.1	14,22	14,22	14,22	20, 60°C
D14Z2	GJ211972.2	13,14,21,22	14	14	25, 60°C
D14Z3/D22Z6	GJ211986.2	13,14,21,22	14,22	14,22	25, 60°C
D15Z3	AF237720.1	15	15,8,22	15	22, 64°C
D16Z2	M58446.1	16	16	16	25, 60°C
D17Z1	M13882.1	17	17,22,Y	17	20, 61°C
D171b	GJ212053.1	17	17	17	25, 60°C
D18Z1	M65181.1	18	18,Y	18	18, 64°C
D18Z2	M65182.1	18	18,Y	18	20, 62°C
D19Z4	AC004164	19	19,22	19,22	25, 60°C
D19Z5	AC006504	19	19,22	19,22	25, 60°C
D20Z2	X58269.1	20	20	20	18, 60°C
D21Z1/D13Z1	D29750.1	13,21	13,15,21,Y	13,15,21	20, 62°C
D21Z1LNA/D13 _{clamp}	D29750.1	13,21	13,21,Y	21	20, 64°C*
D22Z4-1	GJ212162.2	22	22	22	25, 60°C
D22Z4-2	GJ212162.2	13,14,21,22	22	22	25, 60°C
D22Z5	GJ212163.2	13,14,21,22	22	22	25, 60°C
DXZ1	X02418.1	X	X	X	25, 60°C
DYZ3	GJ212193.1	Y	Y	Y	25, 60°C
P82H	Mitchell 2015	All	All	All	25, 60°C

In Table 2, “Chr.” is an abbreviation for “chromosomes”; “Chr. detected” refers to the chromosomes detected by amplification using a non-optimized PCR assay; “Chr. detected (optimized)” refers to the chromosomes detected by amplification using an optimized PCR assay; and “Conditions” indicates the number of cycles and the annealing temperature used for the amplification assay.

Using some embodiments of the technology described herein, identifying specific markers was challenging for certain human chromosomes with extremely similar centromeric sequences. The centromere sequences of chromosomes 1, 5, and 19 have 99% to 100% similarity. Similarly, the pairs of chromosomes 4 and 9; 13 and 21; and 14 and 22 also share centromere sequences with 99-100% similarity (Visel and Choo, 1992; Hayden, 2012; Miga, 2015). Accordingly, efforts were directed to develop embodiments of the technology that target particular nucleotide mutations and to identify amplification conditions for qPCR assays that are chromosome-specific even for the similar α -repeat arrays in chromosomes 1, 4, 5, 9, and 22. Furthermore, embodiments of the technology

were developed by screening annotated α -repeat arrays in the human genome assembly Hg38 for centromeres 13, 14, 21, and 22 (Miga, 2014, 2015). Based on this analysis, α -repeat arrays were identified that are unique to either centromere 14 or 22 (Fig. 2), which provides a specific assay for the arrays of chromosomes 14 and 22.

During the development of embodiments of the technology provided herein, amplification assays (e.g., qPCR assays) were developed to study α -repeat arrays for every human chromosome except chromosome 19. Most of the PCR assays target α -repeat arrays present in centromeric cores (e.g., D1Z7/D5Z2, D2Z1, D3Z1, D4Z1, D5Z1, D6Z1, D7Z1, D7Z2, D8Z2, D9Z4, D10Z1, D11Z1, D12Z3, D13Z1/D21Z1, D14Z1/D22Z1, D15Z3, D16Z2, D17Z1, D17Z1b, D18Z1, D18Z2, D20Z2, DXZ1, and DYZ3). In addition, during the development of embodiments of the technology provided herein, amplification assays (e.g., qPCR assays) were developed that are specific for pericentromeric regions. For example, some embodiments of the assays target pericentric arrays (e.g., D1Z5, D13Z2, D13Z3, D13Z6, D13Z7, D13Z8, D13Z9, D14Z2, D14Z3, D19Z4, D19Z5, D22Z4, and D22Z5). In some embodiments, the assays detect the array p82H that is present in all human centromeres (Mitchell et al., 1985; Alexandre et al., 1987).

Experiments identified the pericentric endogenous retroviruses K111 and K222 (Contreras-Galindo et al., 2011, 2013; Zahn et al., 2015). K111 exists in the pericentromeres (e.g., at the centromere/pericentromere border, e.g., based on data indicating that K111 associates not only with CENP-A and CENP-B, but also with H3K9 trimethylated chromatin) of fifteen human chromosomes and K222 exists in the pericentromere of nine human chromosomes. During the development of the technology provided herein, amplification assays were developed to quantify the abundance of these proviruses. Consequently, the amplification assays find use in some embodiments to study the variation and function of pericentric regions, e.g., to assess specific changes at the centromeric core and at the pericentromere in most human centromeres.

Example 2 – Validation of assays to characterize centromeres

To evaluate the accuracy of embodiments of the technology provided herein, assays described herein were used to determine the size of centromere arrays in DNA isolated from peripheral blood lymphocytes of 5 individuals (Fig. 3). Data collected from the assays indicated that arrays at the centromere core are larger than arrays at pericentromere loci. The assays also indicated that the lengths of arrays vary in these individuals. Estimates of the average sizes of each array in these subjects are shown in Table 3. Interestingly, the sizes of these arrays do not correlate significantly with the

sizes of the chromosomes. For example, centromere arrays in smaller chromosomes, such as chromosomes 18 and 20, are larger than arrays in bigger chromosomes, such as chromosomes 4 and 5.

5

Table 3 – Sizes of human centromere arrays

Centromere Repeat	Real Time PCR (copies/genome)	Illumina 1000 Genomes (copies/genome)	Estimated average size (copies/genome)	Reference
D1Z5	680	749	2573	Liehr, 2013
D1Z7	ND	32170	11695	Liehr, 2013
D2Z1	35047	26388	36842	Liehr, 2013
D3Z1	17137	22880	19923	Liehr, 2013
D4Z1	949	10919	18713	Liehr, 2013
D5Z1	8060	22710	26900	Liehr, 2013
D6Z1	8354	21373	17543	Liehr, 2013
D7Z1	69502	15905	22280	Liehr, 2013
D7Z2	48	1090	584	Liehr, 2013
D8Z2	22145	16797	14912	Liehr, 2013
D9Z4	20720	2657	15789	Liehr, 2013
D10Z1	14103	9030	12865	Liehr, 2013
D11Z1	39225	5853	27836	Liehr, 2013
D12Z3	205	9598	8187	Liehr, 2013
D13Z1/D21Z1	26495	16696	13450	Liehr, 2013
D14Z1/D22Z1	74037	2070	13450	Liehr, 2013
D15Z3	268	10145	14619	Liehr, 2013
D16Z2	14731	30258	11695	Liehr, 2013
D17Z1	916	12596	15789	Liehr, 2013
D18Z1	47908	19852	7953	Liehr, 2013
D18Z2	2858	2020	9941	Liehr, 2013
D19Z4	34	15	45	Lamerdin, unpublished
D19Z5	1	1	11	Lamerdin, unpublished
D20Z2	95484	12039	5964	Liehr, 2013
D21Z1	9342	7979	10438	Liehr, 2013
D22Z4	9	161	316	Liehr, 2013
DXZ1	6780	5769	22262	Liehr, 2013
DYZ3	526	931	1328	Liehr, 2013
K111	24474	47	2000	Contreras, 2013
K222	24	5	50	Contreras, 2015
p82H	32864	0	28	Mitchell, 1985
Top3A	2	2	2	Contreras, 2015

In Table 3, the numbers of α -repeats in each centromere array per diploid genome were determined by real time PCR according to the technology provided herein (column 2), as measured by bioinformatics analysis retrieved from the 1000 Genomes project (column 3), or as estimated by previously published reports (column 4).

5 A reference data set was constructed in silico using sequences generated by the 1000 Genomes project as described in the Examples below (Fig. 4). Comparing the in silico results with the amplification assay indicated that the data collected during the development of embodiments of the technology reflected the trends observed in silico (Fig. 5). In particular, the PCR data indicated that the average sizes of the centromere
10 arrays are: 1) larger in core arrays than in pericentric arrays; 2) vary in human populations; and 3) do not correlate with the chromosome size (Table 3). Importantly, the data indicated a correlation between the average numbers of α -repeats in each array as determined by PCR and by in silico analysis (Fig. 5A and 5B). Data analysis was also performed to correlate the numbers of α -repeats detected by amplification to the average
15 number of α -repeats in each array determined by Southern blotting analysis as previously reported (Liehr, 2013; Table 3). The analysis indicated a positive correlation (Fig. 5C), suggesting that the PCR assays rapidly and accurately determine the size and variation of human centromeres and correlate well with Next Generation Sequencing (NGS) data and Southern blotting analysis.

20 These data and analyses indicate that the present technology provides improvements and benefits relative to extant technologies previously reported in the literature (e.g., Southern blotting, NGS). For example, studying centromere arrays by Southern blotting requires laborious and time-consuming techniques that often comprise use of radioactive compounds and large amounts of DNA. Furthermore,
25 Southern probes often hybridize non-specifically to other centromere arrays, which causes such a technique to yield inaccurate results. Further, Southern blotting is limited to studying a single array and estimating array size within a window of variation between approximately 6000 to 24,000 α -repeats (Liehr, 2013). Accordingly, the Southern blot technique is unreliable for accurately estimating the size of centromere
30 arrays.

In contrast, embodiments of the technologies described herein provide an assay that can measure the full size of each array in a given individual with an associated variation in size that is as low as zero and is at most approximately 1800 α -repeats.

While extant NGS methodologies, analysis of databases (e.g., 1000 Genomes
35 data) (e.g., Fig. 4), and other bioinformatics approaches of the same caliber have

generated useful data, these approaches suffer from deficiencies relative to the technologies provided herein. For instance, although automated NGS studies are efficient, several steps are required to prepare sequencing-ready DNA, including complex sample preparation and costly sequencing technology. Bioinformatics analyses are difficult for highly repetitive elements and require added expertise and cost, which makes bioinformatics unlikely to compete with the assay technologies described herein, which can be done in approximately 30 minutes at relatively low cost. Importantly, in silico sequence analysis of centromere sequences requires analysis of raw (e.g., unprocessed) data from whole genome sequencing (WGS) studies because bioinformatics pipelines exclude repetitive sequence data (including centromere repeats) from analysis, especially if reads do not map to a reference human genome sequence. Also, some NGS methods rely on target enrichment in the library preparation process, which uses blocking reagents that eliminate the production of repetitive sequences (such as centromeric α -repeats) in the library. These techniques greatly decrease the value of some NGS technologies for studying human centromeres.

Embodiments of the technologies described herein offer a rapid assay (e.g., PCR assay) to analyze individual centromere sequences simultaneously. Furthermore, some embodiments of the technologies described herein offer a rapid assay (e.g., PCR assay) to analyze individual centromere sequences simultaneously in real-time. Embodiments of the technologies use primers and probes that specifically detect variations in centromere sequences.

Example 3 – Detecting normal ploidy and nondisjunction defects

Embodiments of the technology (e.g., amplification (e.g., PCR) assays) provide a determination of chromosome number by determining the abundance of α -repeats in the centromere of each chromosome. Thus, embodiments of the technology find use in screening DNA samples to determine the number of sex and somatic chromosomes, e.g., for distinguishing genetic gender and identifying chromosome aneuploidy in nondisjunction congenital defects. During the development of embodiments of the technology described herein, experiments were conducted to quantify the number of α -repeats in arrays of chromosomes X and Y to identify gender. Data collected indicated that the assay detected the DYZ3 array from chromosome Y in the male population but not in the female population (Fig. 6). Centromere X DXZ1, while varying among individuals, was consistently found at about 2 times the level in females than in males. In addition, DXZ1 was found at different levels in one individual with trisomy X and in

one phenotypic female with an XY karyotype. In particular, the content of α -repeats was three times higher for the individual with trisomy X and the content of α -repeats was similar to a male for the phenotypic female with the XY karyotype. DYZ3 was also detected in the second individual. Therefore, as shown by experiments described herein, embodiments of the assays described herein determine the sex chromosome ploidy of an individual and thus find use in identifying and characterizing nondisjunction genetic disorders of sex chromosomes, such as triple X syndrome, quadruple X syndrome, Klinefelter (XXY) syndrome, Turner syndrome (monosomy X), super male (XYY) syndrome, and other disorders associated with sex chromosome ploidy.

Additionally, experiments were conducted during the development of embodiments of the technology described herein to assess α -repeat array number for chromosomes 8 and 18 in karyotypically normal individuals and in individuals with trisomy 8 (Warkany syndrome) or trisomy 18 (Edwards syndrome). The data collected during these experiments indicated an approximately 1.5 $\times$ increase in the α -repeat content of D8Z2 in individuals with trisomy 8 relative to the karyotypically normal population (Fig. 6). A similar increase was detected in D18Z1 and D18Z2 of chromosome 18 in individuals with trisomy 18.

Thus, these data indicated that embodiments of the assay technologies provided herein reliably distinguish individuals having trisomy 8 or trisomy 18 from those without the ploidy defect. Therefore, in contrast to current karyotype analysis in live cells, embodiments of the assays described herein find use in detecting the abundance of centromere arrays and thus provide a measure of the number of chromosomes in any biological sample, regardless of origin or preservation of the specimen. Thus, embodiments of this new technology find use for antenatal diagnostics. Further, as described herein, embodiments of the assays find use in studying centromere instability and the effects thereof in nondisjunction chromosomal defects.

Example 4 – Quantifying centromeric transcript abundance

Human centromere sequences are thought to be tightly packaged into heterochromatin structures and thus transcriptionally inert. Recent evidence suggests that these areas of the genome are not only transcriptionally active, but their transcripts regulate chromosome stability and cell division (Wong et al., 2007). Studies have shown the expression of centromere arrays in chromosomes 4, 9, and 13/21, yet a more complete map of centromere α -repeat array transcripts has not yet been generated. During the development of embodiments of the technology provided herein, amplification (e.g., PCR)

assays were used to measure the levels of transcription of individual human centromere α -repeat arrays. Data collected from these experiments indicated that transcripts originating from all centromere arrays are found in benign epithelial prostate and prostate cancer cell lines (Fig. 7). These transcripts are genuine RNA sequences; these transcripts are detected in samples carefully treated with DNase that do not produce amplification products without a reverse transcription step in the PCR reactions. Therefore, embodiments of the technology provided herein (e.g., PCR assays) find use in studying transcription levels of distinct human α -repeat arrays arising from centromeres.

Example 5 – Studying centromere epigenetics

Recent studies indicated that centromere function is modulated not only epigenetically, but also by a genetic component that drives centromere formation and function (Hayden, 2012; Henikoff et al., 2015). In particular, centromere proteins CENP-A and CENP-B bind to centromere DNA and orient centromeric chromatin in a specific and/or spatial order. The affinity of centromere proteins for binding centromeric arrays has been studied in single stretched chromatin fibers or by ChIP-Seq analysis (Maloney et al., 2012; Aldrup-MacDonald and Sullivan, 2014; Henikoff et al., 2015; Ross et al., 2016). These studies measured the binding of CENP-A and identified CENP-B DNA boxes in centromere arrays from chromosomes 1, 4, 5, 7, 11, 17, 19, X, and Y. However, the question remains whether CENP-A and CENP-B are capable of binding to all centromere arrays or only to specific, selected α -repeats.

Accordingly, embodiments of the technology provided herein finds use in addressing this question. During the development of embodiments of the technology provided herein, experiments were conducted in which chromatin immunoprecipitation (ChIP) assays were performed with the human LnCaP prostate cancer cell line using antibodies that recognize CENP-A and CENP-B. Data were collected to measure the content of α -repeats in each array co-immunoprecipitated with these proteins. The ChIP assay data indicated that CENP-A binds to at least one α -repeat array in each human centromere (Fig. 8); these arrays can therefore be classified as competent centromeric arrays (see, e.g., Henikoff et al., 2015).

Similarly, data collected from experiments performed during the development of the technology provided herein indicated that the list of inactive centromere arrays that do not recruit CENP-A was previously incomplete (Henikoff et al., 2015). That is, the

data indicated that more arrays are inactive and do not recruit CENP-A than was previously known.

Interestingly, data collected during experiments conducted during the development of embodiments of the technology described herein indicated that all competent centromere arrays that bind CENP-A contain CENP-B boxes and bind CENP-B except for chromosome Y (Fig. 8; also see below). In centromeres that comprise more than one array, CENP-A bound predominantly to the largest array. In particular, e.g., CENP-A preferentially bound to D1Z7 relative to D1Z5; CENP-A preferentially bound to D7Z1 relative to D7Z2; CENP-A preferentially bound to D14Z1 relative to D14Z2; CENP-A preferentially bound to D17Z1 relative to D17Z1b; CENP-A preferentially bound to D18Z1 relative to D18Z2; and CENP-A preferentially bound to D22Z1 relative to D22Z4 or D22Z5.

Data also indicated that CENP-A binds to α -repeat arrays in the core centromeric areas but not in the pericentric areas. Experiments were conducted to estimate the fraction of α -repeats in each array loaded by CENP-A by determining the fraction of α -repeats immunoprecipitated with CENP-A antibodies. The data collected indicated that CENP-A loads between approximately 10 to 40% of a given array in each centromere, at least in the cell line LNCaP. These studies confirm and extend previous investigations focusing on a few centromere arrays (chromosomes 1, 17, X, and Y) (Maloney et al., 2012; Aldrup-MacDonald and Sullivan, 2014; Ross et al., 2016). CENP-A also binds to approximately 20% of the ubiquitous p82H repeat and occupied up to 70% of the array DYZ3 in chromosome Y (Fig. 8).

Experiments conducted during the development of embodiments of the technology described herein collected data from ChIP assays of CENP-B, which revealed that CENP-B binds to all centromere arrays except for D7Z2, D19Z4, D19Z5, D22Z4, D22Z5, and DYZ3 (Fig. 8). Previously, data indicated that CENP-B does not bind the array DYZ3 in chromosome Y (Fachinetti et al., 2015), which is confirmed by the data collected in experiments conducted during the development of the technology described herein. In contrast to CENP-A, data collected during these experiments indicated that CENP-B binds to arrays in either the centromere core or in the pericentric domain.

CENP-B binds to a specific 17-nt CENP-B DNA box; data collected during the experiments conducted during the development of embodiments of the technology described herein indicated the existence and characterized the distribution of CENP-B boxes along the linear sequences of these arrays (Fig. 9). Analysis of the data identified 4 types of CENP-B boxes that CENP-B binds and did not identify intact CENP-B boxes

in centromere arrays that were not co-immunoprecipitated with CENP-B protein (see, e.g., Table 4). In Table 4, the CENP-B box sequences are:

	Name	Sequence (5' to 3')	SEQ ID NO:
5	CENP-B Box 1	<u>CTTCGTTGGAA</u> <u>ACGGGA</u>	108
	CENP-B Box 2	<u>CTTCGTTGGAA</u> <u>ACGGGT</u>	109
	CENP-B Box 3	<u>TTTCGTTGGAA</u> <u>ACGGGA</u>	110
	CENP-B Box 4	<u>TTTCGTTGGAA</u> <u>ACGGGA</u>	111

- 10 The underlined portions of the CENP-B Box nucleotide sequences promote (e.g., may be necessary for or are necessary for) CENP-B binding. Thus, box sequences having mutations at one or more of the underlined bases will not recruit CENP-B or will recruit CENP-B less efficiently than a non-mutated sequence. In Table 4, a check mark indicates that the indicated CENP-B Box sequence is present in the denoted
- 15 chromosomal α -repeat sequence.

Table 4 – CENP-B Boxes found in human centromere α -repeats

α -repeat	CENP-Box 1	CENP-Box 2	CENP-Box 3	CENP-Box 4
D1Z5-1				
D1Z5-2				
D1Z7/D5Z2/D19Z3	✓			
D2Z1				✓
D3Z1	✓			✓
D4Z1				✓
D5Z1/D19Z2		✓		
D6Z1	✓			
D7Z1	✓			
D7Z2	✓			
D8Z2				✓
D9Z4				✓
D10Z1	✓			
D11Z1				✓
D12Z3	✓			
D13Z1/D21Z1			✓	✓
D13Z2				
D13Z3				
D13Z6	✓			
D13Z7				
D13Z8				
D13Z9				
D14Z1/D22Z1				✓
D14Z2				✓

D14Z3/D22Z6				
D15Z3			✓	✓
D16Z2	✓			
D17Z1				✓
D171b				✓
D18Z1			✓	✓
D18Z2				✓
D19Z4				
D19Z5				
D20Z2				✓
D21Z1/D13Z1			✓	✓
D22Z4-1				
D22Z4-2				
D22Z5				
DXZ1				✓
DYZ3				
P82H	✓			

Furthermore, the data collected indicated that CENP-B boxes exist approximately every 340 nt in the α -repeat arrays to which CENP-B binds, confirming and expanding recent observations (Henikoff et al., 2015). Data analysis of the fractional occupancy of CENP-B in each array provided an estimate of CENP-B occupancy to be approximately 20 to 100% of the full size of the arrays (Fig. 8). Despite the observation that inactive α -repeat arrays do not recruit CENP-A, some α -repeat arrays recruit CENP-B, thus suggesting that some α -repeat arrays and CENP-B are competent to maintain cohesion of sister chromatids. Therefore, data collected from experiments conducted during the development of technology provided herein indicate the presence of a relationship between centromere genomics to epigenetics and further indicate that this relationship exists in most or all centromeres.

Example 6 – Distinguishing centromeres 13 and 21

Aneuploidy of chromosome 21 causes Down syndrome; accordingly, enumeration of chromosome 21, e.g., by detecting and quantifying chromosome centromeres, provides a diagnostic tool for Down syndrome and a research tool for studying the pathogenesis of this syndrome. However, analysis of chromosome 21 based on detection of the centromere is complicated by the high similarity (e.g., nearing 100% identity) of the centromere α -repeat arrays of chromosome 13 D13Z1 and chromosome 21 D21Z1. The centromere α -repeat arrays of chromosome 13 D13Z1 and chromosome 21 D21Z1 differ by two single nucleotides (Pellestor et al., 1994; Nilsson et al., 1997).

Accordingly, embodiments of the technology provide amplification (e.g., PCR) assays and modified primers that recognize the nucleotide variations in D13Z1 or D21Z1 (Fig. 10A). To detect these mutations (and chromosomes comprising these mutations), experiments were conducted to design primers containing a locked nucleic acid (LNA) modification at the bases targeting the nucleotide change and amplification conditions were developed to detect the D13Z1 and/or D21Z1 nucleotide substitution (Skronsky et al., 2013; Tables 1 and 2, Fig. 11).

In some embodiments, the assay for detecting D13Z1 comprises a LNA primer that specifically binds to a D13Z1 base substitution and another primer that binds to both D13Z1 and D21Z1 arrays (Fig. 11A). In some embodiments, the assay for detecting D21Z1 comprises primers with LNA modifications that detect a nucleotide substitution of D21Z1 and a primer clamp that recognizes a base substitution of D13Z1 and prevents amplification of D13Z1 (Tables 1 and 2). This combination of primers, primer modifications, and a primer clamp substantially reduced the amplification of D13Z1 while allowing the detectable amplification of D21Z1 (Fig. 11B). Experiments were conducted during the development of embodiments of the technology in which amplification reactions designed to detect either D13Z1 or D21Z1 were evaluated in the context of all human chromosomes; The data collected from these experiments indicated that the assay is specific for D13Z1 and/or D21Z1 (Fig. 10B).

During the development of embodiments of the technology, *in silico* analysis was performed on samples derived from the Human Genome Diversity Project (HGDP) to examine the existence of these centromere 13 and 21 nucleotide substitutions in human populations, extinct hominids, and apes. The data produced from these experiments indicated that the D13Z1 and D21Z1 substitutions were present in all human populations studied (Fig. 10C). Accordingly, these data indicate that the embodiments of the assay technology described herein (e.g., PCR assays) differentiate centromeres 13 and 21 in diverse human populations. Furthermore, these data indicated that D13Z1 and D21Z1 substitutions do not exist in modern apes, except for a very few copies of a D21Z1 substitution detected in chimpanzees and gorillas, indicating that these nucleotide variations accumulated after the split of the Homo-Pan ancestor (Fig. 10C). Additional analysis of these nucleotide substitutions in extinct hominids indicated that D13Z1 and D21Z1 substitutions accumulated in Denisovan and Neanderthal individuals. Interestingly, the D13Z1 substitution was detected at a higher frequency in the Denisovan population (Fig. 10C).

Example 7 – Centromeric instability in trisomy 13 and trisomy 21

Some embodiments of the technology described herein provide a rapid and comprehensive assay (e.g., amplification (e.g., PCR) assay) to examine human centromeres (e.g., one or more human centromeres; e.g., each distinct human centromere). In some embodiments, the technology finds use to assess the role of these sequences in developmental disorders. Accordingly, experiments were conducted during the development of embodiments of the technology to detect nucleotide substitutions present in D13Z1 and D21Z1 in individuals with unbalanced numbers of chromosome 13 or chromosome 21. Data were collected to quantify the number of D13Z1 and D21Z1 specific substitutions in genetically normal individuals and in individuals with trisomy 13 (Patau syndrome) or trisomy 21 (Down syndrome). The data indicated that the D13Z1 array was detected at a level that is approximately $1.5 \times$ higher in individuals with trisomy 13 than in those without the defect (Fig. 5A). The data also indicated that some individuals with trisomy 21, but not all, have lesser amounts of D13Z1 than the amount of D13Z1 detected in the normal population.

Analysis of the D21Z1 sequences did not demonstrate an increased number of centromere 21 substitutions in individuals with trisomy 21 as compared to individuals without the defect. In fact, the number of D21Z1 specific sequences is dramatically lower in individuals with trisomy 21, indicating that a loss of D21Z1 α -repeat arrays or a partial loss of centromere 21 DNA exists in individuals with trisomy 21 (Fig. 12B). Experiments were conducted with the trisomy 21 DNA samples assayed as described above to confirm that the samples had a trisomy 21 karyotype. In particular, a PCR genetic test was used for the chromosome 21 short tandem repeat D21S167 and the S100B gene (Yang et al., 2005), which verified that the trisomy 21 DNA samples had extra copies on chromosome 21 in contrast to samples from individuals with a normal karyotype (not shown). Taken together, these data indicate that instability of centromere 21 exists in individuals with trisomy 21, which is a previously unsuspected contribution to this genetic defect.

Example 8 – Pericentric instability in trisomy 13 and trisomy 21

As described herein, individuals with trisomy 21 have genome instability at the centromere of chromosome 21. Accordingly, experiments were conducted during the development of embodiments of the technology described herein to evaluate instability at the pericentromere of chromosome 21 in individuals having trisomy 21. In particular, experiments were conducted using the pericentric markers K111 and K222 described

herein. While K111 and K222 exist in several pericentromeres as single copies, multiple copies of K111 have accumulated in the pericentromere of chromosome 21 and K222 has accumulated in pericentromeres of chromosomes 13 and 14 (Contreras-Galindo et al., 2011, 2013; Zahn et al., 2015). During the development of embodiments of the technology provided herein, experiments were conducted in which pericentric K111 and K222 were quantified as a surrogate for pericentromere size. The data collected from these experiments indicated that, in contrast to the DNA of healthy individuals, individuals with trisomy 13 and 21 have shorter pericentromeres (Fig. 12C). These data indicate that pericentromere instability or deletion of pericentric areas is present in individuals with trisomy 13 and trisomy 21.

Furthermore, during the development of embodiments of the technology described herein, experiments were conducted to develop an assay using LNA-modified primers to detect K111 specifically and, furthermore, to quantify K111. The specificity of the assay detected K111 but did not detect K222. Indeed, the data indicated that K111 copy numbers, which are abundant in the pericentromere of chromosome 21, are significantly reduced in individuals with trisomy 21 (Fig. 12D), indicating with additional data that pericentric instability is present in individuals with trisomy 21.

Additionally, experiments were conducted during the development of embodiments of the technology in which sequencing was used to assess the diversity of K111 proviruses in trisomy 21 individuals. Sequence analysis indicated that, in contrast to genetically normal individuals, K111 sequences specific for chromosome 21 are present at reduced numbers in trisomy 21 individuals (Fig. 13). The analysis also shows that K111 sequences from other chromosomes are detected in trisomy 21 individuals, similar to detection of these sequences in healthy individuals. Interestingly, the nucleotide sequences of chromosome 21 DNA from trisomy 21 individuals comprised novel K111 sequences that do not match known K111 sequences. Phylogenetic analysis indicated that these sequences are the result of homologous recombination at the pericentromere (Fig. 13). Recombination analysis indicated that the new K111 phylogenetic branch detected in the tree of trisomy subject C (black star in Fig. 13) is the result of recombinational deletion between distinct K111 sequences from chromosome 21 (not shown). Recombination analysis did not identify the origin of the novel K111 sequences in trisomy 21 subjects A and B (Fig. 13).

Accordingly, the data indicate that instability at pericentromeric and/or centromeric loci is associated with and/or causes missegregation of chromosomes during cell division and thus that instability at pericentromeric and/or centromeric loci is a

causative factor for trisomy. Importantly, the methodologies of extant technologies would not have been able to produce these data.

Example 9 – Centromere instability in scleroderma

5 Scleroderma (“SSc”) is a disease of unknown etiology, but exposure to chemicals appears to play some fundamental role in the pathogenesis of this disease. SSc is an autoimmune disease characterized by a complex interplay of autoimmunity, vasculopathy, and fibrosis of the skin and may involve internal organs. There are two main types of the disease: limited cutaneous SSc (“lcSSc”) and diffuse cutaneous SSc
10 (“dcSSc”) (49, 50). LcSSc usually has skin involvement limited to hands, forearms, feet, and legs. DcSSc is associated with disease in the upper arms, thighs and torso, and is a rapidly progressive condition. Both subsets are associated with internal organ involvement, usually of the kidneys, heart, intestines, and lungs. SSc can be quite disabling to patients. About 150,000 individuals suffer from SSc in the United States
15 and it affects mostly women between the ages of 30 to 50 (51).

Although genetic factors and exposure to environmental agents seem to play roles in SSc, the cause of this disease remains unknown. Chemicals used for chemotherapy (e.g., bleomycin) and chemicals found in metal degreasing and dry cleaning products (e.g., benzene, perchloroethylene (PCE), and trichloroethylene (TCE))
20 have been shown to induce SSc-like fibrosis. (53-55). Limited studies indicate that personnel exposed to these solvents during work or accessing contaminated water supplies have an increased incidence to develop SSc or other autoimmune disorders (50-55). There is also concern that veterans exposed to herbicides such as Agent Orange might have an increased risk of developing SSc (56). This is a major medical concern, for
25 example, with respect to millions of active-duty military workers and reserve veterans have been assigned at some time to places where exposure to these environmental contaminants took place. Nonetheless, the pathogenesis of SSc remains unclear, including the role of chemical agents in inducing SSc.

SSc diagnosis relies on physical exams and clinical evaluations accompanied by
30 screening for antinuclear antibodies (ANAs) (57-59). Among the ANAs tested, antitopoisomerase-I antibodies (ATAs) (usually called anti-Scl-70) are seen predominately in dcSSc patients (prevalence 20-40%) and are associated with poor prognosis, pulmonary fibrosis, and disease progression (52-54). Anticentromere antibodies (ACAs), another analyte tested by the ANA test, are found in up to 60-70% of
35 lcSSc (57-62). ACAs recognize centromere proteins CENP-A to CENP-T, predominately

CENP-B. ACAs are associated with a more favorable prognosis as compared with the positivity of ATAs (57, 60). Anti-RNA polymerase I and III antibodies detected in 10-25% of dcSSc patients are used as a predictive marker of rapid onset of the disease, skin thickening, and renal crisis (57). Other autoantibody tests have prognostic utility, but
5 only for a small percentage of SSc patients.

Despite their clinical value for prognosis, the etiology attributed to the production of ATAs and ACAs in SSc patients is unknown. Further, whether nuclear and/or centromere protein defects and antigen presentation arise in cells implicated in SSc fibrosis have not been explored. The pathophysiology of SSc includes immune
10 activation, vascular injury, and loss of epithelium, and culminates in fibrosis, a thickening of the connective tissue produced by excessive accumulation of extracellular matrix (ECM) proteins, including collagen type-I, type-III, and fibronectin (63, 64). Activation of fibroblasts, the resident cells of the connective tissue, is a key contributor to fibrosis in SSc (63). After tissue injury, active fibroblasts migrate to the ECM. In an
15 environment rich in growth factors secreted by immune and blood cells, such as fibroblast growth factor (FGF) and connective tissue growth factor (CTGF), active fibroblasts differentiate into secretory myofibroblasts in a process called fibroblast to myofibroblast transition (63). SSc fibroblasts also secrete growth factors to recruit more
20 resident fibroblasts to these lesions. Unlike fibroblasts, myofibroblasts express alpha smooth muscle actin (α -SMA) (63). Accumulation of a large number of myofibroblasts is responsible for the excessive synthesis of ECM proteins. Myofibroblasts usually arise from mesenchymal stem cells (MSCs) of fibroblastic lineage, such as resident fibroblasts and pericytes, but could also arise from other cellular sources including smooth muscle
25 cells (63, 64). Recent studies show that in contrast to healthy cells, fibroblasts and/or myofibroblasts isolated from fibrotic tissue from idiopathic pulmonary fibrosis (IPF) or oral submucous fibrosis (OSF) show genome instability (65-67).

During the development of embodiments of the technology described herein, experiments were conducted indicating that the stability of centromere sequences is quite different in SSc with respect to healthy fibroblasts. Data collected during these
30 experiments indicated that major losses of centromere material occurred in several chromosomes from diffuse cutaneous Scleroderma (dsSSc) fibroblasts (5043, 5060, 5065, 5074) relative to healthy fibroblasts (N59, N60, and N64) (Fig. 14). Further, the data did not indicate that centromere changes occurred in cells cultured at different passages or treated with antibiotics or antimycotics. Accordingly, embodiments of the technology
35 relate to a centromere assay (e.g., by amplification, e.g., by PCR) for detecting

destabilized centromeres in samples from SSc patients and/or tissues. In some embodiments, the centromere assay detects the abundance of alpha-repeats of one or more chromosomes. In some embodiments, the centromere assay detects the abundance of an alpha-repeat designated as demonstrating a significant difference in Figure 14
5 between the SSc and normal fibroblasts (e.g., D1Z5, D2Z1, D3Z1, D4Z1, D5Z1, D6Z1, D7Z2, D8Z2, D9Z4, D10Z1, D11Z1, D13Z1, D15Z3, D16Z2, D17Z1, D17Z1b, D19Z4, D19Z5, D20Z2, D21Z1, p82H, Top3A). In some embodiments, an alpha-repeat designated as demonstrating a significant difference in Figure 14 between the SSc and normal fibroblasts provides a measurement of genome instability in fibrosis.

10

Example 10 – Centromere instability induced by contaminants of concern

During the development of embodiments of the technology described herein, experiments were conducted to investigate causes of centromere deletion in dcSSc. Without being bound by theory, it was contemplated that centromere deletions arise
15 from recombination of centromeres created during repair of double strand breaks (DSBs). Thus, experiments were conducted to test this hypothesis by treating fibroblasts with bleomycin, an agent that induces double strand breaks (“DSBs”) (68-70), produces a SSc-like syndrome in mice (71, 72), and causes Raynaud’s syndrome and pulmonary fibrosis (PF) in humans (73), similar to that seen in SSc. Treatment of CHON-002
20 fibroblasts with different concentrations of bleomycin produced DSBs, as measured by increased production of H2AX (Fig.15A). This assay measures the levels of phosphorylation of histone H2A variant X (H2AX), a protein that is specifically phosphorylated after (e.g., nearly immediately after) the formation of DSBs (74, 75).

Further, bleomycin treatment produced centromere instability as measured by
25 PCR assays, indicating that DSBs in centromere DNA produce centromere mutations that cause SSc-like syndrome and fibrosis. Accordingly, the data indicate that centromere instability causes SSc-like disease and further experiments were conducted to test whether environmental contaminants linked to SSc in military personnel produce centromere instability. Strikingly, data collected during these experiments indicated
30 that treatment of cells with the Agent Orange component 2,4,5-T and the herbicide Paraquat produced centromere instability at 24 hours (Fig. 15B), suggesting that these herbicides produce genome instability that causes fibrosis. In contrast, the data indicated that amphotericin B produced no effect in cells.

Consequently, in some embodiments, the technology provides an assay of
35 centromere instability in samples from many types of patients, e.g., those having

fibrosis, trisomy and disease. Furthermore, in some embodiments the technology relates to screening drugs to identify drugs that produce genome instability and would therefore produce undesirable disease states.

5

REFERENCES

1. Aldrup-Macdonald, ME., Sullivan, BA. (2014). The past, present, and future of human centromere genomics. *Genes (Basel)*. 5, 33-50.
2. Aleixandre, C., Miller, DA., Mitchell, AR., Warburton, DA., Gersen, SL., Disteche, C., Miller, OJ. (1987). p82H identifies sequences at every human
10 centromere. *Hum Genet.* 77, 46-50.
3. Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V., Yurov, Y. (2001). Alpha-satellite DNA of primates: old and new families. *Chromosoma*. 110, 253-266.
4. Amor, DJ., Choo, KH. (2002). Neocentromeres: role in human disease, evolution, and centromere study. *Am. J. Hum. Genet.* 71, 695-714. Review.
- 15 5. Ballantyne, KN., van Oorschot, RA., Mitchell, RJ. (2008). Locked nucleic acids in PCR primers increase sensitivity and performance. *Genomics*. 91, 301-305.
6. Bersani, F., Lee, E., Kharchenko, P.V., Xu, AW., Liu, M., Xega, K., MacKenzie, OC., Brannigan, B.W., Wittner, B.S., Jung, H., et al. (2015). Pericentric satellite repeat expansions through RNA-derived DNA intermediates in cancer. *Proc.*
20 *Natl. Acad. Sci. USA*. 112, 15148-53.
7. Carroll, CW., Straight, AF. (2006). Centromere formation: from epigenetics to self-assembly. *Trends Cell. Biol.* 16, 70-8. Review.
8. Choo, K.H., Vissel, B., Earle, E. (1989). Evolution of alpha-satellite DNA on human acrocentric chromosomes. *Genomics*. 5, 332-44.
- 25 9. Contreras-Galindo, R., Kaplan, MH., Contreras-Galindo, AC., Gonzalez-Hernandez, MJ., Ferlenghi, I., Giusti, F., Lorenzo, E., Gitlin, SD., Dosik, MH., Yamamura, Y., Markovitz, DM. (2011). Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J. Virol.* 86, 262-76.
- 30 10. Contreras-Galindo, R., Kaplan, MH., He, S., Contreras-Galindo, AC., Gonzalez-Hernandez, MJ., Kappes, F., Dube, D., Chan, SM., Robinson, D., Meng, F., et al. (2013). HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses. *Genome Res.* 23, 1505-1513.

11. Fachinetti, D., Han, JS., McMahon, MA., Ly, P., Abdullah, A., Wong, AJ.,
Cleveland, DW. (2015). DNA Sequence-Specific Binding of CENP-B Enhances the
Fidelity of Human Centromere Function. *Dev. Cell.* 33, 314-27.
12. Falk, SJ., Guo, LY., Sekulic, N., Smoak, EM., Mani, T., Logsdon, GA., Gupta, K.,
5 Jansen, LE., Van Duyne, GD., Vinogradov SA, et al. (2015). CENP-C reshapes
and stabilizes CENP-A nucleosomes at the centromere. *Science.* 348, 699-703.
13. Hanai, R., Caron, PR., Wang, JC. (1996). Human TOP3: a single-copy gene
encoding DNA topoisomerase III. *Proc. Natl. Acad. Sci. USA.* 93, 3653-3657.
14. Hayden, KE. (2012). Human centromere genomics: now it's personal.
10 *Chromosome Res.* 20, 621-33.
15. Hayden, KE., Willard, HF. (2012). Composition and organization of active
centromere sequences in complex genomes. *BMC Genomics.* 13, 324.
16. Henikoff, JG., Thakur, J., Kasinathan, S., Henikoff, S. (2015). A unique
chromatin complex occupies young α -satellite arrays of human centromeres. *Sci.*
15 *Adv.* 1, pii: e1400234.
17. Hood, L., and Rowen, L. (2013). The Human Genome Project: big science
transforms biology and medicine. *Genome Med.* 5, 79.
18. Horvath, JE., Gulden, CL., Bailey, JA., Yohn, C., McPherson, JD., Prescott, A.,
Roe, BA., de Jong, PJ., Ventura, M., Misceo, D, et al. (2003). Using a pericentric
20 interspersed repeat to recapitulate the phylogeny and expansion of human
centromeric segmental duplications. *Mol. Biol. Evol.* 20, 1463-79.
19. Huelsenbeck, JP., Ronquist, F., Nielsen, R., Bollback, JP. (2001). Bayesian
inference of phylogeny and its impact on evolutionary biology. *Science* 294, 2310-
2314.
- 25 20. Jørgensen, AL. (1997). Alphoid repetitive DNA in human chromosomes. *Dan Med*
Bull. 44, 522-34.
21. Kirsch, S., Weiß, B., Miner, TL., Waterston, RH., Clark, RA., Eichler, EE.,
Münch, C., Schempp, W., Rappold, G. (2005). Interchromosomal segmental
duplications of the pericentric region on the human Y chromosome. *Genome Res.*
30 15, 195-204.
22. Koumbaris, G., Kypri, E., Tsangaras, K., Achilleos, A., Mina, P., Neofytou, M.,
Velissariou, V., Christopoulou, G., Kallikas, I., González-Liñán, A, et al. (2016).
Cell-Free DNA Analysis of Targeted Genomic Regions in Maternal Plasma for
Non-Invasive Prenatal Testing of Trisomy 21, Trisomy 18, Trisomy 13, and Fetal
35 Sex. *Clin. Chem.* 62, :848-855.

23. Liehr, T. (2013). Benign & Pathological Chromosomal Imbalances. Academic Press, Elsevier.
24. Malik, HS., Henikoff, S. (2009). Major evolutionary transitions in centromere complexity. *Cell*. 138, 1067-1082.
- 5 25. Maloney, KA., Sullivan, LL., Matheny, JE., Strome, ED., Merrett, SL., Ferris, A., Sullivan, BA. (2012). Functional epialleles at an endogenous human centromere. *Proc. Natl. Acad. Sci. USA*. 109, 13704-13709.
- 26 Miga, KH. (2015). Completing the human genome: the progress and challenge of satellite DNA assembly. *Chromosome Res.* 23, 421-426.
- 10 27. Miga, KH. (2015). Completing the human genome: the progress and challenge of satellite DNA assembly. *Chromosome Res.* 23, 421-426.
28. Miga, KH., Newton, Y., Jain, M., Altemose, N., Willard, HF., Kent, WJ. (2014). Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res.* 24, 697-707.
- 15 29. Mitchell, AR., Gosden, JR., Miller, DA. (1985). A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. *Chromosoma*. 92, 369-377.
30. Nilsson, M., Krejci, K., Koch, J., Kwiatkowski, M., Gustavsson, P., Landegren, U. (1997). Padlock probes reveal single-nucleotide differences, parent of origin and in situ distribution of centromeric sequences in human chromosomes 13 and 21. *Nat. Genet.* 16, 252-255.
- 20 31. Ohzeki, J., Nakano, M., Okada, T., Masumoto, H. (2002). CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J. Cell. Biol.* 159, 765-775.
- 25 32. Pellestor, F., Girardet, A., Andréo, B., Charlieu, JP. (1994). A polymorphic alpha satellite sequence specific for human chromosome 13 detected by oligonucleotide primed in situ labelling (PRINS). *Hum. Genet.* 94, 346-348.
33. Roizès, G. (2006). Human centromeric alphoid domains are periodically homogenized so that they vary substantially between homologues. Mechanism and implications for centromere functioning. *Nucleic Acids Res.* 34, 1912-24.
- 30 34. Ronquist, F., Huelsenbeck, JP. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572-1574.
- 35 35. Rosandić, M., Paar, V., Basar, I., Glunčić, M., Pavin, N., Pilas, I. (2006). CENP-B box and pJalpha sequence distribution in human alpha satellite higher-order repeats (HOR). *Chromosome Res.* 14, 735-53.

36. Ross, JE., Woodlief, KS., Sullivan, BA. (2016). Inheritance of the CENP-A chromatin domain is spatially and temporally constrained at human centromeres. *Epigenetics Chromatin*. 9, 20.
37. Scott, KC., Sullivan, BA. (2014). Neocentromeres: a place for everything and everything in its place. *Trends Genet.* 30, 66-74. Review.
38. Skronski, M., Chorostowska-Wynimko, J., Szczepulska, E., Szpechcinski, A., Rudzinski, P., Orlowski, T., Langfort, R. (2013). Reliable detection of rare mutations in EGFR gene codon L858 by PNA-LNA PCR clamp in non-small cell lung cancer. *Adv. Exp. Med. Biol.* 756, 321-331.
39. Stimpson, KM., Sullivan, BA. (2010). Epigenomics of centromere assembly and function. *Curr. Opin. Cell. Biol.* 22, 772-80. Review.
40. Sullivan, LL., Boivin, CD., Mravinac, B., Song, IY., Sullivan BA. (2011). Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosome Res.* 19, 457-70.
41. Talbert PB, Henikoff S. (2010) Centromeres convert but don't cross. *PLoS Biol.* 8(3):e1000326.
42. Verdaasdonk, J.S., Bloom, K. (2011). Centromeres: unique chromatin structures that drive chromosome segregation. *Nat. Rev. Mol. Cell. Biol.* 2011 12,320-32.
43. Vissel, B., Choo, KH. (1992). Evolutionary relationships of multiple alpha satellite subfamilies in the centromeres of human chromosomes 13, 14, and 21. *J. Mol. Evol.* 35, 137-46.
44. Vos, LJ., Famulski, JK., Chan, GK. (2006). How to build a centromere: from centromeric and pericentric chromatin to kinetochore assembly. *Biochem. Cell. Biol.* 84, 619-39. Review.
45. Wong, LH., Brettingham-Moore, KH., Chan, L., Quach, JM., Anderson, MA., Northrop, EL., Hannan, R., Saffery, R., Shaw, ML., Williams, E., Choo, KH. (2007). Centromere RNA is a key component for the assembly of nucleoproteins at the nucleolus and centromere. *Genome Res.* 17, 1146-1160.
46. Yang, YH., Nam, MS., Yang ES. (2005). Rapid Prenatal Diagnosis of Trisomy 21 by Real-time Quantitative Polymerase Chain Reaction with Amplification of Small Tandem Repeats and S100B in Chromosome 21. *Yonsei Medical Journal.* 46, 193-197.
47. Zahn, J., Kaplan, MH., Fischer, S., Dai, M., Meng, F., Saha, AK., Cervantes, P., Chan, SM., Dube, D. et al. (2015). Expansion of a novel endogenous retrovirus throughout the pericentromeres of modern humans. *Genome Biol.* 16, 74.

48. Zeitlin, SG. (2010). Centromeres: the wild west of the post-genomic age. *Epigenetics*. 5, 34-40.
49. Denton CP, Khanna D. Systemic sclerosis. *Lancet*. 2017 Apr 13. pii:S0140-6736(17)30933-9.
- 5 50. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med*. 2009 May 7;360(19):1989-2003.
51. Mayes MD, Lacey JV Jr, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, Schottenfeld D. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum*. 2003 Aug;48(8):2246-55..
- 10 52. Redd D, Frech TM, Murtaugh MA, Rhiannon J, Zeng QT. Informatics can identify systemic sclerosis (SSc) patients at risk for scleroderma renal crisis. *Computers in biology and medicine*. 2014;53:203-205.
- 53 Long-term health effects of early life exposure to tetrachloroethylene (PCE)-contaminated drinking water: a retrospective cohort study *Ann Aschengrau, Michael R Winter, Veronica M Vieira, Thomas F Webster, Patricia A Janulewicz, Lisa G Gallagher, Janice Weinberg, David M Ozonoff Environ Health*. 2015; 14: 36.
- 15 54. Rubio-Rivas M, Moreno R, Corbella X. Occupational and environmental scleroderma. Systematic review and meta-analysis. *Clin Rheumatol*. 2017 Mar;36(3):569-582.
- 20 55. Mayes MD. Epidemiologic studies of environmental agents and systemic autoimmune diseases. *Environmental Health Perspectives*. 1999;107(Suppl 5):743-748.
- 25 56. Noakes R. The Aryl Hydrocarbon Receptor: A Review of Its Role in the Physiology and Pathology of the Integument and Its Relationship to the Tryptophan Metabolism. *International Journal of Tryptophan Research : IJTR*. 2015;8:7-18.
57. Affandi AJ, Radstake TR, Marut W. Update on biomarkers in systemic sclerosis: tools for diagnosis and treatment. *Semin Immunopathol*. 2015 Sep;37(5):475-87.
- 30 58. Kuwana M. Circulating Anti-Nuclear Antibodies in Systemic Sclerosis: Utility in Diagnosis and Disease Subsetting. *J Nippon Med Sch*. 2017;84(2):56-63.
59. Sticherling M. Systemic sclerosis-dermatological aspects. Part 1: Pathogenesis, epidemiology, clinical findings. *J Dtsch Dermatol Ges*. 2012 Oct;10(10):705-18.

60. Perosa F, Prete M, Di Lernia G, Ostuni C, Favoino E, Valentini G. Anti-centromere protein A antibodies in systemic sclerosis: Significance and origin. *Autoimmun Rev.* 2016 Jan;15(1):102-9.
61. Gunn J, Pauling JD, McHugh NJ. Impact of anti-centromere antibodies on pulmonary function test results in patients with systemic sclerosis without established or suspected pulmonary disease. *Clin Rheumatol.* 2014 Jun;33(6):869-71.
62. Song G, Hu C, Zhu H, Wang L, Zhang F, Li Y, Wu L. New centromere autoantigens identified in systemic sclerosis using centromere protein microarrays. *J Rheumatol.* 2013 Apr;40(4):461-8.
63. Gilbane AJ, Denton CP, Holmes AM. Scleroderma pathogenesis: a pivotal role for fibroblasts as effector cells. *Arthritis Res Ther.* 2013;15(3):215.
64. Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. *Annu Rev Pathol.* 2011;6:509-37.
65. Selman M, Pardo A. Revealing the pathogenic and aging-related mechanisms of the enigmatic idiopathic pulmonary fibrosis. an integral model. *Am J Respir Crit Care Med.* 2014 May 15;189(10):1161-72.
66. Thannickal VJ. Mechanistic links between aging and lung fibrosis. *Biogerontology.* 2013 Dec;14(6):609-15.
67. Teh MT, Tilakaratne WM, Chaplin T, Young BD, Ariyawardana A, Pitiyage G, Lalli A, Stewart JE, Hagi-Pavli E, Cruchley A, Waseem A, Fortune F. Fingerprinting genomic instability in oral submucous fibrosis. *J Oral Pathol Med.* 2008 Aug;37(7):430-6.
68. Benítez-Bribiesca L, Sánchez-Suárez P. Oxidative damage, bleomycin, and gamma radiation induce different types of DNA strand breaks in normal lymphocytes and thymocytes. A comet assay study. *Ann N Y Acad Sci.* 1999;887:133-49.
69. Chen J, Ghorai MK, Kenney G, Stubbe J. Mechanistic studies on bleomycin-mediated DNA damage: multiple binding modes can result in double-stranded DNA cleavage. *Jingyang Chen, Manas K. Ghorai, Grace Kenney, JoAnne Stubbe Nucleic Acids Res.* 2008 Jun; 36(11): 3781–3790.
70. Jiang M, Yu Y, Luo J, Gao Q, Zhang L, Wang Q, Zhao J. Bone Marrow-Derived Mesenchymal Stem Cells Expressing Thioredoxin 1 Attenuate Bleomycin-Induced Skin Fibrosis and Oxidative Stress in Scleroderma. *J Invest Dermatol.* 2017 Jun;137(6):1223-1233.

71. Liang M, Lv J, Zou L, Yang W, Xiong Y, Chen X, Guan M, He R, Zou H. A modified murine model of systemic sclerosis: bleomycin given by pump infusion induced skin and pulmonary inflammation and fibrosis. *Lab Invest.* 2015 Mar;95(3):342-50.
- 5 72. Azhdari M, Baghaban-Eslaminejad M, Baharvand H, Aghdami N. Therapeutic potential of human-induced pluripotent stem cell-derived endothelial cells in a bleomycin-induced scleroderma mouse model. *Stem Cell Res.* 2013 May;10(3):288-300.
73. Cooper JA Jr, Matthay RA. Drug-induced pulmonary disease. *Dis Mon.* 1987 Feb;33(2):61-120.
- 10 74. Scarpato R, Castagna S, Aliotta R, Azzarà A, Ghetti F, Filomeni E, Giovannini C, Pirillo C, Testi S, Lombardi S, Tomei A. Kinetics of nuclear phosphorylation (γ -H2AX) in human lymphocytes treated in vitro with UVB, bleomycin and mitomycin C. *Mutagenesis.* 2013 Jul;28(4):465-73.
- 15 75. Eberlein U, Peper M, Fernández M, Lassmann M, Scherthan H. Calibration of the γ -H2AX DNA double strand break focus assay for internal radiation exposure of blood lymphocytes. *PLoS One.* 2015 Apr 8;10(4):e0123174.

20 All publications and patents mentioned in the above specification are herein incorporated by reference in their entirety for all purposes. Various modifications and variations of the described compositions, methods, and uses of the technology will be apparent to those skilled in the art without departing from the scope and spirit of the technology as described. Although the technology has been described in connection with specific exemplary embodiments, it should be understood that the invention as claimed

25 should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the following claims.

CLAIMS

WE CLAIM:

1. A method for detecting a human chromosome in a sample, the method
5 comprising:
 - a) producing an amplicon comprising a nucleotide sequence from a target α -repeat array; and
 - b) detecting the human chromosome in the sample by detecting the amplicon comprising the nucleotide sequence from the target α -repeat array.
- 10 2. The method of claim 1 further comprising hybridizing a first oligonucleotide and a second oligonucleotide to the nucleic acid comprising the target α -repeat array.
3. The method of claim 1, wherein:
 - 15 a) the human chromosome is chromosome 1 and the target α -repeat array is D1Z5, and/or D1Z7;
 - b) the human chromosome is chromosome 2 and the target α -repeat array is D2Z1;
 - c) the human chromosome is chromosome 3 and the target α -repeat array is
20 D3Z1;
 - d) the human chromosome is chromosome 4 and the target α -repeat array is D4Z1;
 - e) the human chromosome is chromosome 5 and the target α -repeat array is D5Z1;
 - 25 f) the human chromosome is chromosome 6 and the target α -repeat array is D6Z1;
 - g) the human chromosome is chromosome 7 and the target α -repeat array is D7Z1 and/or D7Z2;
 - h) the human chromosome is chromosome 8 and the target α -repeat array is
30 D8Z2;
 - i) the human chromosome is chromosome 9 and the target α -repeat array is D9Z4;
 - j) the human chromosome is chromosome 10 and the target α -repeat array is D10Z1;

- k) the human chromosome is chromosome 11 and the target α -repeat array is D11Z1;
 - l) the human chromosome is chromosome 12 and the target α -repeat array is D12Z3;
 - 5 m) the human chromosome is chromosome 13 and the target α -repeat array is D13Z1;
 - n) the human chromosome is chromosome 14 and the target α -repeat array is D14Z1 and/or D14Z2 and/or D14Z3;
 - o) the human chromosome is chromosome 15 and the target α -repeat array is D15Z3;
 - 10 p) the human chromosome is chromosome 16 and the target α -repeat array is D16Z2;
 - q) the human chromosome is chromosome 17 and the target α -repeat array is D17Z1 and/or D171b;
 - 15 r) the human chromosome is chromosome 18 and the target α -repeat array is D18Z1 and/or D18Z2;
 - s) the human chromosome is chromosome 19 and the target α -repeat array is D19Z4 and/or D19Z5;
 - t) the human chromosome is chromosome 20 and the target α -repeat array is D20Z2;
 - 20 u) the human chromosome is chromosome 21 and the target α -repeat array is D21Z1;
 - v) the human chromosome is chromosome 22 and the target α -repeat array is D22Z4, and/or D22Z5;
 - 25 w) the human chromosome is chromosome X and the target α -repeat array is DXZ1; and/or
 - x) the human chromosome is chromosome Y and the target α -repeat array is DYZ3.
- 30 4. The method of claim 1 wherein:
- a) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6;

- b) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 7 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 8;
- 5 c) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 9 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 10;
- d) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 11 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 12;
- 10 e) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 13 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 14;
- f) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 15 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 16;
- 15 g) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 17 or 19 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 18 or 20;
- h) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 21 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 22;
- 20 i) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 23 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 24;
- 25 j) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 25 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 26;
- k) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 27 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 28;
- 30 l) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 29 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 30;
- m) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and the second oligonucleotide
- 35

comprises a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47;

n) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53;

o) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 54 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 55;

p) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 56 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 57;

q) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 58 or 60 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 59 or 61;

r) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 62 or 64 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 63 or 65;

s) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 66 or 68 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 67 or 69;

t) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 70 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 71;

u) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 72 or 74 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 73 or 75;

v) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81;

w) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 82 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 83; and/or

x) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 84 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 85.

5. The method of claim 1 further comprising inhibiting amplification of an amplicon comprising a nucleotide sequence from a non-target α -repeat array.
- 5 6. The method of claim 5 wherein inhibiting amplification of an amplicon comprising a nucleotide sequence from a non-target α -repeat array comprises hybridizing a clamp oligonucleotide to a nucleic acid comprising the non-target α -repeat array.
- 10 7. The method of claim 1 wherein the first oligonucleotide and/or the second oligonucleotide comprises a locked nucleic acid nucleotide.
8. The method of claim 1 wherein a centromere core comprises the target α -repeat array.
- 15 9. The method of claim 1 wherein a pericentromeric region comprises the target α -repeat array.
10. The method of claim 1 wherein the target α -repeat array is p82H.
- 20 11. The method of claim 1 wherein the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 86 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 87.
- 25 12. The method of claim 1 further comprising quantifying the amplicons produced from the target α -repeat array to provide a quantity of α -repeat amplicons.
- 13 13. The method of claim 12 further comprising calculating the relative number of α -repeats in the α -repeat array per human diploid genome by comparing the
- 30 quantity of α -repeat amplicons to the quantity of a single-copy gene.
14. The method of claim 13 further comprising calculating the ploidy of the human chromosome in the sample by comparing the relative number of α -repeats in the α -repeat array per human diploid genome calculated for the sample to a relative

number of α -repeats in the α -repeat array per human diploid genome for a normal sample.

15. The method of claim 14 wherein the relative number of α -repeats in the α -repeat array per human diploid genome for a normal sample is a previously determined known value.

16. The method of claim 14 wherein the relative number of α -repeats in the α -repeat array per human diploid genome for a normal sample is determined experimentally using a sample from a normal human.

17. The method of claim 14 wherein the ploidy of the human chromosome indicates that the human source of the sample has an aneuploidy.

18. The method of claim 14 wherein the ploidy of the human chromosome indicates that the human source of the sample is male.

19. The method of claim 14 wherein the ploidy of the human chromosome indicates that the human source of the sample is female.

20. The method of claim 17 wherein the aneuploidy is a trisomy.

21. The method of claim 16 wherein the aneuploidy is trisomy 21, trisomy 13, trisomy 8, trisomy 18, or trisomy X.

22. A method for detecting an aneuploidy of a chromosome in a subject, the method comprising:

- a) determining a pericentromere size of a chromosome;
- b) comparing the pericentromere size of the chromosome to the pericentromere size for the chromosome of a normal subject; and
- c) identifying the chromosome as aneuploid by determining that the pericentromere size of the chromosome of the subject is smaller than the pericentromere size for the chromosome of a normal subject.

23. The method of claim 22 wherein determining a pericentromere size of a chromosome comprises quantifying copy number of K111 and/or K222.
24. The method of claim 22 wherein determining a pericentromere size of a chromosome comprises amplifying K111 using a first oligonucleotide comprising a sequence provided by SEQ ID NO: 98 or 104 and a second oligonucleotide comprising a sequence provided by SEQ ID NO: 99 or 105.
25. The method of claim 22 wherein determining a pericentromere size of a chromosome comprises amplifying K222 using a first oligonucleotide comprising a sequence provided by SEQ ID NO: 101 and a second oligonucleotide comprising a sequence provided by SEQ ID NO: 102.
26. A method for determining the number of α -repeats in an α -repeat array of a chromosome, the method comprising:
- hybridizing a first oligonucleotide and a second oligonucleotide to a nucleic acid comprising a target α -repeat array;
 - producing an amplicon comprising a nucleotide sequence from the target α -repeat array;
 - quantifying the amplicon; and
 - calculating the number of α -repeats in the α -repeat array of the chromosome from the quantity of the amplicon.
27. The method of claim 26 wherein:
- the human chromosome is chromosome 1 and the target α -repeat array is D1Z5, and/or D1Z7;
 - the human chromosome is chromosome 2 and the target α -repeat array is D2Z1;
 - the human chromosome is chromosome 3 and the target α -repeat array is D3Z1;
 - the human chromosome is chromosome 4 and the target α -repeat array is D4Z1;
 - the human chromosome is chromosome 5 and the target α -repeat array is D5Z1;

- 5 f) the human chromosome is chromosome 6 and the target α -repeat array is D6Z1;
- g) the human chromosome is chromosome 7 and the target α -repeat array is D7Z1 and/or D7Z2;
- h) the human chromosome is chromosome 8 and the target α -repeat array is D8Z2;
- i) the human chromosome is chromosome 9 and the target α -repeat array is D9Z4;
- 10 j) the human chromosome is chromosome 10 and the target α -repeat array is D10Z1;
- k) the human chromosome is chromosome 11 and the target α -repeat array is D11Z1;
- l) the human chromosome is chromosome 12 and the target α -repeat array is D12Z3;
- 15 m) the human chromosome is chromosome 13 and the target α -repeat array is D13Z1;
- n) the human chromosome is chromosome 14 and the target α -repeat array is D14Z1 and/or D14Z2 and/or D14Z3;
- o) the human chromosome is chromosome 15 and the target α -repeat array is D15Z3;
- 20 p) the human chromosome is chromosome 16 and the target α -repeat array is D16Z2;
- q) the human chromosome is chromosome 17 and the target α -repeat array is D17Z1 and/or D17b;
- 25 r) the human chromosome is chromosome 18 and the target α -repeat array is D18Z1 and/or D18Z2;
- s) the human chromosome is chromosome 19 and the target α -repeat array is D19Z4 and/or D19Z5;
- t) the human chromosome is chromosome 20 and the target α -repeat array is D20Z2;
- 30 u) the human chromosome is chromosome 21 and the target α -repeat array is D21Z1;
- v) the human chromosome is chromosome 22 and the target α -repeat array is D22Z4, and/or D22Z5;

- w) the human chromosome is chromosome X and the target α -repeat array is DXZ1; and/or
- x) the human chromosome is chromosome Y and the target α -repeat array is DYZ3.

5

28. The method of claim 26 wherein:

- a) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6;
- 10 b) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 7 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 8;
- c) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 9 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 10;
- 15 d) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 11 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 12;
- e) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 13 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 14;
- 20 f) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 15 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 16;
- 25 g) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 17 or 19 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 18 or 20;
- h) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 21 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 22;
- 30 i) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 23 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 24;

- j) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 25 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 26;
- 5 k) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 27 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 28;
- l) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 29 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 30;
- 10 m) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47;
- n) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53;
- 15 o) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 54 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 55;
- 20 p) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 56 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 57;
- q) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 58 or 60 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 59 or 61;
- 25 r) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 62 or 64 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 63 or 65;
- s) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 66 or 68 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 67 or 69;
- 30 t) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 70 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 71;

- u) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 72 or 74 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 73 or 75;
- 5 v) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81;
- w) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 82 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 83; and/or
- 10 x) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 84 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 85.
29. A reaction mixture comprising a nucleotide sequence from a target α -repeat array.
- 15 30. The reaction mixture of claim 29 further comprising an amplicon comprising the nucleotide sequence from the target α -repeat array.
- 20 31. The reaction mixture of claim 29 further comprising a first oligonucleotide complementary to the nucleotide sequence from the target α -repeat array.
32. The reaction mixture of claim 29 further comprising a second oligonucleotide complementary to the nucleotide sequence from the target α -repeat array.
- 25 33. A composition comprising one or more of:
- a) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6;
- 30 b) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 7 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 8;
- c) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 9 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 10;
- 35

- d) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 11 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 12;
- 5 e) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 13 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 14;
- f) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 15 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 16;
- 10 g) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 17 or 19 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 18 or 20;
- h) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 21 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 22;
- 15 i) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 23 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 24;
- j) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 25 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 26;
- 20 k) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 27 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 28;
- 25 l) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 29 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 30;
- m) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47;
- 30 n) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53;

- o) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 54 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 55;
- 5 p) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 56 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 57;
- q) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 58 or 60 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 59 or 61;
- 10 r) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 62 or 64 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 63 or 65;
- s) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 66 or 68 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 67 or 69;
- 15 t) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 70 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 71;
- u) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 72 or 74 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 73 or 75;
- 20 v) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81;
- 25 w) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 82 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 83; and/or
- x) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 84 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 85.
- 30

34. A composition comprising a pair of oligonucleotides specific for one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1 and/or

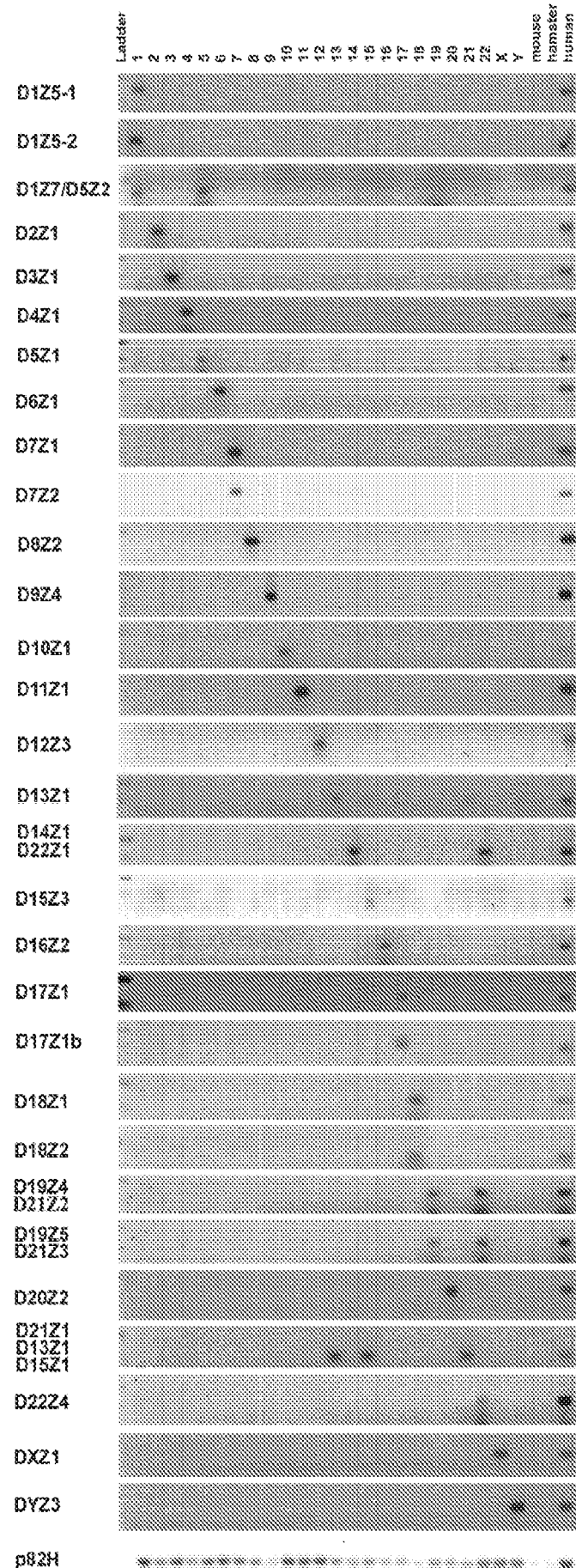
D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2;
D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3.

35. The composition of claim 34 further comprising a nucleic acid comprising a
5 sequence from an α -repeat array from a human chromosome.
36. The composition of claim 34 further comprising a nucleic acid comprising a
sequence from one or more of the following α -repeat arrays: D1Z5, and/or D1Z7;
D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1;
10 D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or
D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or
D22Z5; DXZ1; and/or DYZ3.
37. The composition of claim 34 further comprising a polymerase.
15
38. The composition of claim 34 further comprising an amplicon produced from one
or more α -repeat arrays.
39. A kit comprising one or more of:
20
 - a) a first oligonucleotide comprising a nucleotide sequence provided by SEQ
ID NO: 1, 3, or 5 and a second oligonucleotide comprising a nucleotide
sequence provided by SEQ ID NO: 2, 4, or 6;
 - b) a first oligonucleotide comprising a nucleotide sequence provided by SEQ
ID NO: 7 and a second oligonucleotide comprising a nucleotide sequence
25 provided by SEQ ID NO: 8;
 - c) a first oligonucleotide comprising a nucleotide sequence provided by SEQ
ID NO: 9 and a second oligonucleotide comprising a nucleotide sequence
provided by SEQ ID NO: 10;
 - d) a first oligonucleotide comprising a nucleotide sequence provided by SEQ
30 ID NO: 11 and a second oligonucleotide comprising a nucleotide sequence
provided by SEQ ID NO: 12;
 - e) a first oligonucleotide comprising a nucleotide sequence provided by SEQ
ID NO: 13 and a second oligonucleotide comprising a nucleotide sequence
provided by SEQ ID NO: 14;

- 5
- f) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 15 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 16;
- g) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 17 or 19 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 18 or 20;
- h) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 21 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 22;
- 10 i) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 23 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 24;
- j) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 25 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 26;
- 15 k) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 27 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 28;
- l) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 29 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 30;
- 20 m) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47;
- 25 n) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53;
- o) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 54 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 55;
- 30 p) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 56 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 57;

- q) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 58 or 60 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 59 or 61;
- 5 r) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 62 or 64 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 63 or 65;
- s) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 66 or 68 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 67 or 69;
- 10 t) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 70 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 71;
- u) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 72 or 74 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 73 or 75;
- 15 v) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81;
- w) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 82 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 83; and/or
- 20 x) a first oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 84 and a second oligonucleotide comprising a nucleotide sequence provided by SEQ ID NO: 85.
- 25
40. A kit comprising one or more pairs of oligonucleotides specific for one or more of the following α -repeat arrays: D1Z5, and/or D1Z7; D2Z1; D3Z1; D4Z1; D5Z1; D6Z1; D7Z1 and/or D7Z2; D8Z2; D9Z4; D10Z1; D11Z1; D12Z3; D13Z1; D14Z1 and/or D14Z2 and/or D14Z3; D15Z3; D16Z2; D17Z1 and/or D171b; D18Z1 and/or D18Z2; D19Z4 and/or D19Z5; D20Z2; D21Z1; D22Z4, and/or D22Z5; DXZ1; and/or DYZ3.
- 30

FIG. 1



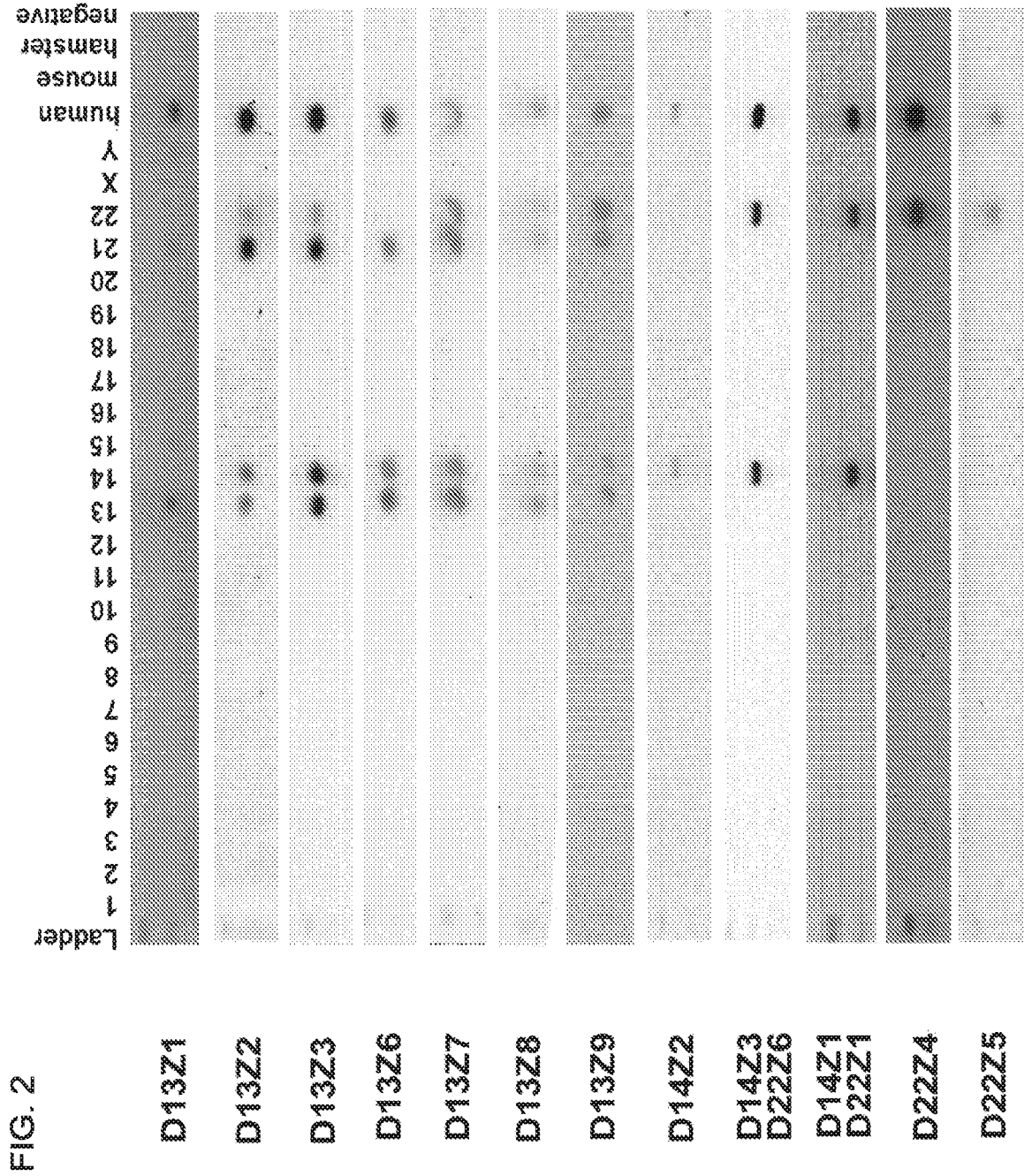


FIG. 3

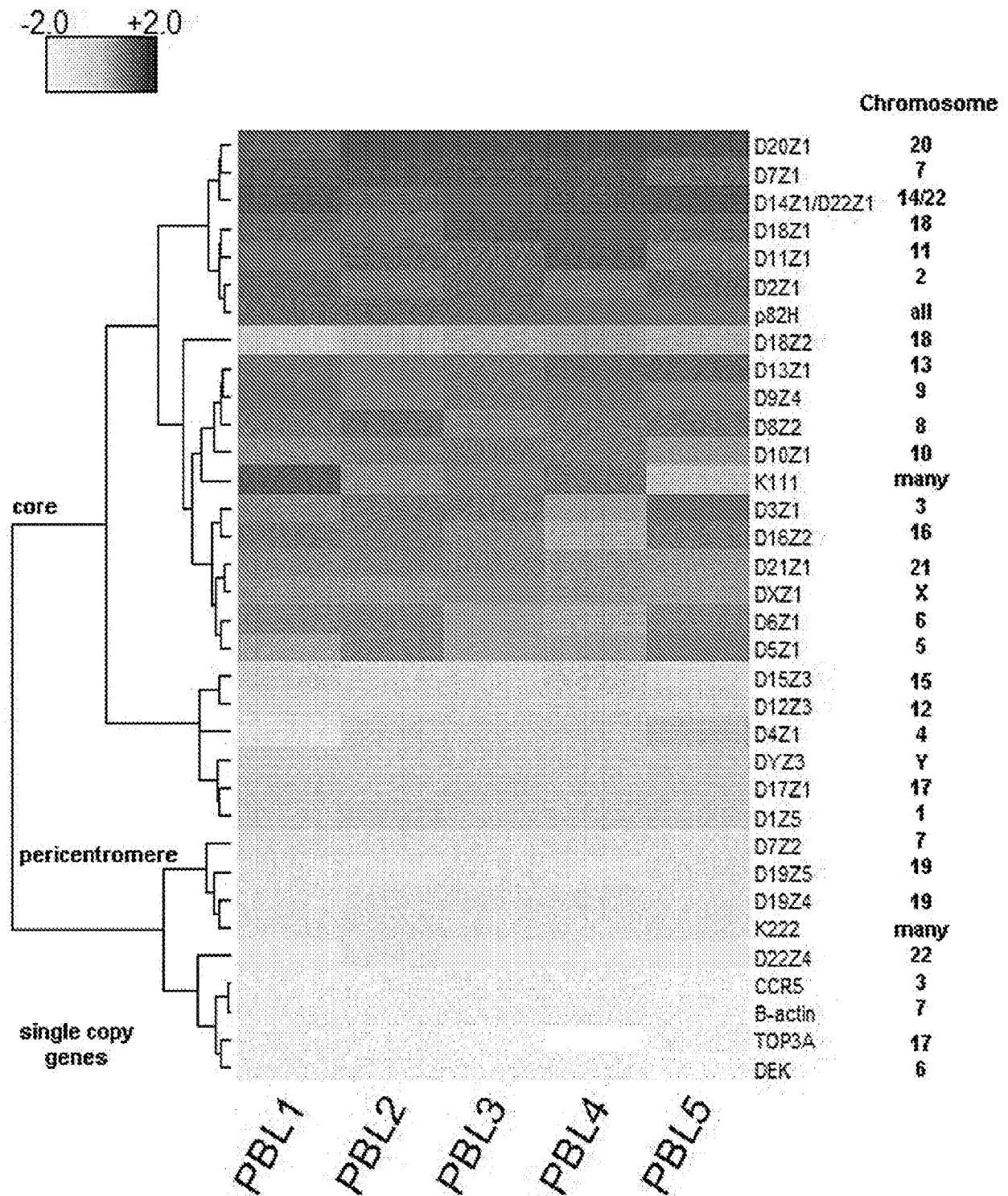


FIG. 4

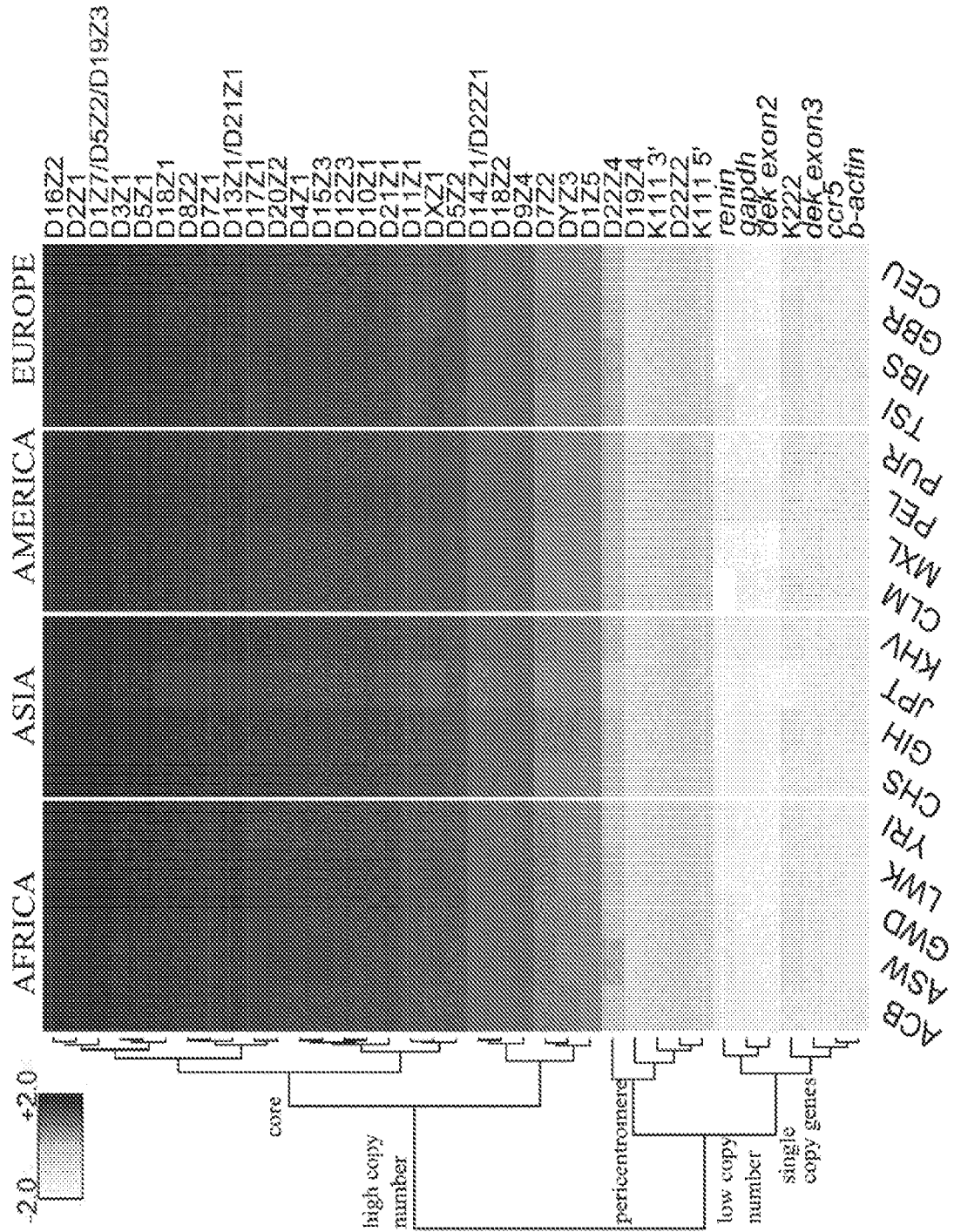


FIG. 5A

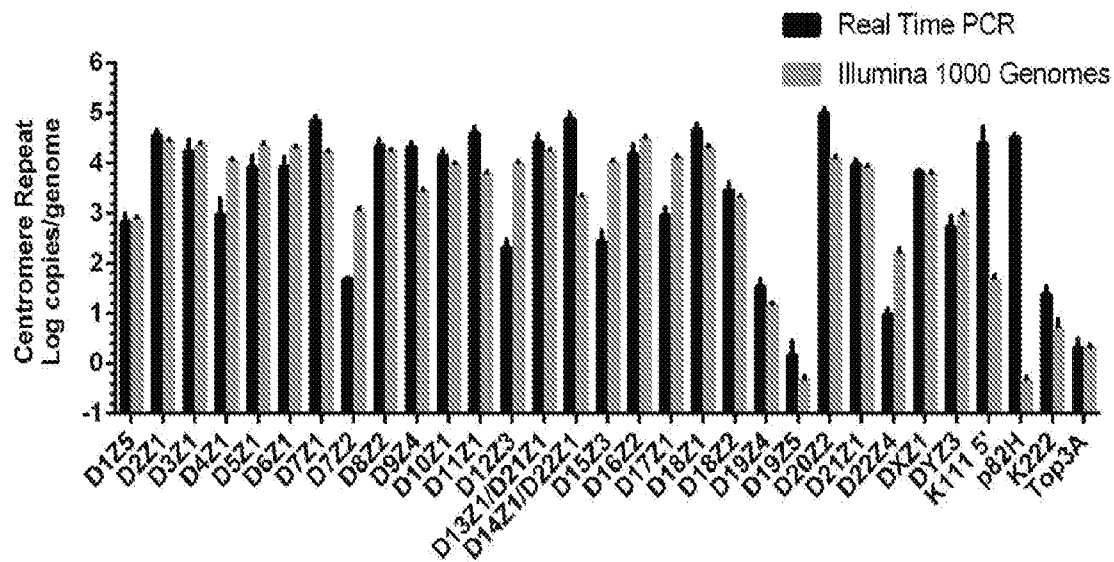


FIG. 5B

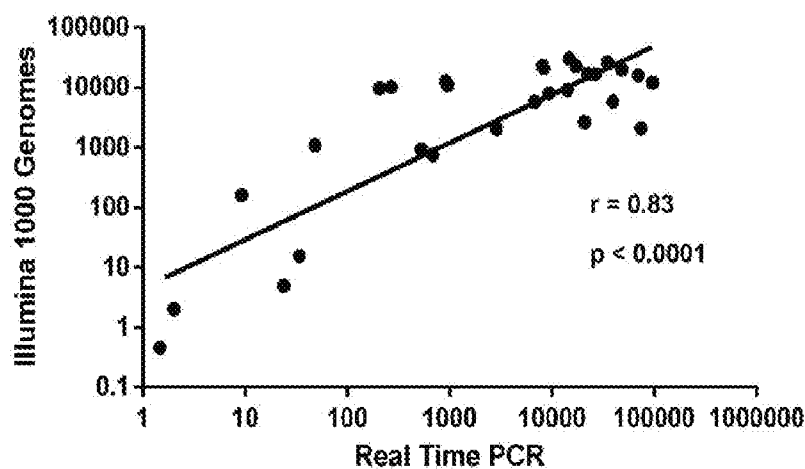


FIG. 5C

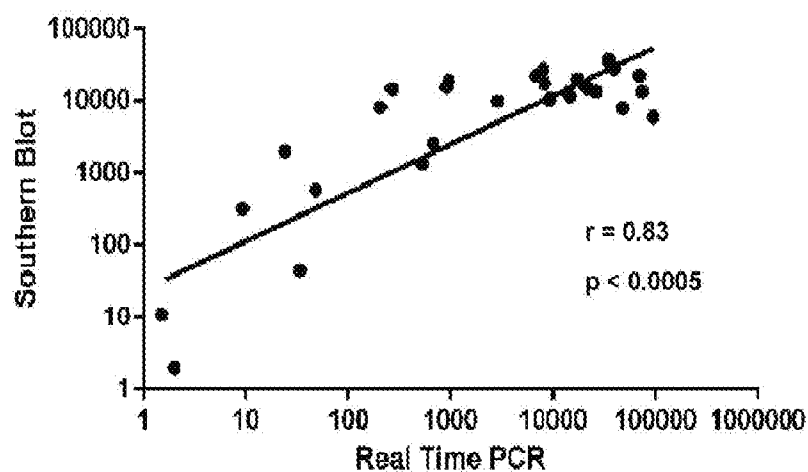
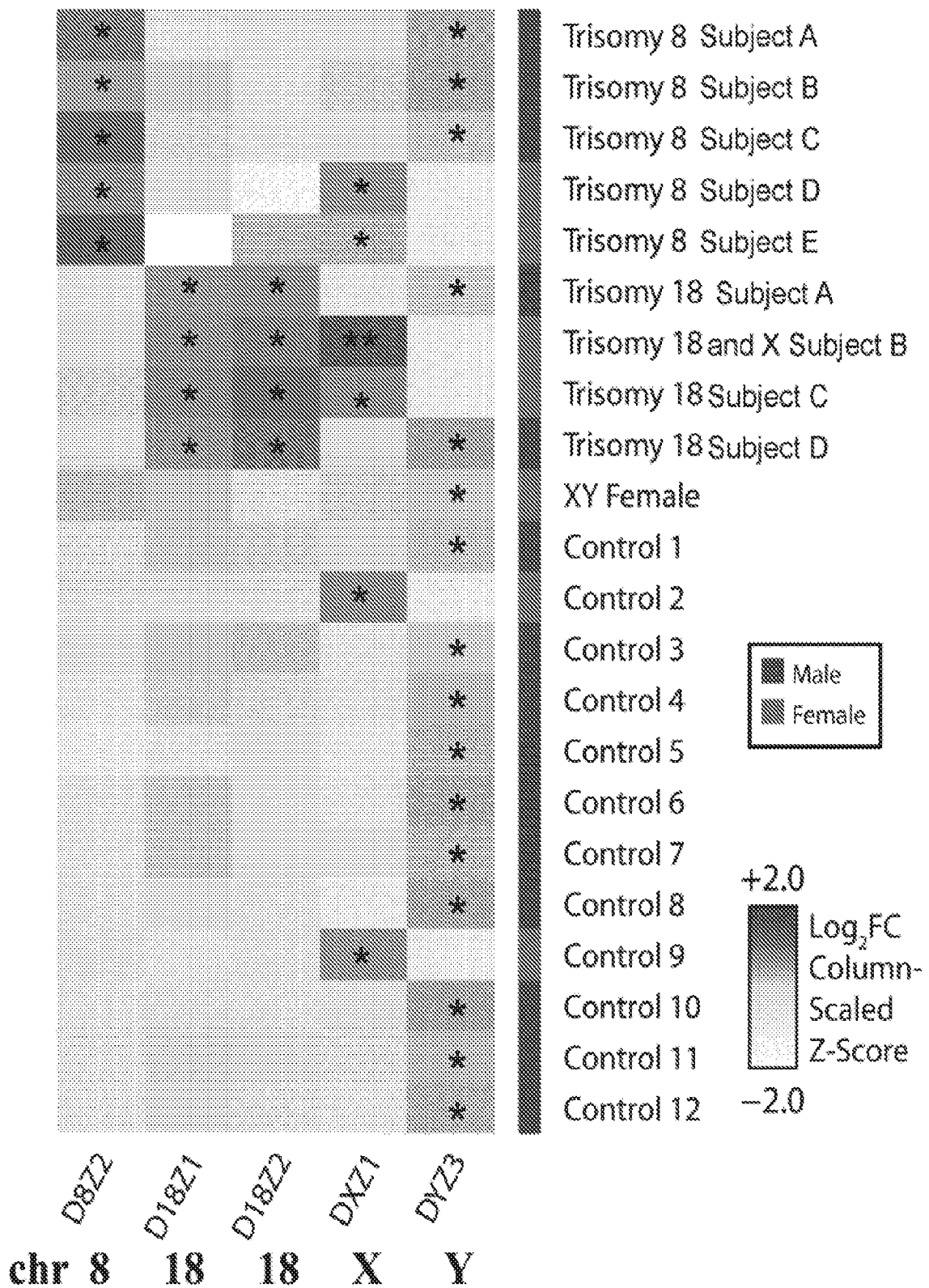


FIG. 6



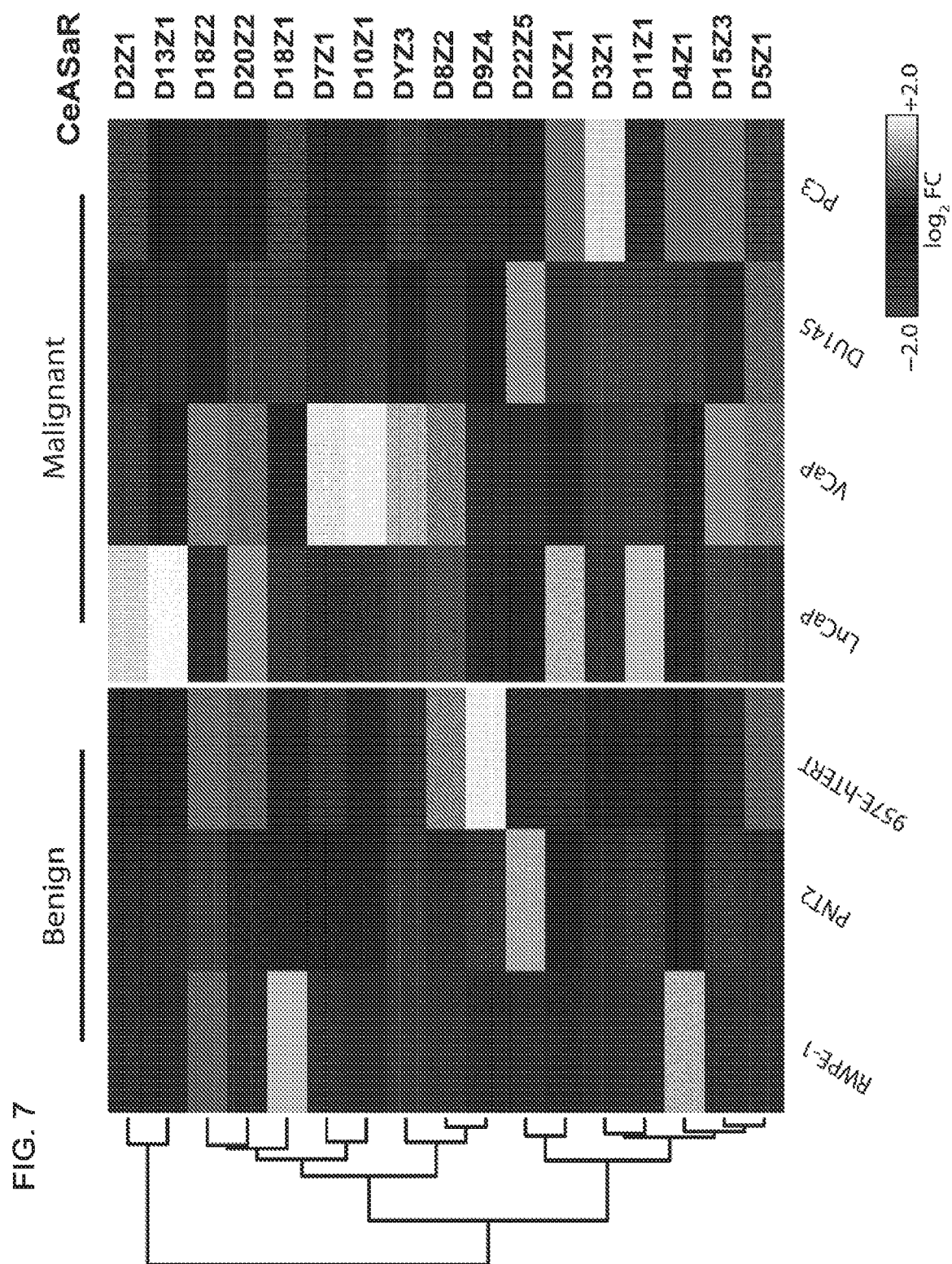


FIG. 8A

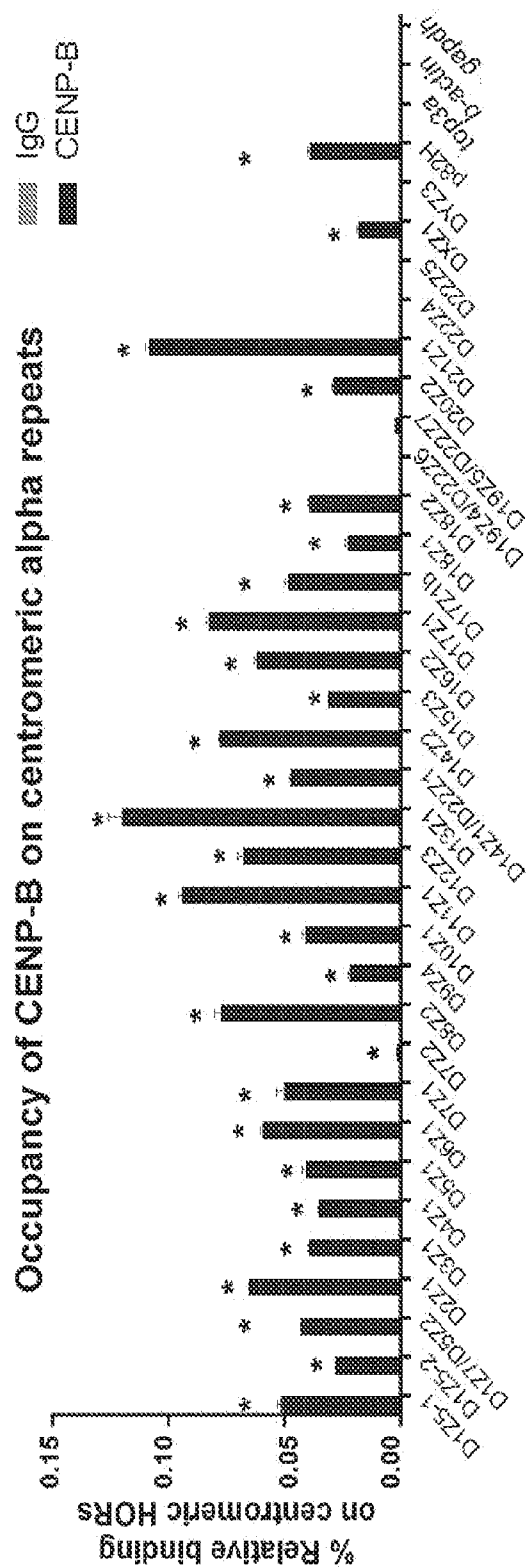
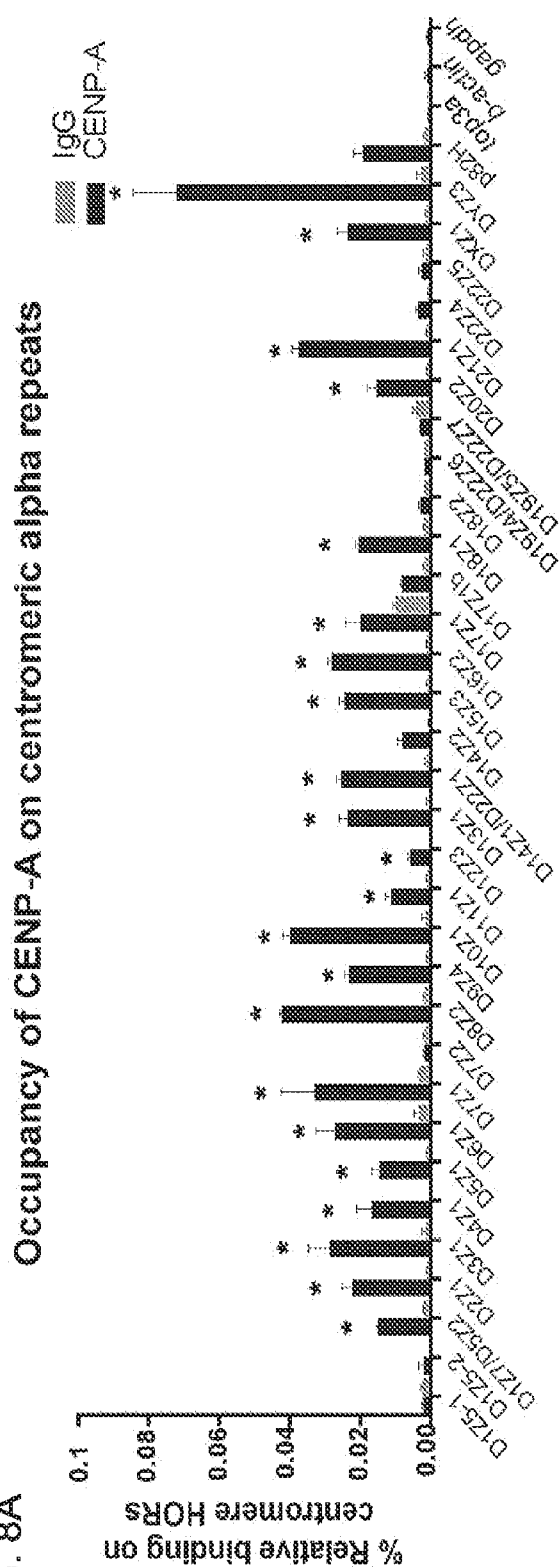


FIG. 9A

Box	sequence	delimiter
5 CENP-B Box 1	<u>CTTCGTTGGAAACGGGA</u>	/
CENP-B Box 2	<u>CTTCGTTGGAAACGGGT</u>	()
CENP-B Box 3	<u>TTTCGTTGGAAACGGGA</u>	[]
CENP-B Box 4	<u>TTTCGTTGGAAACGGGA</u>	{ }
Mutated boxes		* *

>D1Z5 Acc No. BX248407.26 (SEQ ID NO: 112)

CTCATCAAAGCTACATGGAAGCATTCTCAGAACTGGTTTGTGATGATTGCATTCAAGTCACCGAG
TTGAACATTCCCTTTGATGGTTCCGTTTGGAAACACACTTTTGGTAGAATCTTAAAGGGGAGATTT
GAACCGCTTTGAGGCCTATGGCAGTAGAGGATATAACTGCACATAAAAACGAGACAGGAGCATTCC
5 CAGGAAACACTTTGTGACGATTGAGTTCAATTCACAGAGCTGAACATTCCCTTTGGATGGAGCAGTT
TCAAAACACACTTTTTGTAGAATCTGCAAGTGGAGATTTGGACCGCTCTGAGGA*TTTCATTGGAT
ACGGGA*GAAAACTCACCTATGTAAACAGAAGCATTCTCAGAACCTTCTTCGTGATGCTTGCATT
AACTCACAGTGTTCAACTTTCTCTGATAGTTTCAGGTTTTTAAACACTCCTTCTGCAGAATCTGCAA
GTGGAGATTTGGACCTCTTTGAGGCCTATCGTAGTAAAGGATATAACCTCATCTAAAAACAAGACA
10 GAAGCATTCTCAGAAAATACTTTGTGATGATTGAGTTTAAACACACAGAGCTGAGCATATCTTTTGA
TGGAGCATTTTGAAAACACACCTTTTGTAGAATATGCAAGTGGATTTTGGGACTTCTCTGAGAA{T
TTTCGTTGGAAACGGGA}TAAACCTCACATAACTGAAGAGGAACATTCTCAGAAAGTTCTTGGTGACG
TTGGCATTCAACTGACAGAGTTGAACCTTCCCTTGTGAGTTCAGGTTGAAA

>D1Z7 D5Z2 D19Z3 Acc No. AJ295044.1 (SEQ ID NO: 113)

15 GAATTCCCAGTAACTTCCTTGTGTTGTGTGCATTCAACTCACAGAGTTGAACGTTCCCTTAGACAG
AGCAGATTTGAAACACTCTATTTGTGCAATTTGCAAGTGTAGATTTCAAGCGCTTTAAGGTCAATG
GCAGAAAAGGAAATATCTTCGTTTCAAACTAGACAGAATCATTCCCACAACTGCGTTGTGATGT
GTTTCGTTCAACTCACAGAGTTTAACTTTCTGTTCATAGAGCAGTTAGGAAACACTCTGTTTGTA
AGTCTGCAAGTGGATATACAGACCTCCTTGAGC/CTTCGTTGGAAACGGGA/TTTCTTCATATTCT
20 GCTAGACAGAATAATTCTCAGTAACTTCCTTGTGTTGTGTGTATTCAACTCACAGAGTTGAACGAT
CCTTTACACAGAGCAGACTTGAAACACTCTTTTTATGGAATTTGCAAGTGGAGATTTTCAGCCGCTT
TGAGGTCAATGGTAGAATAGGAAATATCTTCCTATAGAACTAGACAGAATGATTCTCAGAACTC
CTTTGTGATGTGTGCGTTCAACTCACAGAGTTTAAACCGTTCTGTTCATAGAGCAGTTAGGAAACAC
TCTGTTTGTAAGTCTGCAAGTGGATATTCAGACCTCTTTGAGGC/CTTCGTTGGAAACGGGA/TT
25 TCTTCATATTCTGCTAGACAGAAGAATTC

>D2Z1 Acc No. J04773.1 (SEQ ID NO: 114)

CCTATGGTGAAAAAGGAAATATCTTCCCCTGAAAACTAGACAGAAGCATTCTCAGAACTCTTATTTG
TGATGTGCGCCCTCAACTAACAGTGTGGAAGCTTTCTTTTGATAGAGCAGTTTGGAAACACTCTTT
TCGTAAAATCTGCAAGAGGATATTTTGATAGCTTTGAGGA{TTCGTTGGAAACGGGA}TTGTCTT
30 CATATAAACTCTAGACAGAAGCATTCTCAGAAGCTTCATTGGGATGTTTCAATTGAAGTCACAGTG
TTGAACAGTCCCTTTTCATAGAGCAGGTTTGAACACTCTTTTTGTAGTATCTGGAAGTGGACATTT
GGAGCGCTCTCAGGACTACGTGGAAAAAGGAAGTATCTTCCAATAAAAAGCTAGATAGAAGCAATGT
CAGAACTTTTTTCATGATGTATCTACTCAGCTAACAGAGTTGAACCTTTCTTTTGAGAGAGCAGTT
TTGAAACACTCTTTTTGTGGAATCTGCAAGTGGATATTTCTCTAGCTTTGAGGA{TTCGTTGGAA
35 ACGGGA}TTACATATAAAAAGCAGACAGCAGCATTCCCAGAACTTCTTTGTGATGTTTGCATTCA
AGTCACACAGTTAAACATTCCCTTTTCATAGAGCAGGTTTGAACACTCTTTTTGTAGTATCTGGAT
GTGGACATTTGGAGCGCTTTCAGGTCTATGGTGAAAAAGGAAATATCTTCCCCTCAAAAGTAGACA
GAAGCATTCTCAGAACTTATGTGTGATGTGCGCCCTCAACTAAGAGTGTGGAACCTTTCTTTTGA
TAGAGCAGTTTGAACACCCTTTTTGTAAAATCTGCAAGAGGATATTTGGATAGCTTTGAGGA{T

TTCGTTGGAAACGGGA} TTGTCTTCATATAAACTCTAGACAGAAGCATTCTCAGAAGCTTCATTGG
 GATGTTTCAATTGAAGTCACAGTGTTGAACAGTCCCTTTTCATAGAGCAGGTTTGAAACACTCTTTT
 TGTAATATCTGGAAGTGGACATTTGGAGAGATCTCAGGAATACGGTGATAAAGGAAATATCTTCCA
 ATAAAAGCTAGATAGAAGCAATGTCAGATACTTTTTTCATGATGTATCTACTCAGCTAACAGAGTTG
 5 AACCTTTCTTTTGAGAGAGCAGTTTTGAAACACTCTTTTTGTGGAATCTGCAAGTGGATATTTGTC
 TAGCTTTGAGGA { TTTCGTTGGAAACGGGA } TTACATATAAAAAGCAGACAGCAGCATTCCCGAA
 ACTTCTTTGTGAAGTTTGCATTCAAGTCACAGAGTTGAACATTCCCTTTTCATAGAGCAGGTTTGAA
 AACTCTTTTCGTAGTATCTGGATGTGGACATTTGGAGCGCTTTCAGG

>D3Z1 Acc No. Z12006.1 (SEQ ID NO: 115)

10 CTGCAGCTGGATATTTGGACCTCTTTGAGGC*CTTCGTAGTAAACGGGA*TTTCTTCGTGTAATGA
 TAGACAATAGAATTCTCAGTGAATTTTTTCTGTGTGTGTATTCAACTCACAGGGTTGAACCTT
 CCTTTAGACAGTGCAGATTTGAGACACTTGTCTGTGGAATTTGCAAGGGGAGATTTCAAGCACTTT
 GAGGCCATTGGTGGAAAAGGAAATATCTTCGTATGAAAACCTAGACAGAATCATTCTCAGGAACCTAC
 TTTGTGATATGTGCATTCAACTCCCAGAGTTTAACTTTCTTTTCATAGATGAGTTTGGAACAGT
 15 CAGTTTGTAATTTCTGCAACTGGATATTTGGACCTCTTTGAGGC { TTTCGTTGGAAACGGGA } TTT
 CTTCACATAATGCTAGACAGAAGAATTCTCAGTAACTTCTTTTGGGATGTATGTATTCAAATCAGA
 GAGTTGAACCTTCCTTTAGACAGAGCGGATTGGAACACTCTTTTTGTGGAATTTGCAAGTGGAAA
 ATTCTAGCAGTATGAGGCCAATGGTACAAAAGGAAATATCTTCGTATAAAAACCTAGACAGTATCAT
 TCTCAGAAACTGCTTTGTGATGTGTGTATTAACTCACAGAGTTGAACATTTCTTTGCATAGAGCA
 20 GTTTGGAAAGACTTAGTTTGTGCAGTGTGCAAGTGGATATTTGGAACCTCTTTGAGGC/CTTCGTTG
GAAACGGGA/TTTCTTCTTATAATTTCTTGAAAAAGAATTCTCAGTAGCTTCTTTGTGTGTGTGT
 ATTCAACTCACAGAGTTGAACCTTCCTTTAGACAGAGCAGATTGGAACACTCTTTTTGTGGAATT
 TGCAAGTGGAGAATTCTAGCGCTTTGACGCCAATGGTAGAAAAGGAAATATCTTCGTATAAAAACCTA
 GACAGTATCATTCTCAGAAGCTACTTTGTGATGTGTGCGTTCAACTCACAGAGTTTAACTTTCTT
 25 TTCATAGAGCAGTTTGGAACCCCTCTGTTTGTGAAGTCTGCAAGTGGATATTTAAACGTCTTTGAG
 GC/CTTCGTTGGAAACGGGA/TTTTTTTCATATAAACCAGGACAGAAGAATTCTCAGAAACGTCTTG
 ATTGTTATGTGTGCATTGGCTCACAGAGTTGAACCTTACTTTGGAAAGAGCAGTTTCTAACACT
 CTTTTTGTAAGTTTCCAAGTGAATACTTTGAGTGCTTTGAAGCCTACGGTTGACAACGAAATATC
 TTCCTGTAAAAACTACAAAGAATCATTCCGAGAAACCACGTTGTGATCTCTGCATTCAACTCACAG
 30 AGTTCAACCTTTCTTCTATAGAGCAGTTATGAAACAGTCTCTTTGTAGAATTTGCAAGGGTGTAT
 TTAGAGGGCATTGAAGCCTACGGTAGAAAAGGAAATATCTTACCATAAAATCTAGTCAGAAGCATT
 CTCAGCAACTGAGTTGTGATGTTTGCATTCAACTCACAGAGTTCAACATTCCTTTTAATGGAGCGG
 TTTTGAAACACTCTTTTTGCAGAATCTGCAAGTGGATATTTGGACCTCTTTGAGGC/CTTCGTTGG
AAACGGGA/TTTCTTCATGTAATGCCAGACAGAAGAATTCTCAGTGAATTCTTTCTGTGTGTGTGT
 35 ATTCAACTCACAGAGTTGAACGTTCTTTAGACAGAGTAGATTGGAACACTCTTTTTGTGGAATT
 TTCAGGTGGAGGTATCAAGCGCTTTGAGGCCAATGATAGAAAAGGAAATACCTTCGTATAATAATT
 AGACGGAATCATTCTCAGAAACCGCTTTGCAATGTGTGCGTTCAACTCACAGTGTTTAACTTTCT
 TTTCATACAAGTTGTTTCGAAACACTCTTTTTGCAGAATCTGCAAGTGGATATTTGGACCTCTTTG
 AAGT*CTTCGTTGGAAATGGGA*TTTCTTCATATAATGCTAGACAGAAGACTTCTCAGTAACTGCT
 40 TTTTCTGGTGTGTATTCAACTCTCAGAGTTGAACTTTCCTTTAGAAACAGCAGATTTGAAACTCTC

FIG. 9B Continued

TTTTTGTGGAATTTGCAAGTGGAGATTTTCAGAGCTTTGAGGCCAATGGTAGAAAAGGAAATATCTT
CGTATGCAAACCTAGACAGAATCATTCTCAGAACTACTTTGGTACGTGTGTGTTCAACTCACAGTG
TTTAACCTTTCTTTTCATAGAGCAGTTTGGAAACACTCAGTTTGTAAAGTCAGCAACTGGATATTT
GGATGTATTTGAGGC/CTTCGTTGGAAACGGGA/TTTCTTCATATAATGCTAGAAATTC AATTCTCA
5 GTAACCTCTTTGGGTTGTGGGTATTCAAGTCACAGAGTTGAAGCTTCCTTTAGGCGGTTGGAAACA
CTTTTTGTGGAATTTTCAGGGGGAGACTTCAAGCGCTTTGAAGTGAATGGTAGGAAAGGAAATACC
TTCGTATAAAAACTAGACGGAGTCATTCTCAGAACTACTTTGTGATGTTTGC GTTCAACTCACAG
AGTTTAACGTTTCTTTTCATAGAGCAGTTTGGAAACACTCTTTTGCAGAATCTGCAAGTGGATAT
TTGGACCTCTTTGTGGC/CTTCGTTGGAAACGGGA/TTTTTCATATAATGCTAGACAGAAGAATTC
10 TCAGTAACCTCTTTTGTGGTGTGTATTCAACTCACAGAGTTGAACCTTCCTTTAGACAGAGCAGA
TTTGAAACTCTCTTTTGTGGAATTTGCAAGTGGAGATTTCAAGCGCTTTGAGGCCAACGGCAGAA
AAGGAAATATCTTCGTAGAAAAAATAGACGGCATCATTCTCAGAACTGCTTGGGATGTGTGATTG
AACTCACAGTGTTTAACACTTCTTTTCATAGAGCACTTTGGAAACACTCAGGTTGTAATGTCTGCA
G

15 >D4Z1 Acc No. Z12011.1 (SEQ ID NO: 116)

CTGCAGGTGGATATGTGGATAGCTTTGAAGATTTTCGTTGGAAACCGGAATATCTTCCTATAAAATC
AAGACAGAAGCATTCTCGGAAACATCTCTGTGATGTTTGCATTCAACTCAGTAGAGTTGAACACGT
CCTTTTCATAGAGCAGGTTTGAACACTCTTTCTGCCCTACCTGGAAGCGGACATTTTCGAGCGCTTT
GAGGCCTATGGTGA AAAAGGAAATATCTTCTCATAAAAACCAGAAAGAAGCATTCTCAGAAACTTC
20 TTTGTGTTGTGTGTACTCAAGTAACAGTGTTGAACCTTCCTTTTGACAGAGCAGTTTGAACACT
CTTTTGGTAGAATCTGCAAGTGGATATTTGGATAGCTTTGAGGATTTTCGTTGGAAACGGGTTATCT
TCATATAAAATCCAGACAGGAGCATTCTCAGAACTTCTTTGTGCTGTATGTCCTCAATTCACAGA
GCTGAACCTTTGTTTGGATACAGCATTTTGGAGACATTCCTTTAGTAGAATCTGCAAGTTGATATT
TAGATAGCTTTGAAGA { TTTCGTTGGAAACGGGA } ATATCTTCATAGAAAATCTAGACGGAAGCAT
25 TCTCAGAACTGCTTTGTGATGTTTGCATTCAAGTCACAGAGTTGAATATTCCCTTTTATAGAGTA
GGTTTGAACACTCTTTTCGGCACTACCTGGAAGTGGATATTTTCGAGCTCTTTGAGGCCTATGGTTA
AAAGGAAATATCTTCCCATAAAAACCTAGACAGAAGCCGTCTCAGAAACTTGTTTGTGATGTGTGTA
TTCAACTACCAGAGTTGAACATTTCTGTTACAGAGCAATTTTAAACACTCTTTTGTGGAATCTG
AAAGTGGATAATTGGATAGCTTTGTGGA { TTTCGTTGGAAACGGGA } TGACGTATAAAATCTAGAG
30 AGAAGCATTCTCAGGAACTTCTTTCTGATGTTTGCATTCAAGTCACAGAATTGAACATTCCTTTTC
AGAGTGCAGGTTTGAACACTCTTTCTGTAGTATCTGGAAGTGGACATTTCAAGCGCTTTTCAGGCC
TACGGGGAGAAAGGAAATATCTTCAAATAAAAACCTAGACAGAAGGATTCTCAGAACTTATTTGTG
ATGTGTGTCCTAAACGAACACAGTTGAACCTTTGTTTTGATACAGCATTTTGGAAACACTCCTTTT
GTAGGATCTGCAG

35 >D5Z1 Acc no. M26920.1 (SEQ ID NO: 117)

AAATTTTCTTTTCATACAGCAGAGTTTGGAAACACTCTGTTTGTAAAGTCTGCACGTGGATAAGTT
GTCCACTTAGAGGC*ATTCGTTGGAAACGGGT*TTTTTTCATGTAAGGCTACACAGAAGAATTCCC
AGTAACTTCCTTGTGTTGTGTGTATCAACTCAAAGAGTTGAACGATCCTTTACACAGAGCAGACTT
CTAACACTCTTTTGTGGAATTTGCAAGTGGAGATTTTCAGCCGCTTTGAAGTCAAAGGTAGAAAAG

FIG. 9B Continued

GAAATATCTTCCTATAAAAACTAGACAGAATGATTCTCAGAAACTCCTTTGTGATGTGTGCGTTCA
ACTCACAGAGTTTAACCTTTCTTTTCATAGAGCAGTTAGGAAACACTCTGTTTGTAAGTCTGCAA
GTGGATATTCAGACCTCTTTGAGGC*CTTCGTTGGAAACGGGT*TTTCATATAAGGCTAGGCAGAGAA
TTCCCAGTAACTTCCTTGTGTTGTGTGTGCAACTCACAGAGTTGACTTTTCAATTTACACAGAGCAG
5 ACTTGAAACACTCTTTTTTGTAATTGCAAGTGGAGATTTCAAGCGCTTTGAGCAAGGCCGAAAAGGA
AATATCTTCGTATAAAAACTAGACAGAATCATTCTCAGAAACTGCTCTGCGATGTGTGCGTTCAAC
TCTCAGAGTTTAACCTTTCTTTTCATCAGCAGTTTGAAACACTCTGTTTGTAAGTCTGCACGTG
GATATTTTGACCACCTTAGAGGC (CTTCGTTGGAAACGGGT) TTTTTCTGTAAAGGCTAGACAGAA
GAATTCCTGTAGCTTCCTTGTGTTGTGTACATTCAACGCACAGAGTTGAACGTTCCCTTAGACAG
10 AGCAGATTTGAAACACTCTTTTTGTGCAATTGGCAAGTGGAGATTTCAAGCGCTTTAAGGTCAATG
GCAGAAAAGGAAATATCTTCGTTTCAAACTAGACAGAATCATTCCCACAACTGCGTGGTGATGT
GTTTCGTTCAACTCACAGAGTTTAACCTTTCTTTTCATAGAGCAGTTAGGAAACAGTCTGTTTTTT

>D5Z2 Acc No. M26919.1 (SEQ ID NO: 118)

GAATTCAGTAGCTTCCTTGTGTTGTGAACATTCAACTCACAGAGTTGAACGTTCCCTTAGACAG
15 AGCAGATTTGAACACTCTTTTTGTGCAATTGGCAAGTGGAGATTTCAAGCGCTTTAAGGTCAATGG
CAGAAAAGGAAATATCTTCGTTTCAAACTAGACAGAATCATTCCCACAACTGCGTTGTGATGTG
TTCATTCAACTCACACAGTTTAACCTTTCTTTTCATAGAGCAGTTAGGAAACAGTCTGTTTGTAAG
TTCTCTAAGTGGATATTCTGACATCTTGTGGC/CTTCGTTGGAAACGGGA/TTTCTTCATATTCTG
CTAGACAGAAGAATTC

20 >D6Z1 Acc No GJ211907.1 (SEQ ID NO: 119)

AGAATTCTCAAACTACTATGTGATGTTTGCATGCAAGTCACAGAGTGCAACATTCTCTTGATA
GAGCAGTTGGGAAACACTCCTTTTGTAGAATTTGCAATGGGATATTTGGACTTCTTTGAGGC*CTT
CATTGGAAACGGGA*TTTCTTCGTATGAATCTAGACAGAAGAATTCTCAGAAACTTCCTTGTGATG
TGTCATTCAACTCAGCGAGTGGCACTTTCTTTGGATACAGCAGTTTGAACACTGTTTTTGTA
25 GTATTTCCAAGCGGATATTTAGAGCGCCTTGAAGCCTATGCTAGAAATGGAATATCTCCCCATAA
AACCAAGACAGAAGCAATCTCAGAACTAATGTGTGATGGCTGCATTCCACACACACGGTGGACCA
TTTCTCTTGATAGAGCAGTTTGAACACTCTTTCTGTAGAATCTGCAAGTGGATAATTGGACCTC
CTAGAGGC/CTTCGTTGGAAACGGGA/TTTCTTCATCTAAACCTACAGAGAAGAATTCTCAGTAAC
TTCTTCGGATGTGTGCATTGCACTCACAGAATGGAACATTCCCTTTATAGAGCAGTTTGTAGACAC
30 CGTTTTTGTAGAATTACCAAGTGGATATTTAGAGCACTTGGAAGTCTCTGCTAGAAAAGGAAACAT
CTTCATGTAAAAAGTAGATAGAATCGTTCTCAGAAAGTGCTTAGTGACGTGTGTGTTCAACTCACA
GAGTTTAACGTTTCTTTTGATAGAGCGTTTCTGAAACACCTTCTTGTAGTAGCTGCAAGTGGATA
TTTGACCTATTTGAGGC*CTTCTTTGTAACGGGA*TTTCTTCATGTAACCTAGATTGAAGAAT
TCTCAGAAACTCCTTTGTGATGTGTGCATTCAATTCAAAGAGTGAAACCTCCCTTTTACAGAGCA
35 GTTTTGAAACACTGTTTTTGTAGGATTTCCAAGGGGATATTTATATCGCATTGAGCCTACGGCAGA
AAAAGAAACATCTTCCTATAAAAACTAGACAGAATGATTCTCAGAAATCTGCTTTGCGATGTGTGCA
TTCAACCCACAGAGTAAACTTTTCTTTTGATAGAGCAGTTTGAACACTCTTTTTGTAGTATTT
GCATGTGTATATTTAGAGCGCGTTGAAGCCACAGTAGAAAAGGAGATAACTTCACCAAAAACCTA
GACGGAAGCAATCTCAGAACTACTTTGTGATGTGTACATTCAACTCACAGAGTGGAACCTTTCCCC

FIG. 9B Continued

TTTACAGAGCAGTGTTGAAACACTCTTTTTGTAGAAACTGCAGGTGGATATTTGGACCTCTTTGAG
GC/CTTCGTTGGAAACGGGA/TTTCTTCTATAACCCTAGACAGAAGAATTTTCAGAAACCTCATT
GTGAGTGTGCGTTCACCTCACAGAGTGGAGTCTTCCGTTTGATAGAGAAGTTTTGAAACCCTGTTT
TTGTAGTATCTCCAAGTGGATATTTAGACCCTTTGAAGCCTATGATAGAAAAGGAAACATCTTCA
5 TGGAAAACATAGATAGAATCATTCTCAGAAACAACCTTTGTGATGTGTGCGTTGAACTCACCGTCTT
TAACCTTTCTTTTGGTAGAGAAGTTTTGAAACA

>D7Z1 Acc No. AC142529.3 (SEQ ID NO: 120)

GGGAAATATCTTCAAATAAAAACCAGACAGAATCATTCTCAGAAAATTCTTTGTGATGTGTGCGTT
CAACTCACATAGTTTAACTTTCTTTTCATAGAGCAGTTTGGAAACACTCTGTTTGTAAGTCTGC
10 AAGTGGATATATGGACCGCATTGAGGC/CTTCGTTGGAAACGGGA/TTTCTTCATTTTCATGCTAGA
CAGAAGAATTCTCAGTAACTTCTTTGTGCTGTGTGATTCAACTCACAGAGTGGAACTCCCTTTG
CACAGAGCAGATTTGAAACACTCTTTTTGTGGAATTTGCAAGTGGAGATTTCAAGCGATTTGATGC
CAACAGTAGAAA

>D7Z2 Acc No. M16037.1 (SEQ ID NO: 121)

15 AGCTTCCTTTTGATAGAGCAGTTTTGAAACTCCCTTTTTGCACAATCTGCAGGAGGATATTTGGAG
CTCTTTGAGTGCTACATTGGAAACGGGAATATCGTCACCTAAAACTAGAAAGAAGCATTCTCTGA
AACCCTTTGTGATGTGTGCATTCATCTCACAGAGTTGAACCTTCCTTTTGATAGAGCAGTTTTGA
AACCCTCTTTTGTACAATCTGCAAGTGGATATTTGGAGCAAATTGAAGCCTTCTTTGGAAATGGG
AATATCTTAATCTAAAAATTAGGCAGAGCATTCTCAGAAACTACTTTGTGATGTGTGCATTCAACT
20 CACAGAATTGAACCTTCCTTTTGATACAGCAGTTTTGAAACACTCTTTGTTTAGAATCTGCAAGTG
GATATTTGGAGCACATTTATGCCTGTGGTAGAAAAGGAAATATCTTCACATAAAAACTAGACAGAA
GCATTCTCAGAAACGACTTTGTGATGTGTGCATTCTACTCCCATAGTTGAAAATTTCTTTTGATAG
AGCAGTCTGGAAACACTCTGTTTCTAAAAATCTGCAAATGGACATTTGGTGCGCTTTGAACGTTATG
ATGGAAGGGAATATCTTCGCATTAAAACTAGACAGAAGCATTCTCAGAAACTTCTTTGTGATGT
25 GTGCATTCAACTCCCAGGTTGAACCTTTCTTTTGTTAGAGCAGTTATGAAACACTCCTTTTGTA
ATCTGCAGGCGGATATTTAGTACTCTTTGAAGCATTCTTTGGAAACGAGAATATCTTCACCTAAAA
CCTAGACAGAAGCATTCTCAGAAACATCTTTATGATGTGTCCATTCATCTCACAGAGTTGATAGAA
CATTTTGATAGAGCAGTTTTGAAACACTCCTTTTAAAGAATCTGCCAGTTCATATGTGCAGTGCTT
TGAGGCTTATGGTAGAAAAGGAAATATCTTCATATAAAAACTAGACAGAAGCATTCTCAGAAACGA
30 CTTTGTGATGTGTGCATTCTACTCGCAAAGTTGAACTTTCTTTTGATAGAGCAGTTTTGAAACCT
TCTTTCCAAAGAATCTTCAAGTGGGCATTTTCGAGGGCTTTAGGACCATTGCGGATAAGGAAATATC
TTCCCATAAGAAGTAGACAGAACTATAATCAGAAACTTCATTTTGATGTGTACATTCAACTCACAA
AGCAGACCCTTACTTTTGATAGAGAAGTTTTGAAACACTCTTTTTGTAGAATCTGCAAATTGATGT
TTGGAGCGCTTTCAGGCCTCTGGTAGAAAAGGAAATACCTTCACATAAAAACTAGACAGAAGCATT
35 CTCAGAAACGACTTTGTGATGTGTGTATTCTACTCCCATAGAAGAACATTTCTTTTGATAGAGCCG
CCTGGAAACAATCTTCTTGTAGAATCTGCAAGTGGACATTTGGAGCTTTTTGAAGGCTGTGGTTGA
AAAGGTAATATCTTCACCTAAAACTAAATGGAAGCATTCTCCCAAACGTTTTTGTGATGTGTGCG
TTCAACTCACAGAGCTGAACCTTCCTTTTCTTAGACCAGTTTTGAATCACTCTTTTTGTAGAATCC
GCATTTAGATATTTGGAGCGCTTTGAAGACTTCATTGGAATCGCGAATACCTTCACATAAAAACTA

FIG. 9B Continued

GACAGAACCATTCTCAGAAACCCCTTTGAGATGTGTGCATTCAACTCACAGAGCTGAACCTTTCTT
TTGATAGTGCAGTTTTGAAACATTCTTTTTTAAAAATCTGCAGTTGGACATTTGGAGCTCTTTTAG
GCTATCGGTTGAAAAGGAAGTATCTTCACATTAACAAGACAGAAGCATTCTCAGAAACTCCTTT
ATGATGTCTGCATTCAACTCACAGAGTTGAACCTTCCTTTTGATAGAGCAGTTTTGAAACACTCTT
5 TCTGTAGAATCTGGAGGCGGATATTAGGGTGCTTTGAAGCCTTCTTGGGAAACAGGATTATCTTCA
CATAAAAATTAGACAGAAGCATTCTCAGAAACTTCTTTGTGATGTGTGCATTCAACTCACAGCGTT
GAAACTTCCTTTTGCTAGAGCAGTTTTGAAACCCTCTTTTGAAGAATCTGAAAGTGCATAATTGC
AGCACTTTGAGGCTTAAGGTAGAAAAGGAAATATCTTCATATAAAAACTAGACAGAAGCATTCTCA
GAAACTACTTTGTGATGTATACATTCTACTCACATAGTTGAAATTTCTTCTGATACTGCAGTTTT
10 GAAACCGTCTTTTTGAGGAATCTTCCAGTGGGCATTTTGAGGGCTTTGGGGGCTATTGTGGATAAG
GAAATATCTTCACATGAAAAGTAGACAGAAGTGTCTCAGAAACTTCATTTTGATGGGTGCATTCA
ACTACAAGGTAAAACCTTACTTTATTGAGCAGTTTTGAAACAGTCTTTTTGTAGACTCTGCAGGTG
GATATTTGGAACGCTTTGAAGC/CTTCGTTGGAAACGGGA/ATATCTTCCCCTTGAAACTAGACAG
AAGCATTCTCAGAAACTTCTTTGTGATGTGGGCATTGAACTCACGGAGCTGAACCTTCCTTTGGAT
15 TGAGCAGTTTTGAAAACTCTTCTTTATAATCTGCAGGTGGATATTTGGAGTGCTTTGAAGCCTT
CTTTGGAAACGGGAGTATCGTCACATAAAAATAGACAGAAGTATTCCCAGAAACTTCTTTGTGATT
TGTGCATTCAACTCACAGAGTTGA

>D8Z2 Acc No. M64779.1 (SEQ ID NO: 122)

AAGCTTTGAAGA{TTTCGTTGGAAACGGGA}ATATCTTCAAATAAAAATCTAGCCAGAAGCATTCTA
20 AGAAACATCTTAGGGATGTTTACATTCAAGTCACAGAGTTGAACATTCCCTTTCACAGAGCAGGTT
TGAAACAATCTTCTCGTACTATCTGGCAGTGGACATTTTGAGCTCCTTGGGGCCTATGCTGAAAAA
GGAAATATCTTCCGACAAAAACTAGACAGAAGCATTTCGAGAATCACGTTTGTGATGTGTGCACTC
AACTGTCAGAATTGAACCTTGGTTTGGACAGAGCACTTTTGAAACACTCTTTTTGTAGAATCTGCA
GGTGGATATTTGGCTAGCTTTGAGGA*TTTCGTTGGAAACGGTA*ATGTCTTCAAAGAAAATCTAG
25 ACAGAAGCATTCTCAGAAACAGCGTCGTGATGTTTGCAATCAAGTCACAGAGTTGAACCTTCCGTT
TCATAGAGCAGGTTGGAAACACTCTTTTTGTAGTATCTGGAAAGTGGACATTTGGAGGGCTTTGTAG
CCTATCTGAAAAAGGAAATATCTTCCCATGAATGCGAGATAGAAGTAATCTCAGAAACATGTTTA
TGCTGTATCTACTCAACTAACTGTGCTGAACATTTCTATTGATAGAGCAGTTTTGAGACACTCTTC
TTTTGGAATCTGCAAGTGGATATTTGGATAGATTTGAGGA{TTTCGTTGGAAACGGGA}TTATATA
30 TAAAAAGTAGACAGCAGCATTCTCAGAAACTTCTTTGTGATGTTTGCATCCAGCTCTCAGAGTTGA
ACATTCCCTTTCATAGAGTAGGTTTGAAACCCTCTTTTTATAGTGTCTGGAAGCGGGCATTTGGAG
CGCTTTCAGGCCTATGCTTAAAAAGGAAATATCTACCTACAGAAACTAGACAGAAGCATTCTGAGA
ATCACGTTTGTGATGTGGGTACTCAACTAACAGTGTGATCCATTCTTTTGATACAGCAGTTTTGA
ACCACACTTTTTGTAGAATCTGCAAGAGGATATTTGGATAGCTGTGAGGA{TTTCGTTGGAAACGG
35 GA}ATGTCTTCAAAGAAAATCTAGACAGAAGCATTCTCAGAAACACCTTCGTGATGTTTGCAATCA
AGTCACAGAGTTGAACCTTCCGTTTCATAGAGCAGGTTGGAAACACTCTTATTGTAGTATCTGGAA
GTGGACATTTGGAGCGCTTTCAGGCCTATGGTGAAAAAGGAAATATCTTCCATAAAAACGACATA
GAAGCTATCTCAGGAACCTGTTTATGATGCATCTAATCAACTAACAGTGTGAGGTTTGTACTGA
CAGAGCAGTTTGAAACACTCTTTTTTTGGAATCTGCAAGTGGATATTTGGATCGCTTTGAGGA{TT
40 TCGTTGGAAACGGGA}TGCAATATAAACGTACACAGCAGCATACTCAGAAAATACTTTGCCATAT

FIG. 9B Continued

TTCCATTCAAGTCACAGAGTGGAACATTGGGATTCATAGAGCAGGTTTGAAACACTCTTTTTGGAG
TATCTGGAAGTGGACATTTGGAGCGCTTTCTGAACTATGGTGAAAAAGGAAATATCTTCCAATGAA
AACAAAGACAGAAGCATTCTGAGAACTTATTTGTGATGTGTGTCCTCAACAAACGGACTTGAACCT
TTCGTTTCATGCAGTACTTCTGGAACACTCTTTTTGAAGATTCTGCATGCGGATATTTGGATAGCT
5 TTGAGGA*TTTCGTTGGAAACGGGC*TTACATGTAAAAATTAGACAGCAGCATTCTCAGAACTTC
TTTGTGGTGTCTGCATTCAAGTCACAGAATTGAACTTCCCCCTCACATAGAGCAGTTGTGCAGCACT
CTATTTGTAGTATCTGGAAGTGGACATTTGGAGGGCTTTGTAGCCTATCTGGAAAAAGGAAATATC
TTCCCATGAATGCGAGATAGAAGTAATCTCAGAAACATGTTTATGCTGTATCTACTCAACTAACTG
TGCTGAACATTTCTATTGATAGAGCAGTTTGGAGACACTCTTCTTTTGGAACTCTGCAAGTGGATAT
10 TTGGATAGATTTGAGGA { TTTCGTTGGAAACGGGA } TTATATATCAAAAGTAGACAGCAGCATTCT
CAGAACTTCTTTGTGATGTTTGCATCCAGCTCTCAGAGTTGAACATTCCCTTTCATAGAGTAGGT
TTGAATCCCTCTTTTTATAGTGTCTGGAAGCGGGCATTGGAGCGCTTTCAGGCCTATGCTGAAAA
AGGAAATATCTACCTATAGAACTAGACAGAAGCATTCTGAGAATCACGTTTGTGATGTGGGTACT
CAACTAACAGTGTTGATCCATTCTTTTGATACAGCAGTTTGAACCACACTTTTTGTAGAATCTGC
15 AAGTGGATATTTGGATAGCTGTGAGGA { TTTCGTTGGAAACGGGA } ATGTCTTCATAGAAAATTTA
GACAGAAGCATTCTCAGAACCTTGATTGTGATGTGTGTTCTCCACTAACAGAGTTGAACCTTTCTT
TTGACAGAACTGTTCTGAAACATTCTTTTTATAGAATCTGGAAGTCGATATTTGGAAAGCTT

>D9Z4 Acc No. M64320.1 (SEQ ID NO: 123)

CCTATGATGAAAAAGGTAATATCTTCCCATAAAACTAGCAGAAGCATTCTCAGAACTGTTTGTGA
20 TGTGTGTATTCAACTAACAGAGATGAACCTTTCTTTTTACAGAGCAGTTTGAACACTCTTTTGTGG
AATCTGAAAGTGGATATTTGGATAGCTTTGCGGA { TTTCGTTGGAAACGGGA } TTACATATAAAAT
CTAGGGAGAAGCATTCTCAGGAACCTTTTGTGATGTTTGCATTCAAGTCACAGAAGTGAACATTC
CCTTTCATAGAGCAGGTTTGAAACACTCTTCTGTAGTATCTGCAAGCGGACGTTTTAAGCGCTTT
CAGG

25 >D10Z1 Acc No. X63622.1 (SEQ ID NO: 124)

TGCAGGTGGATAATTGGCCCTCTTTGAGCC/CTTCGTTGGAAACGGGA/TTTCCTCATATAATGCT
AGACAGAAGAATTCTCAGTAACTTCTTTGTGTTGTTTGTATTCAACTCAGAGATTTGAACCTTCCT
TTAGAGAGAGCAGATTTGAAACACTGTGTTTTTGGAAATTTGCAAGTGCAGATTTCAAGCGCTTCTA
TGCCTATGGCAGAAAAGGAAATATCTTCGTATAAAAACTACACAGAATCATTCTCAACAACTACTT
30 TGTGATGTGTGCGTTCAACTCACAGAGTTAAACCTTCCTTTTCATTGAGCAGTTTGGAAACACTCT
GTTTGTAAAGCCTGCAAGTGCTTTTTTGGACTTCATTGAGGC/CTTCGTTGGAAACGGGA/TTTCA
TCATATAATGCTAGAGAGAAGAATTCTCAGTCACTTCTTTGTGTTGTGTGTTATTCAAGTCACAGAG
TTGAACCTTCCTTTAGACAGAGCAGTTTGGAAAAATCTTTCTGTGGAATTTGCAAGTAGAGATTT
CAAGCGATTTGAGGCTAATCTTTGAAATTGAAATATCTTCGTGTAAAAACGACACAGAAGCATTCT
35 CAGAACTGCTTTGTCATCTGTGCGTTCAAGTTCACAGAGTTTCACCTTTCTCTTCATAGAGCAGTT
TGGAAAGACTGTCTGTAAAGTCTGCAAGTGATTAGTTAGACCCCTTTGAGGC*CTTCGTTGGAAAGC
GGGA*TTTCTCATTTACTGCTAGACAGAAGAATTCTCAGTAAATCCTTTGTGTTGTGTGTTATTCAA
CTCACAGAGTGGAACTTCCTTTATTTCAGAGCAGTTTGGAAAAACACTTTTTGTGGAATTTGCAAG
TGGAGATTTCAAGCGATTTGACGCCAATCTTAGACATGGAAATATCTTCATATTTAAAGT

FIG. 9B Continued

>D11Z1 Acc No. M21452.1 (SEQ ID NO: 125)

CTAGAATCCCCAAGTGGACATTTGGAGGGCTTTGAGGCCTGTGGTGGAAAAGGAAATATCTTCACA
TAAAACCTAGATAGAAGCATTCTCAGAACTACTTTGGGATGATTGCATTAGACTCACAGACTTGA
ACCTTCCAATGGATAGAGCAGTTTGTAAACACTCTTTTTGAGAATCTGTGATTGCTGATTTGGACT
5 GCATTGAGGCCTACGGTACTAATGGAAATAACTTCATCTAAAATCCAAACGGAAGCATTACAGAA
AATTCTTTGTGATGATTGGATTGAACTAAGAGAGCTGAACATTCTTTAGATGGCGCAGTTTCCAA
ACACACTTTCTGTAGAATCTGCACGTGGATATTTGGACTTCTCTGAGGA { TTTCGTTGGAAATGGG
A } TAAACTTCCCAGAACTACACGGAAGCATTCTCAGAACTTCTTTGTACGTTTGCATTCAACTC
ACAGAGTTGAACCTTCTTTTCATAGTTCAGCTTTCAAACACTCTATTTGTAGAATCTGCAAGTGGAA
10 TATTTGGACACTTTGTGGCCTTCTTCGAAACGGGTATATCTTCACATCAAACCTAGACAGAAAACA
TTCTCAGAATGTTTCCTGTGATGACTGCATTCAACTCACAGAGTTGAACAATCCTGCTGATGGAGC
AGTTTGGAAACGCTCTTTCTTTGGATTCTGCAAGTGGATATGTGGACCTCTTTGAAGA * ITTCCTT
GGAAACGGGT * TCCTCTTCACATAAAAAGTAAACAGAAGCATTCTCAGAACTGCTTTGTGATGTT
TGTGTTCCATTTCAAGAATTGAAATTTCTGTTGACAGAGCAGCTCTGAAACCCTCTTTTT

15 >D12Z3 Acc No. M28221.1 (SEQ ID NO: 126)

AAGCTTTCTTTTCATAGAGTAGTTTGGAAACACTCTGTCTGTAAAGTCTGCAAGCAGATATTTGGA
CCTCTTTGGGGC / CTTCGTTGGAAACGGGA / TTTCTTCATAGAACGCTAGAAAGAAGAATACTGAG
TAAGTTCTTTGTGTTGCCTCTATTCAACTCACAGAGGTGAACTGTCCTTTAGACAGAGCAGATGTG
AAACCCTCTTTTTGTGATATTTGCAGGTGGAGATTTCAAGCGCTTTTAGGCCAAATGTAGAAAAGG
20 AAATATCTTCGTATAAAAACCTAGACAGAATCATTCTCAGAACTACTTTGTGATGTGTGCGTTCAA
TTCACAGAGATAACCTTTCTTTTGATGAAGGAGTTTGGAGACACTGTGTTTGTAAAGTCTGCAAGT
GGATATTTGGACCTCTTTGAGGC / CTTCGTTGGAAACGGGA / TTTCTTCCTGTAATGTTTCGACAGA
AGAATTCTCAGTAACTTATTTGTGGTGTGTGATTCAACTCACAGAGTTGAACCTTCTTTTAGACA
GAGCAGATTTGAAACACCCTATTTGTGCAGTTTCCAGTTGGAGATTTCAATCACTTTGAGACCAAA
25 TGTACAAAAGGAAACATCTTCGTATAAAAACCTAGACAGAATCATTCTCAGAACTACTTTGTGATG
TGTGCGTTCAACTCAAGGTGTTTAAGCTT

>D13Z1 Acc No. D29750.1 (SEQ ID NO: 127)

AATTCAAATAAAAAGGTAGACAGCAGCATTCTCAGAAATTTCTTTCTGATGTCTGCATTCAACTCAT
AGAGTTGAAGATTGCCTTTTCATAGAGCAGGTTTGAAACACTCTTTCTGGAGTATCTGGATGTGGAC
30 ATTTGGAGCGCTTTGATGCCTACGGTGGAAAAGTAAATATCTTCCATAAAAACGAGACAGAAGGAT
TCTCAGAAACAAGTTTGTGATGTGTGTACTCAGCTAACAGAGTGGAAACCTTTCTTTTACAGAGCA
GCTTTGAACTCTATTTTGTGGATTCTGCAAATTGATATTTAGATTGCTTTAACGA * TATCGTTG
GAAAAGGGA * ATATCGTCATACAAAATCTAGACAGAAGCATTCTCACAACTTCTTTGTGATGTGT
GTCCTCAACTAACAGAGTTGAACCTTTCTTTTGTGATGCAGCAGTTTGGAAACACTCTTTTGTAGAA
35 ACTGTAAGTGGATATTTGGATAGCTCTAACGA { TTTCGTTGGAAACGGGA } ATATCATCATCTAAA
ATCTAGACAGAAGCACTATTAGAACTACTTGGTGTATCTGCATTCAAGTCACAGAGTTGAACAT
TCCCTTACTTTGAGCACGTTTGAAACACTCTTTTGGAGAATCTGGAAGTGGACATTTGGAGCGCT
TTGACTGCCTTTGTTGAAAAGGAAACGTCTTCCAATAAAAAGCCAGACAGAAGCATTCTCAGAACT
TGTTTCGTGATGTGTGTACTCAACTAAAAGAGTTGAACCTTTCTATTGATAGAGCAGTTTGAACA

FIG. 9B Continued

CTCTTTTTGTGGATTCTGCAAGTGGATATTTGGATTGCTTTGAGGA [TTTCGTTGGAAGCGGGA] A
TTCGTATAAAAACTAGACAGCAGCATTCCCAGAAATTTCTTTTCGGATATTTCCATTCAACTCATAG
AGATGAACATGGCCTTTTCATAGAGCAGGTTTGAAACACTCTTTTTGTAGTTTGTGGAAGTGGACAT
TTCGATCGCCTTGACGCCTACGGTGAAAAAGGAAATATCTTCCCATAAAAAATAGACAGAAGCATT
5 TCAGAAACTTGTGGTGATATGTGTCTCAACTAACAGAGTTGAACTTTGCCATTGATAGAGAGCAG
TTTTGAAACACTCTTTTTGTGGAATCTGCAAGTGGATATTTGGATAGCTTGGAGGA [TTTCGTTGG
AAGCGGGA] ATTCAAATAAAAGGTAGACAGCAGCATTCTCAGAAATTTCTTTCTGATGACTGCATT
CAACTCATAGAGTTGAACATTCCCTTTTCATAGAGCAGGTTTGAAACACTCTTTCTGGAGTATCTGG
ATGTGGACATTTGGAGCGCTTTGATGCCTATGGTGAAAAAGTAAATATCTTCCCATAAAAAACGAGA
10 CAGAAGGATTCTGAGAAACAAGTTTGTGATGTGTGTAAGTGGATATTTGGATAGCTCTAATGA
GATGCAGCAGTTTGGAACACTCTTTTTGTAGAACTGTAAGTGGATATTTGGATAGCTCTAATGA
{TTTCGTTGGAACCGGGA} ATATCATCATCTAAAATCTAGACAGAAGCCCTCTCAGAACTACTTT
GTGATATCTGCATTCAAGTCACAGAGTTGAACATTCGCTTTCTTAGAGCACGTTGGAAACACTCTT
TTTGTAGTGTCTGGAAGTGGACATTTGGAGCGCTTTGATGCCTTTGGTGAAAAAGGGAATGTCTTC
15 CCATAAAAACTAGACAGAAGCATTCTCAGAACTTGTTTTTGATGTGTGTACCCAGCCAAAGGAGT
TGAACATTTCTATTGATAGAGCAGTTTTGAAACACTCTTTTTGTGGAAAATGCAGGTGGATATTTG
GATAGCTTGGAGGA [TTTCGTTGGAAGCGGGA]

>D13Z2 Acc No. GJ211955.2 (SEQ ID NO: 128)

AGCATTCTGAGAAATTACTTTGTGATGTGTGCATTTCATCACAAAGAGTTGAACCTTTCTTTTGGTT
20 GAGCAGTTTTGAAGCACTCTTTTTGTAGAATCTGTAAGTGGATATTTGGAGTTCTTTGAGGCTTAT
GGTGGTAAAGGAAATATCTTCACACAAAACTACACAGAACCATTCTGAAATACCTCTTTGTGATG
CTTGCAATTCATCTCACATAGTTGAACCATTCTTTTTATTGAGCAGTTTTGAAACAATCTCCTTGTA
GAATGTGCAAGTGGATATTTGGAATGCTTTGATGAGTATGGTGGAATGAAAAATCTTCACATAA
AACTAGACAGAATTACTCTGAGAACTTCTTTGTGATGTGCACATTCATCTCACAAATTTGAAAA
25 TTTCTTTTGATTGAGCAGTTTTGAAACGCTCTTTTTCTAGAATCTGCCAGTGGTTATTTGGAGTGC
TTTGAGTCCTATGGTGGAGAAGGAAATATCCTCACATAAAAACTAGAGAGAAGCATTCTGAGAAAC
TTCTTTCTGATGTGTGCATACATCTCACGGAGTTGAACTTTCTATTGATTTAGCATTTTTTTATAC
ACTTTTTGTAGGATCTGCAGTTGCTATTTGGAGCCCTTTGGGGCCAATGGTGGAAAAGTACTATCT
TCTCATAAAAACTAGACAGAAGCATTCTGAGAACTTCTTTGGTGTGATGTGTGCATTCATCTCACAGT
30 AGTTGAACCTTTCTTTTGATTGAGCAGTTTTGAAACGCTCTTTTCGTTGAATCTGCAAGTGCATAT
TTAGAGTGCTTTGAGGCACGTGGTGGAAAAGGAAATATCTTCACATAAACTAGACAGAAGCATT
CTGAGAAATGTCTTTGTGATGTGTCCATTCACTTCACAGAGTTGAACTTTCTTTTCATTGAGCAG
TTTTGAAACACTCTTTTTATAGAATATGCAAGTGGATATTTGGAGCGTTTTGGAGAGAATGGTGG
AATGGAAATATCTTCATATAAAAACTACGGAGAAGCATTTCGGAGAAGCGCTTTGTTATGTGTGAC
35 TTCAGCTCACACAGTTGAACCTTTCTTTTGATTGAGCATTTTTGATTCCCTCTTTTTGTAGAATCT
GCAAGTGGATATTTGGAGAGCTTTAGGGCCTACGGTGGAAGGAAATATCTTCACATAAAAACTA
CACAAAAGCATTCTGAGAACTTCTTTCTGATGTGTGCATACAACTCACAGAGTCGAACTTTCTT
TTGATTGTGCAGTTTTGAAACACTTCTTTTGTAGAATCTGCAAGTGGATATTCAGAGGGCTTTGTG
GAGTATAGTGGAAAAGGAAATAACTTTGGATAAAAGCTAGACAGCAGAATTCTGAGAACTTCTTT
40 GTGATGTGTGCATTCAACGTACAGAGTTGAACCTTTCTTTAGATTTGGCAGTTTTGAAACACTACT

FIG. 9B Continued

TTTGTAAATATCTGCAAGTGGATATTTGGTGACCATTGCAGCCTATGGTGGAAAGGCCAAATATCTTC
ACATAAAACTAGACCAAGGCATTCTGAGAATCTTCTTTGTGATGTGTGCATTCTTCTCACACAGT
TCAACTTTTCTTTTGATTGAGCAGTTTGGAAACAGTATTTTCTACAATCTGCAAAGGGATACTTC
TTAGCCGATTTAGGCCTATGGTGAATTAGGAAATATCTTCACATAAAAAATAAACAGAACTTTCT
5 GAGAACTTCTTTGGGATGTGTGTTTTTCATCTCA

>D13Z3 Acc No. GJ211961.2 (SEQ ID NO: 129)

AACATGATGAGAACTGCTTTGTGATGCGTGCATTTCATCACCAGTAGTTGAGTTTCTCTTTTGATT
GAACAGTTTTTGAACACTCTTTCTGAAGAATCTGAAAGGGATATTTGGAGCGCTTTGCAGCCTATG
GTGAAAAAGGAAATATCTTCACATAAAAGCTAGACAGAAGCATTCTAAGAAAGTGCTTTGTGACGT
10 GTGCATTCATCTCACAGTGTTGAAGCTTTCTTTTGATTGAGCAGTTTGAACACTCTTATTGTAG
AATCTGCAAGTGGATATTTGGAGAGTATGAGGCCACTGGTGGAAAAGCAAATATCTTCACATCAAA
ACTAGACAGAATCATTATAAGTAATCTCTTTGAGATGCGTGCATTCAACTCACAGAGTTGGACATT
TCCTTTGATTGAGCAGTGTGGAAACAGTCTTTTGCAGTATCTGCAAACGGATATTTGGAGCACTT
TCAGGCCTATAGTAGGAAAGGAAATATCTTCACATAAAAACTACACAGAAAATTACTGAGAACTT
15 CTTAATGATGTGTGCATTTCATCTCACAGAGTTGAACTTTCTTTTGATTGAGCAGTTTGGAAACAC
TCTTTTAGTAGAACTGCAAGGGGATATTTGGAGAGTTTTGTGGTCTATGGTAGAAAAGGATATAT
CTTCACATAAAAAATAGAAGCATTCTGAGGAACTTCATGATGTGTGCATTTCGTCTCAAGGAGTTGAA
CTTTTCTTTTGATTGAGCAGCTTTGAATAACTCTTTCTGCAGAATCTGCAAGTTGATATTTGGAGT
GCTTTGTGGCCTATAGTAGAAAAGGAAATATCTTTACATAAAACTAGACAGAAGCATTCTGAGAAA
20 ATTTTTTGTGATGTGTGCATTCAACTCACAGAGTTGAACCTTTCTTTTGATCGAGGAGTTTGGAAAG
CAGTCTTTTTGTAAATATCTACAAATGGATATTTGCATCACTTTACTTCATGGAATGGAAAAGGAAA
CATCTTAACATAAAAACTAGACAAAAGCATTCTGAGAACTTCTTTGTGACGTGTGCATTCAACTC
ATGGAGTTCAACCTTTCTTTTGATTCAGCAGTTTGGAAACAGTCTTTTACAGTATCTGCAAATGG
CTATTTGGAGAGCATTGACGCCTATGGTGGAAAAGGAAATCTCTTCATAAAAACTAGACAGCAG
25 CATTCTGAGAACTTATTTGTGATCTGTGCATTTCATCTCACAGAGTTGAACCTTTCTTTTGATTCA
GCAGTTTTGAACTGTCGTTTTGTAGAATCTGCAAAGGAATATTTGTGAGCCCATTGAGGCTTCTG
GGGTGATAGGAAATATCTTCACATTAAAACTAGACAGATACTTTCCGAGAAACTATTTTGTGATGT
GTGACTTCTACTCACAGGGTTGAACTTTCTCTTGATTGAGCAGTTTGGAAACAGTCTTTTTGTAG
AATCTGCAAATTGATATTTGGAGTGCTTTTGGCCTACGTTGAAAAACGAAATATCTTCCCATAAAA
30 AGTAGGCAGAAGTTTTGGAGAACTTATTTTGATGTGTGCATTTCATCTCACACAGTTGAAATTTTC
TTTTGATTGAGCAGTGTGGATACACTCGTTTTGTAGAGTCTGCAAGTGGATATTTGGAGCACTTTG
TGGCCTATAGTAGAAAAGGAAATATCTTCACATAAAAACTAGATAGAAGAATTCTGAGAACTTCC
TTTGAATGGGCGCATTTCATCTCACACTGTTGAAC

>D13Z6 Acc No. GJ211965.2 (SEQ ID NO: 130)

35 TGCATTCTCAGAAAGTTCTTTGTGATGTGTGCATTCAAATCACAGATTTGAACATACCTTGTGATA
GAGCAGTTTTTGAACACTCGTTTCGTAGAATCTGCAACTGGGTATTTGGACTTCTTTGAGGCCTTC
GTCGGAACCGGAATATCTTCACATAAGAACTAGACAGAAGAATTCTGGGGAATTTCTTTGTGATG
TGTGCATTCAACTCACAGAGTTGAACCTTTCTGTTGATAGAGCAGTTTGGAAACACTCTTTTCGCA
AAATCTGCAGAGTGGATATTTGTACTGCTTAGAGGC/CTTCGTTGGAAACGGGA/ATATCTCCACA

TAAAACTAGACAGAAGCATTCTCAGAACTTCTTTGTGATCTGCACATTCAACACAAAGAGTAGA
ATCTTCCTTTTGTAGAGAGCAGTTTTTAAACACTCTTTTTGTAGAATCTGCAAGTGGACATTTGGAA
AGCTTTGAGGCCTGTGGTGGAAAAGGAAATACCTTCACATAAAAACCAGACGGAAGCATTCTCAGA
AACTTCTTTGTATTGTTTGCATTCAACCCACTGAGTTGAACACACCTTTTTCACAGAGCAGTTTTGA
5 AACACTCTTTTTGTAGAATCTGCAAGTGGATATATGGAGTGCTTTGAGGCCTTCTTTGTAAACGGG
AATATCTTCACATAAAAACCTAGAGAGAAGCATTCTCAGAGCCTTCTTTGTGATGTGTGCATTCAAC
TCACAGAGCTGAACCTTTCTTTTGTAGAGGTGTTTGAAGCACTGTTTTTTTAGAATCTGCAAGTG
GATATATTGAGTGCTTTGAGGCCTTCTTTGTAAACGGGAATATCTTCACATAAAAACCTAGAGAGAA
GCATTCTCAGAGCCTTCTTTGTGATGTGTGCATTCAACTCACAGAGCTGAACCTTTCTTTTCATAG
10 AGCTGTTTGAAGCACTGTTTTTTTAGAATCCGCATGTGGAAATTTTCAGAGCTTCGAGGCCTGTG
GTGGAGAAGGAAATATCTTCACATAAAAACCTAGACAGAAGCATTCTCAGAAACTTGTGTGTGACGT
TTGCATTCAACTCACAGAGTTGAACATACCTTTTCATAGAGCAGTTTTGAAACACTCTTTTCGTAG
GATCTGCAAGTGGATATTTGGAAGTCTTTGAGGCCTTCGTTGGAAAGAGGAATATCTTCACATAAA
AACTAGACGGAAGCATTCTCAGAAACTTCTTTGTGATGTGTGAATTCAACTCACAGAGTTGAAGCT
15 TCCTATTGATAGAGCAGTTTTGAAAACCGTTTTTGTAGAATCTGCCAGTGGATATTTGGAGAGCT
TTGAGGCCTACGGTGGAAAAGGAAATATCTTCACATAAAAACCAGACACAAAGATTCTCAGAAACT
TCTTTGTGACGCTTGCACTCAACTCACAGAGTTGAACACACCTTTTCATAGAGCAGTTTTGAAGCA
CTCTTTTCGTAGAATCTGCAAGTGTATATTTGGAATGCTTTGAGGCCTTCATTGTAAACGAGAATA
TCTTCACATAAAAACGAGACAGAAGAATTCTCAGCAACTACTTTGTGATGATTGCATTCAACTCAC
20 TGTGTTAATCTTTATTTTGTAGAGGCAGTTTTGAAACACTGTTTTTGTAGCATCTGCAAGTGGTCA
TTTGGAGAGCTTTGAGGCCTATGGTGGAAAAGGAAATATCTTCACATAAAAACAGGACAGAAGCAT
TTTCAGAATCTCCGCTGTGATGTTTGCATTGAACTC

>D13Z7 Acc No. GJ211967.2 (SEQ ID NO: 131)

AGCATTATAGAAACTTCTTTGTGATGTATGCATTCAACTCACAGAGTTGAAACTATCTTATTATT
25 GAGCAGTTTTGAATCTCTCTTTTTTGCAGAATCTGCAACTGGATATTTGGAGCGCTTTGAGGCCTAC
CGTGGAAAAGCAAATATCTTCAGATAAAAGCTACACAGAAGCTTTCTGAGAAACTTTTTTGCATG
TGTGCATTCAACTCACAGAGTTGAAACTTTCTTTTGAATTGAGCAGATTTGAAACACTCTTTTTGTA
GAAACTGTAAGTTGATATTTGGAGCCCTTTGAGGCCTATTGTGGAAAAGGAAATATCTTCACGTAA
AACTACATAGAACCATTCTGAGATACTTCTTTTTGATGTTTGCATTCTCTCACAGTGTGAAAG
30 TTTCTTTTGAATTGAGCAGTTTTTGAACACTCTTTTTGTAGAATCTGCAAGTGAATAATTGGAGCCC
TTTGAGGGCTATGGTAGAAAAGGAAATATCTTCAAATAAGAACTACAAAGAAACATTCTCAGAAAC
TTATTTGTGATGTGAGCATTCAACTCACAGACCTGAACATATCTTTGATTTAGCACTTTTGAATT
TCTCTTTTTGTAGAATTTGCAAGTGGATATTTGGAGCGCTGTGAGACCTACTGTGGGAAATGAAAT
ATGTTTCACATAAAAACCTACTCAGAACCATTCTGAGAAACTTCTTTGTGTCGTGTGCATTGCACTCA
35 CAGAGTTGAACATATGTCCTCTTTGAGCAGTTTTGCGTCTCTTTTTGTAGAATGTACAAGTGGGA
TATTTGGAGCCCATTTGTGTCCTATGGTGGAAAAGGAAATATCTTCAGATAAAAATTACACAGAAGA
ATTCTGAGAACTTCTTTGTGATATGTGCATTTATCTCACAGGTTTGAACCTACCGTTTTATTGAG
CAGTTTTGAAACACTGTTTTTGTAGAATCTGCAAGTGGATATTTAGAGGGAATTGAGGCCTACCGT
GGAAAAGCATATACCTACAAACAAAACCTAAACAGAAGCATTCTGAGAAACTTCTTAGTGATGTGT
40 GCATTCGTCTCACAGAGTTGAAACTTTCCTTTGATTGAGCAGTTTTGAAACACTCTTTTTGTAGAA

FIG. 9B Continued

TCTGCAACTGGATATTTGGAGCCCTTTGAGGAATATTGTGGAAAAGGAAATATCTTCACATAAAAA
CTACACAGAAGCATTCTGAGAACTTCTTTATGAGGAGTCCATTCAACCCACAGAGTTAAACTTTT
CTTCTCATTGAGCAGTTTTGAATCTCTCTATTTGTAGAATCTGCAAGTGGATATTTGCTGCCCTTT
GAGGCATACTGAGGAAAAGCAAATATCTTCATATAAAAACTACACAGAAGCATTCTGAGAACTTC
5 TTTGGGATGTGTGCATTCAACTCACAGAGTTGAACCTATCTTTTGATTGAGCAGATTTGAATCTCT
CTTTTGGCAGAACTGCAAGTAGATATTTGGAGCCATTTGCGGCCTTTTGTGGAAAAGAAATATT
TTCAAATAAAAACTAAACAGAAACATACTGAGAACTTCTTTGTGATGTGTGCATTCATCTCACAG
GGTTGAAACTATCTTATGATTGAGCAGTTTTGAAACACTCTTTTGTAGAATCTGCAACTGGATAT
TTGGAGCCCTTTGAGGGCTATTGTGGAAAAGTAAATATCTTCACATAAAAACTATTCAGGAGCATT
10 CTGATAAACTTCTTTGTGATGTATGCATTCAACT

>D13Z8 Acc No. GJ211968.2 (SEQ ID NO: 132)

GGCATTCTGAGAACTTCTTCGTGATGTGTGCATTCATCTCACAGAGTTGAACCTATCTTATGATT
GAGCAGATTTGAAACACTGGCTTTGTAGATGCTGCAAGTGGATATTTGGAGCGCTTTGAGGCCTAC
TGTGGAAAAGGAAATATTTTCACATAAAAAACCACACAGAAGCCTTCTGAGAACTTCTTTGTGATG
15 TGTCCATTCAACTCACTGCCTTGAACCTATCTTTTGATTGAGCAGTTTCGGATCTCTCTTTTTCAG
GAACCTGCAAGTGGATGTTTGGAGCCCTTTGCGGTCCATGGAGGCAAAGGAAATATCTTCAAATAA
AACTACACAGAAGCGTTCTGAGAACTTCTTTGTGATGTGTGCATTCATCTCACAGAGAAGAACC
TATCTTATGACTGAGCAGCTTTGAAACACTCTCTTTCAGGAATCTTCAGGTGTATATTTGGAGTGC
TTTTTGGCCTATTTTGTAAAAGGAAATATCTTCACATAAAACCTACACAGATTTATTTCTGAGAAAC
20 TACTTTTTGTTGTGTGTATTCATCTAACAGAGTTGTACGTTTCTTTTGATTGAGCAGTTTTGAAAC
ACAGTTTTTACAGAATGTGCAAGTGGATATTTGGAGCGCTTTGGGGCCTACTGTGGAAAAGCAAAT
ATCTTCACGTAAAAACCACACAGAAGCATTCTAAGAACTTCTTTGTGATGTGTGCATTCATCCCA
CAGAGATGAAAGTTTCTTTTGATTGAGAACTTTTGAACACTCTTTTGTAGAATCTGCAAGTGGAT
TATTTGGAGCGCTTTTAGACCTACTGTGGAAAAGGAAATATCTTCCATAAAAACTACACAGAAGC
25 ATCTGAGAACTTCTTTGCAATGTGTGCATTCAACTCATAGAGTTGAACCTATCTTTTGATTGAG
CAGTTTTGAATTTCTCTGTTTGCAGAATCTGCAAGTGGATATTTGAAACCTTTGCAGCCAATGGA
GGATAAGGAAATATCTTCAAATAAATACTACACAGAAGCATTCTGAGAACTTCTTTGTGATGTGT
GCATTCATCTCATAGATTTGAACATATCTTATGATTGAGTACATTTGAAACACTCTCTTTGTAGAA
TCTGCAAGTGGATACTTTGAGGGCTTTTCAGGACTGTTGTGGAAAAGGAAATATCTTCACGTAAAAA
30 CTTACAAAAGCATTCTGAGATACTTCTTTATGATGTATGCACTCAACTCAGAGAGTTGAACCTAT
CTTTTGATTGAGAAGTTTGAACACTCTGTTTGTAGAATCTGCAAGTGGATATTTGGAGTGCTTT
GAGGCCTATTGTGGAAAAGGTAATATCTTCACATAAAAACTACAAATAAGCATTCTGAGAACTTC
TTTGTGATGTGTGAATTCAAGTCACAGAGTTTAACCTATCTTTTGATTGAGCAGTTTTGAATCTCT
CTTTTTGTAGAATCTGCAAGTGGATATTTGGTGGGCTTTGAGGCCTATTTTGGAAAAGGAAACATC
35 TTCACATAAAAACTACACAGAAGCCTTCTGAGAACTTCTTTGTGATGTGTCCATTCAACTCACAG
CCTTGAACCTATCTTTTGATTGAGCAGTTTCGGATCTCTCTTTTTCAGGAACCTGCAAGTGGATGT
TTGGAGCCCTTTGCGGTCCATGGAGGCAAAGGAAATATCTTCAAATAAAAACTACACAGAAGCATT
CTAAGAACTTCTTTGTGATGTGTGCATTCATCC

>D13Z9 Acc No. GJ211969.2 (SEQ ID NO: 133)

GGAATTCTGAGAGACTTCTTTGTGATGCGTGTACTCATCTTACAGAGTTAAACCTTCCTTTTGAAT
GAGCAGATTTTAAAAGTGTCTTTTGTAGAATCTGCAAGTGGACATTTTGAGCGCCTTGAGGCCTAT
GGTGGAAAAGAAAATGCCTTCACATGAAAAGTACAGAGAAGAATTCTGAGAACTTCTTTCTTATG
5 TGTGCGTTAATCTCACACAGTTGAACCTTTCTTTTGATTGAGCAGTTTCAAACACTCTTTTGTAG
AATCTGCAAGTGGACTTTTGGAGCACTTTGTGGCCTACGGTAGAAAAGGAAATATCGTCACATAAA
ATCTAGACAGAAGCAATCTGAGACTTCTTTGTGATGTGTGCATTACCCACACATTGTGTAACCTTT
CCCTTGATTGAGCAGTTTGAAGTCTTTTGTAGAATCTACAAGTCTACATTTGGAGTGCTTTGA
GGCCTATGGTAGAAAAGGAAATATCTTCACATAAAAAGTACTGAAAAGAATTCTGAGAACTGCTT
10 GGTGATGTGTGCGTTACCCACACAGAGCTGAACCATTTGTTTGTATTGAGCAGTTTGGAAACCTCT
TTTTGTAGAATATGCAAGTGGACATTTGGAGTACTTTGATGCCCTGGTCGAAAAGGAAATATCTT
AACTTAAAAAGTACAGAGAATAATTCTGGGAACTTCTTTCTGATGTGTGCGTTTCATCTCACAGAG
TTAAACTTTTCAATTTATTGAGCAGTTTGGAAACACTCTTTTGTAGAATCTGCAAGTGGACATTT
GGAGCGCATTGTGGTATGCAGTAGAAAAGGAAATGTCTCCACAAAAATGTAGACAGAAGCAGTCT
15 TATAAACTTCTTTGTGATGTGTGCATTCAATTCACAGATTTGAACCTATCTTTAGATTGAGCAGTT
TGGAACACTCTTTTGTAGAATCTGCAAGTGCACACGTGGAGAGATTTGCGGCCAATGGTAGAGA
AGCAAATATCTTCGCATAAACTCTAGACAGAAGCATTCTGACAACTTCTTTGTGATGTGTGCATT
CATCTCACAAAGAATTGAACTTTCTTTGATTTCAGGAGCTTTGAAACACTCTTTTGTAGAATCTG
CAAGTGTACATTTGGAGCGCTTTGAGGCCTATGGTGGAAAAGGGAACATCTTCACATATAGAACAG
20 ACAGAAGCCTTCTGACAACTTCTTTTCAATGTGTGCGTTCAACTCAAAGATTTGAACCTTACTTT
TCATTGAGCAGATTTGAAACACTCTTTTGTAGAATCTGCAAGTGGACAATTGGACCGCTTTCTGG
CCTATGGTGGAAAAGGATGTATCGTCACATAAAAAGTACAGAAAATCTTCTGACAACTTCTTTG
TTATGCATGCATTCACTTTTCAGAGTTGAACCTTCCTTTTGTATTGAGCACTTTGAAACACTCTTT
TTGTAGAATCTGCAAGTAGTCATTTGTAGCGCTTTGGAGACTATGGCGAAAAGGAAATATCTTCC
25 CATAAAAAGTACAGAGGCATTCTGACAACTTCTTTGTGATGTGGGCATTCACTCACAGAGTT
GAACCTTACTTTTCAATTGAGCTATTTTGAACACTCTTTTGGAGAATCTGTAAGTGGACTTTTGT
AGGGCTTTGACGCACGTGGTGGAAAAGGAAATATCTTCATATAAAAAGTACAGAGAAGGATTCTGA
GAACTTCTTTGTGATGTGTGCATTCACTCACA

>D14Z1/D22Z1 Acc No. M22273.1 (SEQ ID NO: 134)

30 CAATCTCAGAATCTTCTTTGGGATATATGCACGCAGCTAACAGAGTTGAACTTTTCTATTGACAGA
GCAGTTTGTAAACAGTCTTCTGTGGAATCTGCAAGTGGATATTTGGATAGCTTGGAGGA {TTTCG
TTGGAACGGGA} TTACGTATAAAAAGCAGACAGCAG

>D14Z2 Acc No. GJ211972.2 (SEQ ID NO: 135)

AGCATTCTGTGAACTTGTTTGTGATGTGTGTACTCAACTAACAGTGTTGAACCTTTCTTTTACA
35 GAGCAGTTTGTAAACACTCTTTTGTAGAATCTGCGAGGGGATATTTGGATACATTTTCAGGA {TTT
CGTTGGAACGGGA} ATATCTTCATATAAAAATCTCGACAGAAGCATTCTCAGAACTTCTTTGTGA
TATGTGCATTCAAGTCACAGAGTTGAATATTCCTTTTCACAGAGTAGGTTTGAACACTCTTTTGT
TAGTATCTGAAAGTGGACATTTGGAGCGCCTTGACACCTACGGTGAAAAGGGAATATCTTCCCAT
AAAAAGTACAGAGAAGCAATCTCAGAATCTTCTTTGGGATATATGCACGCAGCTAACAGAGTTGAA

FIG. 9B Continued

CCTTTCTATTGACAGAGCAGTTTTGAAACAGTCTTTCTGTGGAATCTACAAGTGGATATTTGGATA
GCTTGGAGGA { TTTCGTTGGAAACGGGA } TTACGTATAAAAAGTAGACAGCAAGCATTCTCATAAA
CTTGTTTGTGATGTGTGAACTCAGCTAACAGGCGTGGATCTTTCTTTTGATACAGCAGTTTTGAAA
AACACTTTTTGTTGAATCTGCAAGTGGACATTTGGATAGATTTGAAGA { TTTCGTTGGAAACGGGA
5 } ATATCTTCATATCAAATCTAGACAGAAGCATTCTCAGAAACGTCTTTGTGATGTTTGCATTCAAC
TCATAGAGTTGAACATTCCGTTTCAGAGAGCAGGTTTGAAGCACTCTTTTTGTAGTATGTGCAAGT
GGATATTTGGAGCGCTCTGAGGCCTACGGTGAAAAAGCAG

>D14Z3 Acc No. GJ211986.2 (SEQ ID NO: 136)

AATATTCTGGGAAAGTTCTTTGTGGTGCCTGCATTCATGTCATAGAGTTGAAACTTTCTTTTGATG
10 GAGCAGTTTTGAAACACTCTTTTTGTACAATCTGCTAGTGGATAATTGGAGCCCTTTGAGGACTAT
TGTGGAAAAGGAAATATCTTCACGTAAAACTACATAGAACCATTCTGAGATACTTCTTTTTGATG
TTTGCATTCATCTCACAGTGTGAAACTTTCTTTTGATTGAGCAGTTTTGAAACACTCTTTTTGT
GAATCTGCAAGTGAATAATTGGAGCCCTTTGAGGGCTATGGTAGAAAAGGAAATATCTTCAAATA
GAACTACAAAGAAACATTCTCAGAAACTTATTTGTGATGTGTGCATTCAACTCACAGGGCTGAACA
15 TATCTTTTGATTTAGCAGTTTTGAATTTCTCTTTTGCAGAATCTGCAAGGGGATGTTTGGAGAGC
TTTCAGGCATATTGTGGAAAGGGAAATATTTTCACATAAAACTACACAGAACCATTCTGAGAAAC
TTCTTTGTGTGCTGTGCATTCAACTCACAGAGTTGAACATATGTCCTCTTTGAGCAGTTTTGCGTC
TCTCTTTTGTAGAATGTACAAGTGGATATTTGGAGCCCATTGTGTCCTATGGTGGAAAAGGAAAT
ATCTTCAGATAAAAATTACACAGAAGCATTCTGATAAACTTCTTTGTGATGTATGCATTCAACTCA
20 CAGACTTGAACCTATCTTAAGAATGAGCAGTTTTGAATCTCTCTTTTGCAGAATCTGCAACTGGA
TATTTTGGAGGGCCTTAAGGCCTACCGTGGAAAAGCAATTATCTTCAGATTAAAACCTACACAGAAGC
ATTCAGAGAAACATCTTTGTGATGTTTGCATTCATCTCACAGAGTTAAAACCTTCTCTTGATGGAG
CAGTTTTGAAACACTCTTTTTGTAGAATCTGCAAGTGGATATTTGGAGCCCTTTGAGGCCTGTTGT
GGAAAAGGAAATATCTTCCCATGAAAACCTACATAGAAGTATCTGAGAACTTCTTTGCAATGTGT
25 GCATTCAACTCACAAGAGTTGAACCTATCTTTGATTGAGGATTTTTGAATCTTTCTTTTGCAGA
ATCTGCAAGTGTATGTTTGCAAAGCTTTGTGGCCTATTGTGGAAAAGGAAATGTCTTCACATAAAA
ACTACACATA

>D15Z3 Acc No. AF237720.1 (SEQ ID NO: 137)

ATTCTCAGAACTTATTTGTGATGTGTGCTTCAACTAACAGAGTTGAACCTTTGTTTTGATACAG
30 CAGTTTGGAAACACTCTTTTTGTAGAATCTACAAGTGGATATTTGAGAGCATTGAAAA { TTTCGT
TGGAAACGGGA } ATACCTTCATATAAAATCTAGACAGAAGCATTCTCAGAACTTCTTTGTAATGT
TTGCATTCAACTCATAGAGTTGAACATTCCCTTTACACAGCAGGTTTGAAACACTCTTTTTGTAG
TATGTGGAAATGGACATTTGAGCACTCTTAGGCCTAAGTTGAAAAGGGAAATATCTTCAAATAAG
AACTAGTCAGCAGCATTCTCAGAAAGCTCTTTGTGATGTGTGTAAGTACTCAGCTAACAGAGTTGAACCT
35 TCCATTTACAGAGCAGTTTGGAAACACTCTTTTTGTGGCATTGCAAGTGGACATTTGGATAGCT
TTGAGGA*TTTCTTTGAAACGGGA*TAATTTTCATATAAAATCCTAGACAGAAGCATTCTCAGAAAT
CTTCCTTGTGATGTATGCCCTCAATTCACAGAGTTGAACCATTGTTGGGATACAGCATTTTGGGA
AACATTCCTTTTGTAGAATCTGCAAGTTGATATTTGTATAACTTTGAGGA { TTTCGTTGGAAACGG
GA } ATATCTACATATAAAATCTAGACAGAAGCATTCTCAGAAACCTGTTTAAATGTTTTCATTCAA

FIG. 9B Continued

CTCATACGTTTCAACAATCCCTAACCTAAATTAAGGTTTGAACACTCCTTTTGTANTATGTGGAA
 ATGGACATTTGGANCGCTTTGANGCCTACGGTGAAAAAGGAAATATCCTTCCCATAAAAACTNNGA
 CAGGAAGCATTCTCANACANTTGAAAATTTTCGATTANTTCTTGAGGAT*TTTCGTTTGGAAACGG
GA*NTTACCAANTTGAAAGTTGACCAGCCNGCCTTTTCTCAGAAAATTGCTTTGTGATGTCTGCC
 5 ATTCANGTCACATTGGTTGAACATTCCTTTCTTGAGGCCAGNTTTGAATCACTTGTCTCTNGTAG
 TATCTTGGAAGTGGATATTTTCGAGCGCTTTTGAGGCCTTAAGNTGAGAAAGGAAATGTCCTTCAA
 TTAGGAACTGGACAGANGCATTCTCAGAACTTATTTGTGATGTGTGTCTCACTAACAGAGTTG
 AATCTTTGTTTTGATACAGCAGTTTGAAACACTCTTTTTGTAGAATCTACAAGTGGATATTTTGA
 GAGCATTGAAAA [TTTCGTTGGAAGCGGGA] AAACCTTCATATAANTTCTAGACAGAAGCATTNTA
 10 AGAAACTACTTTGCAATGTTTGCATTCACTCATAGAGTTGCACATTCCGGTTCATACAGCAGGTT
 TGAAACACTCTTTTGTAGTATGTGGAAATGGACAATTGGAGCCCTTTGAGGCCTACGGTGAAAAA
 GGAAATATCTTCCCATAAAAACTAGACAGAAGCATTCTCAGAACTTGTTTGTGACGTGTGTATTC
 AACTAACAGAGTTGAACCTTTCTTTTACAGGGCAGCTTTGAAACACTCTTTTGTGGAATCTGCA
 ATTGAAAATTTCCGTAGTTCTGAGGA { TTTCGTTGGAACGGGA } TTACAAATAGAAAGTAGACAG
 15 CAGCATTCTGAGAACTGCTTTGTGATGTTTGCATTCAAGTCACCTAGTTGAACATTCCCTTTTAT
 AGAGCAGGTTTGAATCACTATTTCTGTACTATCTGGAAGTGGATATTTTCGAGCGCTTTCAGGCGTA
 AGGTGAGAAAGGACACGTGCTTCNATGNCAGAACTAACAGAGATTA

>D16Z2 Acc No. M58446.1 (SEQ ID NO: 138)

GAATTCTCAGTAACTTCCTTGTGTTGTGTGTTATTCAACTCACAGAGTTGAACGATCCTTTACACAG
 20 AGCAGATTTGAAACACTGTTTTTCTGGAATTTGCAAGTGGAGATTTAGCCGCTTTGAGGTCAATG
 GTAGAAAAGGAAATATCTTCGTATAAAAACTAGACAGAATGATTCTCAGAACTCCTTTGTGATGT
 GTGCGTTCAACTCACAGGGTTTAACTTTCTTTTACAGAGCAGTTAGGAAACACTCTGTTTGTGA
 AGCCTGCCAGTGGATATTCGGACCTCTTTGAGGC/CTTCGTTGGAACGGGA/TTTCTTCATATTA
 TGCTAGACAGAAGATTTCTCAGTAACTTCTTTGTGTTGTGTGTATGCAACTCACAGAGTTCAACCT
 25 TCCTTTAGACAGAGCAGATTTGAAACACTCTTTTGTGGAATTTGCAAGTGGAGATTTCAAGCGCT
 TCGATGCCAATGGTAGAAAAGGAAATATCTTCGTATAAAAAACAAGACAACTCGTTCCAGACACT
 GCGTAGTGATGTGTGCTTTTAACTCACAGAGTTTACCTTTCTTTTACATACAGCATTCTGGAAACC
 CTCTGTTTGTAAAGTCTGCAAGTGGATATTTGGACCTCTTAGATGC*CTTCGTTGCAAACGGGA*T
 TTCTTCATATAATGCTAGAGGGAG

30 >D17Z1 Acc No. M13882.1 (SEQ ID NO: 139)

A*ATTCGTTGGAACGGGA*TAATTTAGCTGACTAACAGAGCAGTCTCAGAATCTTCTTTGTG
 ATGTTTGCATTCAAATCCCCGAGTTGAACTTTCTTTTCAAAGTTACGTTTGAACACTCTTTTGTG
 CAGGATCTACAAGTGGATATTTGGACCACTCTGTGTCTTCGTTTCGAAACGGGTATATCTTCACAT
 GCCATCTAGACAGAAGCATCCTCAGAAGCTTCTCTGTGATGACTGCATTCAACTCACGGAGTTGAA
 35 CTCTCCTTTTGTAGAGCGCAGTTTGAACCTCTTTTCTGTGGCATCTGCAAGGGGACATGTAGACC
 TCTTTGAAGA*TTTCGTTGGAACGGGA*ATCATCTTCACATAAAAACTACACAGATGCATTCTCAG
 GAACTTTTGGTGATGTTTGTATTCAACTCCCAGAGTTGAACTTTCTTTTGGAAAGAGCAGCTATG
 AAACACTCTTTTCTAGAATCTGCAAGTGGACGTTTGGAGGGCTTTGTGGTTTGTGGTGGAAAAGG
 AAATATCTTCACCTAAATACTAGATAGAAGCATTCTCAGAACTGCTTTGTGATGATTGCATTAC

FIG. 9B Continued

CTCACAGAGTTGAACATTCCTATTGATAGAGCAGTTTGGAAACACTCTTGTTGTGGAATGTGCAAG
TGGAGATTTGGAGCGCTTTGAGGCCTATGGTAGTAAAGGGAATAGCTTCATAGAAAACTAGACAG
AAGCATTCTCAGAAAATACTTTGTGATGATTGAGTTTAACTCACAGAGCTGAACATTCCTTTGGAT
GGAGCAGGTTTGAGACACTCTTTTTGTACAATCTACAAGTGGATATTTGGACCTCTCTGAGGA {TT
5 TCGTTGGAAACGGGA} TAACTGCACCTAACTAAACGGAAGCATTCTCAGAACTTCTTGGTGATGT
TTGCATTCAAATCCCAGAGTTGAACCTTCCTTTGATAGTTTCAGGTTTGAAACACTCTTTTTGTAGG
ATCTGCAAGTGGATATTTGGACCACTCTGTGGCCTTCGTTTCGAAACGGGTATATCTTCGCATAAAA
TCTAGACAGAAGCCTTCTCAGAACTTCTCTGTGATGATTGCATTCAACTCACAGAGTTGAACCTT
CCTATGGATAGAGCAGTGTGAAACTCTCTTTTTGTGGAATCTGCAAGTGGATATGTGGACCTCTC
10 CGAAGA *TGTCTTTGGAAACGGGA* ATATCTTCACATAAAAACTAAACAGAAGCATTCTCAGAAAC
TTCTCTGTGATGTTTGTGTTCAACTCCCAGAGTTTCACATTGCTTTTCATAGAGTAGTTCTGAAAC
ATGCTTTTCGTAGTGTCTACAAGTGGACATTTGGAGCGCTTCAGGCCTGTGGTGGAAAACGAATT
ATGGTCACATAAAAACTGGAGAGAAGCATTGTCAGAACTTCTTTGTGATGATTGCATTCAACTCA
CAGAGTTGAAGGTTCCCTTTTCAAAGAGCAGTTTCCAATCACTCTTTGTGTGGAATCTGCAAGTGG
15 TATTTGGACCTTTTTGAAGA {TTTCGTTGGAAACGGGA} GAATCTTCACAGGAAAGCTAAACAGAA
GCATTCTCAGAACTTCTTTGTGATGCTTGCATTCAACTCACAGAGTTGAACTTTCTTTTCGAGAG
AGAAGCTTTGAAACACTCTTTTTTCCAGAATCTGCAAGTGGACATTTGGAGGGCTTTGAGGCCTGTG
GTGGAAGGAATTATCTTCCCGTAAAAGCTAGATAGAAGCATTCTCAGAACTACTTTGTGATGA
TTGCATTCAAGTCACAGAGTTGAACATTCCCTTTGACAGAGCAGTTTGGAACTCTCTTTGTGTAG
20 AATCTGCAAGTGGAGATATGGACCGCTTTAGGCCTATGGTAGTAAAGGAAATAGCTTCATATAAAA
GCTAGACAGTAGCATTACAGAAAACTCTTGGTGACGACTGAGTTTAACTCACAGAGCTGAACATT
CCTTTGGATGGAGCAGTTTCGAAACACACTATTTGTAGAATGTGCAAGTGGATATTTAGGCCTCTC
TGAGGA {TTTCGTTGGAAACGGGA} TAAACCGCACAGAACTAAACAGAAGCATTCTCAGAACCTTC
TTCGTGATGTTTGCATTCAACTCACAGTGTGAACTTTCTTTGATAGTTTCAGGTTTGAAACGGTC
25 TTTCTGTAGAACTGCAAGTAGATATTTGGACCGCTCTGAGGA {TTTCGTTGGAAACGGGA} TAAC
CCGCACAGAACTAAACAGAAGCATTCTCAGAACCTCTTCGTGATGTTTGCATTCAACTCACAGTG
CTGAACCTTTCTTTGATAGTTTCAGCTTTGAAACACTCTTTTTGTAGAACTGCAAGTGGATATTTG
GTCCTCTCTGAGCA {TTTCGTTGGAAACGGGA} TAACTGCACAGAACTAAACAGAAGCACTCTCA
GAACCTTCTTCGTGATGTCTGCATTCAACTCACAGTGTGGAACCTTTCTTTGATAGTTTCAGGTTG
30 AAACACTCTTTTTGTAGAACTGCAAGGGGATCATTGCACTCTTTGAGGAGTACCGTAGTAAAGGA
AATAACTTCTATAAAAAGAAGACAGAAGCTTTCTCAGAAAATTCTTTGGGATGATTGAGTTGAAC
TCACAGAGCTGAGCATTCCCTTGCGATGTAGCAGTTTAGAAACACACTTTCTGCAGAACTGCAATT
GCATATTTGGACCTCTTTGAGG

>D17Z1b Acc No. GJ212053.1 (SEQ ID NO: 140)

35 ATACTAGATAGAAGCCTTCTCAGAACTACTTTGTGATGATTGCATTCACCTCACAGAGTGGAGCA
TTCCTATTGACAGAGCAGTTTGGAAACACTCTTGTTGTAGAATCTGCTAGTGGAGATTTGGAGCGC
TTTGAGGCCTATGGTAGTAAAGGGAAGAGCTTCACATAAAATCAAGACAGAAGCATTCTCAGAAAA
TACTTTGTGATGATTGAGTTTAAACACACAGAGCTGAACATTCCTTTGGATGGAGCAGGTTTGAAAC
ACACTTTCTATAGAATCTGCGAGTGGATATTTGGACCTCTCTGAGGA {TTTCGTTGGAAACGGGA}
40 TAACTGCACCTAACTAAACGGTAGCATTCTCACAAAATTCTTCGTGATGTTTGCATTCAAATCCCA

GAGTTGAACCTTCCTTTGATAGTTCAGGTTTGAAACACTCTTTTTGTAGGATCTGCAAGTGGATAT
TTGGACCACTCTTTGGCCTTCCTTCGAAACGGGTACATCTTCAAATAAAATCTAGACAGAAGCCTT
CTCAGAAACTTCTCTGGGATGATTGCATTCAACTTTGAGTTGAACCCTGCTATGGATAGAGCAGTT
TAGAATCTCTCTTTTTGTGGAATCTGCAAGTGGATATGTCGTCCTCCTTGAAGA*TGTCTTTGGAA
5 ACGGGA*ATATCTTCACATAAAAACTAAACAGAAGCATTCTCAGAAACTTCTCTGTGATGTTTGTG
TTCAAATCACAGAGTTTCACGTTGCTTTTCATAGAGCAGATGAGAAACATGCTTTTTCGTAGGGTCT
GCAAGTGGACATTTGGAGAGATTTCAAGCCTGTGGTGGAAAACGAATTATCGTCACGTAAAACTA
GAGGGAAGCATTGTCAGAAACTTGTGTGATGACTGCATTCAACTCACAGAGTTGAAGGTTCCCTT
TTCAAACAGCAGTTTCCAAACACTCTTTCTGTGGCATCTGCAAGTGGATGTTTGGGCCTCTTTGAA
10 GA*TTTCGTTGGAAAAGGGA*TAATCTTCACAGAAAAGCTAAACAGAAGCATTCTCAGAAACCTCT
TTTTGATGTTTGTCTTTAACTCACAGAGTTGAACTTTCCTTTTGAGAGAGAAGCTTTGAAACACTC
TTTTTCTAGAATCTGCAAGTGGATATTTGGAGGGCTTTGAGGACTGTGGTGGAAAAGGAATTATCT
TCCCGTTAGAACTAGATAGATGCATTCTCAGAAACTACTTTGTGACGATTGCATTCAAGTCACAGA
GGTGAACATTCCCTTTTCACAGAGCACTTTGGAAACTCTCGTTGTGTAGAATCTGCAAGTGGAGATA
15 TGGACCGCTTTGAGGCCTACGGCAGTAAAGGAAACAGCTTCATATAAAAACTAGACAGCAGCATT
TCAGAAACTCTTTCTGACGACTGAGTGTAACCTCACAGGGTTGAACATTCCCTTTTGATGGAGCAGT
TTGGAAACACACTATCTGTAAGATCTGCAAGCGGATACTTGGGCCTCTCTGAGGA { TTTCGTTGGA
AACGGGA } TAAACCGCACAGAACTAAACAGAAGCATTCTCAGAACTTCTTCGTGACGTTTGCATT
CAACCCACAGTGTTGAACCTTTCTTTGATAGTTCAGGTTTGAAACACTCTTTTTGTAGAAAGTGCA
20 AGAGTATAACTGCACTTCTTTGAGGCCTATCGTAGTAAAGGAAATAACTTCCTATAAAAACAAGAC
AGAAGCTTCTCAGAAAATTCTCTCAGATGATTGAGTTGAACTCACAGAGCAGTACTTTCCTTGGG
ATGGAGTAGTTTTCGAAACACACTTT

>D18Z1 Acc No. M65181.1 (SEQ ID NO: 141)

CTGCAGCGTTCTGAGAAACATCTTTGTGATGTTTGTATTTCAGGACACAGAGTTGAACATTCCCTAT
25 CATAGAGCAGGTTGGATCACTCCTTTTTGTAGTATCTGGAAGTGGACATTTGGAGCGCTTTCAGGCC
TATTTTGGAAAGGGAAATATCTTCCCGTAACAACTATGCAGAAGCATTCTCAGAAACTTGTGTG
ATGTGTGCCCTCTACTGACAGAGTTGAACCTTTCTTTTCATAGAGCACTTTTGAAACACTCTTTTT
GTAGAATCTGCAAGAGGATATTTGCATAGCTTTGAGGA*TTTCGTGGGAAACGGGA*TTGTCTTCA
GGCAAAGTCTAGACAGAAGCATTCTCAGAACTTCTTTGGGATGTTTGCATTCAAGTCACAGAGTA
30 GAACATTCCCTTTGGTAGAGCAGGTTTGAAACACTCTTTTTGTAGTATCTGGAAGTGGACATTTGG
AGCGCTTTCAGGCCCATGTTGGAAAGGGAAATATCTTCCCGTAACAACTAGGCAGAAGCATTCTCA
GAAACTTATTTGAGATGTGTGTACTCAACTAAGAGAATTGAACCTCCGTTTTGAAGGAGCAGTTTT
GAAACACTCTTTTTCTGCAATCTGCAAGAGTATATTTGCCTAGCCTTGAGGA { TTTCGTTGGAAAC
GGGA } TTGTCTTCAGAGAAAATCTAGAGAGAAGCATTCTCAGAACTTCTTTGGGATGCTTGCATT
35 CAAGTCACAGAGTAGAACATTCCCTTTGGTAGAGCAGGTTTGAAACACTCTTTTGTAGTATCTGGA
AGTGGACATTTGGAGCGCTTTCAGGCCTACGTTGGAAAAGGAAATATCTTCCCATAACTAGAC
AGAAGCATTCTCAGAACTAGTTTCTGATGTGTGTCCTCAACTAACAGAGTTGAACATTTCTTTAG
ACAGAACAGTTTTGAAACACTCTTTTTGTGGAATCTGCAAGTGGCTATTTGGCTAGATTTGAGGA {
TTTCGTTGGAAACGGGA} TTACATATAAAAAGCAGTCAGCAGCATTCTCAGAAAGTTCTTTGTGAT
40 GATTGCATTCAAGTCACAGAATTGAACATTCCCTTTCACAGAGCAGGTTTGAAACACTCTTTTTGT

FIG. 9B Continued

AGTGTGTGTAAGTGGACATTTGGAGCACTTACCGGCCTAAGGTGAAAAAGGAAATATCTTCCCATA
 AAAACTGGACAGAAGCATTCTCAGAACTTACTCGTGATGTGTGTCCTCAACTAAAGGAGTAGAAC
 CTTTCTTTTCATAGAGAAGTTTTGAAACGCTCTTTTGTGAAATCTGCAAGTGAATATTTGGCTGG
 TTTTGAGGA [TTTCGTTGGAAGCGGGA] ATTCATACAAATTGCAGACTGCAG

5 >D18Z2 Acc No. M65182.1 (SEQ ID NO: 142)

CTGCAGATGGATATTCGGATAGCTCTGAGGA*TTTCGTTGGAGACGGGA*ATACATAAAGAAAGTA
 GACAGCAGCATTCTCGGGAGATTCTTTGTGATGTTTGCTTTGAAGTCACAGAGTTGAATATTCCT
 TCAATAGAGCAGGTTTGAAACACTCTTTCTGTAGTATCTGGAAGTGGACATTTTCGATCGATTTTCAG
 GCCTATGTTGAAAAAGGAAATATCTTAACATAAAAACTAGACAGAAGCATTCTCAGAAACGCTTTT
 10 GTGATGTGTGTCCTCAACTAACAGAGTTCAACCTTTCTTATGATACAGCAGTTTGGAACACTCTT
 TTTATAGAATTTGCAAGTTGATACATGGATAGCCCTAACTA { TTTCGTTGGAACCGGGA } ATATCT
 TCACATAAAACCTAGACAGAAGCACTCTCAGAACTACTTTGTGATATCTGCATTGATATCAGAGA
 GTTGAATATTCCTTTCTAAGGGCAGGCTTGAAAGCGTATTTTCGTGGAATCTGCAG

>D19Z4 Acc No. AC004164 (SEQ ID NO: 143)

15 TTGTTTCTAGTTTTTATCATGGGATATTCCGTTTTTTACTAAAGACCTCAATGGGCTCAGGACTGT
 CCCTAAGTAGATTTTACAGGAAGAGTGTTTTGAACCTGTTGAAACAAAACAAAGGTTAACGTCTGT
 GAGATGAATCCACACATCACAAAGCATTTCACAGAA

>D19Z5 Acc No. AC006504 (SEQ ID NO: 144)

ATTTTTTCTAGGTTTATTGCAGGATATTTAGTTTTTCACTATTGGCCTCAATGGGTTTCAGAAATGA
 20 CCCTTCGTAGCTTCTATAACAAGAGTGTTTCCAGCCTGTTGAATCAAAATAAAGTTTGAAATCTGT
 GAGATGGATCCATGCATCACA

>D20Z2 Acc No. X58269.1 (SEQ ID NO: 145)

ACTAACAGAACTAAACCATCGTTTTGAAGGAGCAGTTTTGAAACACTCTTTTTGCGGAATCTGCAA
 GTGGATATTTGGCTAGCTGGGAGGA { TTTCGTTGGAACCGGGA } TTACATACAAAAAGCAGACAGC
 25 AGCATTCTCAGAACTTCTTTGTGATGTTTGCATTCAAGTCACAGAGTTGAACATTCCCTTTTCATA
 GAGCAGGTTTGAAACACTCTTTTTGTAGTATCTGGATGTGGACATTTGGATCGCTTTCAGGCCTAT
 GGTGAAAAAGGAAATATCTTCCCATGAAAGCTAGATAGAAGCTATCTCAGAACTTATTTGTGATG
 TGTGCCTCAACTGACAGTGTTGAACCTATATTTTGATAGAGCAGTTCTGAAACACACTTTTTGTAA
 AATCTGCAAGAGGATATTTGGATAGCTTTGAGGA*TTTGCTTGGAACCGGGA*ATGTCTTCATATG
 30 AACTCTGGACAGAAGCATTCTCAGAACTGCTTTGGGATGTTTCAATTGAAGTCCCAGTGTTGAAC
 ATTCCTTTTCATAGAGCAGGTTTGAAACACTCTTTTTGT

>D13Z1/D21Z1 Acc No. D29750.1 (SEQ ID NO: 146)

AATTCAAATAAAAGGTAGACAGCAGCATTCTCAGAAATTTCTTTCTGATGTCTGCATTCAACTCAT
 AGAGTTGAAGATTGCCTTTTCATAGAGCAGGTTTGAAACACTCTTTCTGGAGTATCTGGATGTGGAC
 35 ATTTGGAGCGCTTTGATGCCTACGGTGGAAGTAAATATCTTCCATAAAAACGAGACAGAAGGAT
 TCTCAGAAACAAGTTTGTGATGTGTGTACTCAGCTAACAGAGTGGAACCTTTCTTTTTACAGAGCA

GCTTTGAAACTCTATTTTTGTGGATTCTGCAAATTGATATTTAGATTGCTTTAACGA*TATCGTTG
GAAAAGGGA*ATATCGTCATACAAAATCTAGACAGAAGCATTCTCACAACTTCTTTGTGATGTGT
 GTCCTCAACTAACAGAGTTGAACCTTTCTTTTGATGCAGCAGTTTGGAAACACTCTTTTGTAGAA
 ACTGTAAGTGGATATTTGGATAGCTCTAACGA { TTTCGTTGGAAACGGGA } ATATCATCATCTAAA
 5 ATCTAGACAGAAGCACTATTAGAACTACTTGGTGATATCTGCATTCAAGTCACAGAGTTGAACAT
 TCCCTTACTTTGAGCACGTTTGAACACTCTTTTGGAGAATCTGGAAGTGGACATTTGGAGCGCT
 TTGACTGCCTTTGTTGAAAAGGAAACGTCTTCCAATAAAAGCCAGACAGAAGCATTCTCAGAACT
 TGTTTCGTGATGTGTGTACTCAACTAAAAGAGTTGAACCTTTCTATTGATAGAGCAGTTTGAACA
 CTCTTTTTTGTGGATTCTGCAAGTGGATATTTGGATTGCTTTGAGGA [TTTCGTTGGAAGCGGGA] A
 10 TTCGTATAAAAACTAGACAGCAGCATTCCCAGAAATTTCTTTCGGATATTTCCATTCAACTCATAG
 AGATGAACATGGCCTTTTCATAGAGCAGGTTTGAACACTCTTTTTGTAGTTTGTGGAAGTGGACAT
 TTCGATCGCCTTGACGCCTACGGTGAAAAAGGAAATATCTTCCCATAAAAAATAGACAGAAGCATT
 TCAGAACTTGTGGTGATATGTGTCTCAACTAACAGAGTTGAACTTTGCCATTGATAGAGAGCAG
 TTTTGAACACTCTTTTTTGTGGAATCTGCAAGTGGATATTTGGATAGCTTGGAGGA [TTTCGTTGG
 15 AAGCGGGA] ATTCAAATAAAAGGTAGACAGCAGCATTCTCAGAAATTTCTTTCTGATGACTGCATT
 CAACTCATAGAGTTGAACATTCCCTTTTCATAGAGCAGGTTTGAACACTCTTTCTGGAGTATCTGG
 ATGTGGACATTTGGAGCGCTTTGATGCCTATGGTGAAAAAGTAAATATCTTCCCATAAAAAACGAGA
 CAGAAGGATTCTGAGAAACAAGTTTGTGATGTGTGTACTCAGCTAACAGAGTGGAACTCTCTTTT
 GATGCAGCAGTTTGGAAACACTCTTTTTGTAGAACTGTAAGTGGATATTTGGATAGCTCTAATGA
 20 { TTTCGTTGGAAACGGGA } ATATCATCATCTAAAATCTAGACAGAAGCCCTCTCAGAACTACTTT
 GTGATATCTGCATTCAAGTCACAGAGTTGAACATTCGCTTTCTTAGAGCACGTTGGAAACACTCTT
 TTTGTAGTGTCTGGAAGTGGACATTTGGAGCGCTTTGATGCCTTTGGTGAAAAAGGGAATGTCTTC
 CCATAAAAACTAGACAGAAGCATTCTCAGAACTTGTTTTTGTGATGTGTGTACCCAGCCAAAGGAGT
 TGAACATTTCTATTGATAGAGCAGTTTTGAACACTCTTTTTGTGGAAAATGCAGGTGGATATTTG
 25 GATAGCTTGGAGGA [TTTCGTTGGAAGCGGGA]

>D22Z4 Acc No. GJ212162.2 (SEQ ID NO: 147)

AGCATTTCCAAGAAATTTTTTGTGATGTGTCCATTTACGTACAGAGTTGAACCTCTCCTTTGATTG
 GGCAGTTTGGAAACAGTCTTTTTGTAGAACCTGCAGAGGGATATTTGTGAGCCCTTTATGGCCTGT
 AGTGAAATACGAAGTATCTTCACCTAAAACTAGACAGAAGGTTTCTGAGAACTTCTTGGTGATG
 30 TGTGCCTTCATCTCACAGTGTTGAACCTTTCTTTTGATTGAGCAGTTTGCAAAGTCTTTCTGTAGA
 ATCTGTAAATGGATATTTGGAGATATTTGAGGCCCGTGGTGAAAAAGGAAGTATCTTCACCTAAAA
 ACCAGACAGAAAGATTTCTGAAAAACCTCTTTGTTATGTGTGAATTCATGTACAGAATTCAACCT
 TTCTTTCACTTGAGCAGTTTGGAAACAGTCTTTGGTAGAAGATGCAGAGGGAAATTTCTTAGCTGC
 TTCAGGCCTATGGTGAAAAAGAAATATCTTCACAGAAAACTAGACAGAAGCTTTCTGAGAACTT
 35 CTTTGTGATGTGTCCATTCATCGCACAGAGTGAAACCTTTCTTTTGATTGAGGAGTTTGGAAAAGG
 TCTTTTCTTAGAATCTGCAAAGGGATATTTGTGAGCCCTTTATGGCCTTTGTTGAAATATGAAATA
 TCTTCACGTAAAAAGTAGACAGAAGCTTTCTGACAAATTTCTTGGTGATGTGCACGTTTGTACAC
 GGAATTGAACCTTCTTCTGATTGAGCAGTTTGGAAATCAGTCTTTTTGTAGAATCTGTGAATGTGT
 ATTTAGAGAGTTTTAAGGCCTAGGGTGCAAGAGGCAATGTCTTCACATAAAAAACGACACAGTAGCA
 40 TTTTGAGAAACTCTTTGTGACATTTCCATTCATCTCTAATAGTTGGCCGTTTCCTTTCATTGAGC

FIG. 9B Continued

AGTTTGAAGCAGTCTTTTTCTACAACCTGCAAAGGGATATTTCTGAGCGGTTTGGGGCCAACGGT
GAAAAATAAATATCTTCCCATGAAAAGTACAGAGCAGTATTTGAGAACTTCTTTTTGATGTGTG
TATTCATCTCACAGATTTGAACCTTTCTTTAGATTTAGCAATTTGGAGAAAGTCTCTTGGTAGTAT
AAGTGGAGTTATATTTGCGAGCGGTTTAAGGCCTATGGTGCCAAAGGAAATACCTTCACATAAAAT
5 GCAGACAGAGGCTTTCCGAGAACTTCTTTGTGATGTGTGCTTTTCGTCTCACAGAGTTGTGCCTTT
CTTTTGATTGACCAGTTTGGGAACATTCTTTTTGTAGAATCTGCAAATGGATATTTGGAGCAATTT
GTGGCCTACGGTGAAAAAGGAAATATCTTCACAGAAAACTAGACAGGCAGACTCCTGAGAACTT
CTTTTTGATGAGTGCATTCATTTACATAGTTGAAACATGCCATAGGGGCCAGTTTGGAAACAGTC
TTTTGGTAGAGTCTGCAGACAGATATTTTTGAGTGGCTTAAAGACTATGGTGAAAAAGGAAACATC
10 TTCATAGCAACCAGACAGAAAGCAACCTGAGAACTTCTTTGGGATGTGTTTCATTCATCTCACAA
TGTTGAACGTTTCTTTTGATTGAGAAGTTTGTAAAGAGAACTTTGTAGAATCCGCAAAGAGATAT
GTGTGAGTCCCTTGATTCCTATGGCAAAATAGGAATTATCTTGAGATAAAAGCGAGACAGAAGATT
TCTGAGAACTTTTTTGTGATGTGTGCTTTCATC

>D22Z5 Acc No. GJ212163.2 (SEQ ID NO: 148)

15 AGCTTTCTGAGCAAACCTTCTTTGTGATGTGTGCATTCATCTCACAGTGTGAACTTTATTTTATT
TGAGCAGTTTAGAGACAGACTTTTTCTGCAATCTGCAAAGGTATATTTCTGAGCCATTGAGGTCT
GTGGTGAAAAAGGAATATCTTCACATTTAACTAGACAGAAAGATTCTGAGAACTTCTTTATGAT
GTGTGCATTCATCTCAGGTAGGTGAAATTTCTTTTGATGGAGCAGTTTGGAAACAGTCTTTTTCT
AGTATCTGCAGAAGGATATTTGTGAGCGGTGTAAGGACTACGCTGAAAAAGGAAATATCTTCACAT
20 AAAAACTAGACAGAAGATTTCTGAGAACTTTTTTGTGATGGGTGCTTTCATCTCACAGAGTTGAA
AATTTCTTTTGATTGAGCAGTTTGGAAACAGTCTTTTCGTATCATCTGCAAAGGGATGTGTGGAGC
GCTTTGTGGCCTAACGTGAAAATGGAAATATCTTCACATAAAATCTAGACAGAAGCATTCTGAGAA
ACTTCTTTGTGATGTGTTTCATTCATCTCACAATGTTGAACGTTTCTTTTGATTGAGAGGTTTGTA
ACACAACTTTTGTAGAATCTGCAAAGGGATATTTGTGAGCCCCTTGATTCCTATGGCAAAATAGGA
25 ATTATCTTGTACATAAAAACTAGACAGGAGAATTCTGAGAACTTCTCTTTGATGAGTGCATTCATT
TCACATAGTTGAAACATGCTATATGGGCCAGTTTGGAAACAGTCTTTTTGTAGGGTCTGCAGACAG
ATATTTTGTGAGTGGCTTAAAGACTGTGGTGAAAAAGAAATATCTTCACAGAGTAACCAGACAGAA
GCTTTCTGAGAACTTCTTTATGATGTGTGCTTTTCGTCTCAGAGAGTTGAGCCTTTCTGTTGATTG
ACAGTTTGGAAACATTCTTTCTGTAGAATCCGCAAATGGATATTTGGAGCAATTTGCGGCCTACG
30 GTGAAGAAGGAAATATCTTCACATAAAAACTAGACAGAAGCATTTTGAGAACTTCTTTTTGATGT
GTGTATTCATCTCACAGAGTTGAACGTTTCTTTTGATTTAGCGATTTGGAGAAAGTCTCTTGGTAG
TATAAGCGGAGTTATGTTTGTGAGTGGTTTAAGGCCTACGGTGCCAAAGGAAATACCTTCACATAA
AATGTAGACAGAAGCATTTTGAGAGAACTCCTTGTGACATTTCCATTCATCTCTAATAGTTGACCA
TTTCTTCTCATTGAGCAGTTTGGAAACAGTCTTTTCCTACAACTGCAAAGGGATATTTCTGAGCC
35 GTTTGGGGCCAATGGTGAAAAATAAATATCTTCACATGAAAAGTGGCAGAAAGCTTTCTGAGAAAC
TCCTTTGTGGTGTGCACGTTTGTATCACAGAGTTGAACCTTTCATTTGATTGAGCAGTTTGGAAAC
AGTCTTTTTGTAGAATCTGCAAATGTATATTTGGAGTGTTTTAAGGCCTATAGTGAAAAAGGAAAT
ATCTTCACATAAAAACTACACAGTAGCTTTCTGAGAACTTCTTTGTGATGTGTCCCTTCATCGCA
CAGAGTGAAACCTGTCTTTTGATTAGGAGTTTGA AAAATGTCTTTCTTAGAATCTGCAAAGGGA

FIG. 9B Continued

TATTTGTGAGCCCTTTATGGCCTTTGTTGAAATATGAAATATCTTCACGTAAAAAGTAGAGAGAAA
GATTTCTGAAAAACCTCTTTGTGATGTGTGAATT

>DXZ1 Acc No. X02418.1 (SEQ ID NO: 149)

GATCCGCAAGGGATATTTGGACCTCTTTGAAGG{TTTCGTTGGAAACGGGA}TAATCTTCACCTAA
5 AAGCTAAACGGAAGCATTCTCAGAACTTCTTTGGGATGTTTGCATTACCTCACAGAGTTGAAGT
TTCCCTTTGATAGCGCAGTTTGACACACTTTTTCTACAATGTGCAAGTGGCTATTTAGCGGGCTTG
GAGGACTGTGTTGGAAAAGGAAATATCTTCTCCTAAAAACGACATAGAAGCATTCTCAGAACTGC
TCTGTGATGATTGCATTCAACTCCCAGAGTTGAACATTCCCTTTTGATAGAGCAGTTTGCAAACACT
CTTTTTGTAGAATCTGCAAGTGGAGATTTGGACCGCTTTGAGGCCTGTGGTAGTGAAGGAAAGAAC
10 TTCATATAAAAACCAGACGGTAGCACTCTCAGAAAATTCTTTGTGACGATGGAGTTTAACTCAGGG
AGCTGAACATTTCGTTATGATGGAGCAGATTCCAAACACACGTTTTGTAGAATCTGCGAGGGGATAT
TTGGACTCTCTCTGAGGA{TTTCGTTGGAAACGGGA}TCACCTTCCCATAACTGAACGGAAAGCAA
ACTCAGAACATTCTTTGTGATGTTTGTATTCAACTCACAGAGTTGAACCTTCCTTTGATAGTTTCA
GTTTGCAACACCCTTGTAGTAGAATCTGCAAGTGTATATTTTGACCACCTTTGTAGCCTTCGTTTGA
15 AACGTCTATATCTTCACATCAAACCTAGACAGAAGCATTCTCAGAAAGTTTTCTGCGATGACTGCA
TTCCACTCACAGAGTTGAACAATCCTTTGTGATGGAGCAGTTTTTGAAACCCTCTTTCTTTGGAATC
TGCAAGGGGATATGTGGACCTCTTTGAAGA*TTTCACTGGAAACGGGA*TCATCTTCACATAAAAA
CTAAACAGAAGCATTCTCGGAAACTATTTTGTGATGTTTGTATTCAACTCCCAGAGTTGAACCTTC
CTTTTGAAAGAGCAGCTATGAAACACTCTTTTTTCGAGAATCTGCAAGTGGACGTTTGAGGGGCTTT
20 GAGGCTGTGGTGGAAAAGGAAATATCTTCACACAAAAACCAGATAGAAGCATTCTCAGAAACGACT
TTGTGAGGATGGCATTCAACTACTGGAGTTGACAATCCTATTGATAGAGCAGATTGGAATCACTCT
TTTTGTAGAATCTGCAAATGGAGATTTGGACTGCTTTGAGGCCTACGGTGGTACAGGAAGGAACTT
CATATAAAAGGCAAACGGAAGCATTCTCAGAATATTCTTTGTGATGATGGAGTTTCACTCACAGAG
CTGAACATGCCTTTTGATGGAGCAGTTTCCAAATACACTTTTGGTAGAATCTGCAGGTGGATATTT
25 GGAGCTCTCTGAGGA{TTTCGTTGGAAACGGGA}ATAATTTCCCATAACTAAACACAAACAATCTG
AGAAAGTTCTTCATGATGAATGCATTTAACTCGCAGAGATGAACCTGCCTTTGAGAGTTCAGGTTT
GAAACACTCTTTCTGTAGAATCTGCAAGTGGATATTTGGACCACTGGGTGCCTTCGTTGAAACGG
GTATATGCTCACGTAAAAACTAAAGAGAAGCATTCTCAGAACTTCTGAGTGATGATTGCATTCAA
GTCACACAGTTGAACCCTCCTTTTGATGGAGCAGTTTTGAACTGTCTTTTTGTAGAATCTGTAAG
30 TGGATACGTGGACCTCTTTGAAGA*TTTCTTTGGAATCGGGA*ATATTTCCACAGAAAACTAAAC
TGAAGCATTCTCAGAACTGCTTTGTGGTGTGTGTTTCGAGCCGAGAGTTTAACATTGCTTTTC
ATAGAGCAGTTTTGAAATATTCTTTTGGCAGAATCTGCAAGTGGACATTTGGAGCGCTTTCAGGCC
TGTGGTGGAAAAGCCTGAAAGCCTTTTCGCTTTATCTTCACAGAAAGACGAGAGAGAAGCATTGTC
AGAAACTTCTTTGTGATGATTGCATTCAACTCACAGAGTTGAAGATTCCTTTTGAAACAGCAGTTT
35 CGAAACACTCTTTCTGTGG

>DYZ3 Acc No. M29723.1 (SEQ ID NO: 150)

TTTAAAGTGTCTTTCTGTAGAGTCTGCAAGTGGTTACTTTGGAGACCTTTGTGGAAGATGGTGGAAA
AGGAAATATCTTCCCGTAAAACTACACAGATGCATTCTGAGAACTTCTTTGTGATGTGTGCATT
CATCTCACAGAGTTCAACCTATCTTTTCGTAGAGCAGTTTGAAACTCTCTTTTCCTAGAATCTGT

FIG. 9B Continued

AAGTTGATATTTGGAGCCCTTTGCGGCCTATTGTGGAAAAGGAAATAACTTCACATGAAAACCTACA
CAGAAGCTGAGAACTTCTTTGTGATGTGTGCATTAATTTCCCAGAGTCGAACCTTTCTTTTGATT
GAGCAGTTTTGAAACACTCTTTTGTAGAATCTGCAAGTGGACATTTGAAGCACTTTGAGGCCTAT
TGTTGAAAAGGAAACATCTTCATATAAAAAACAACAAGGAAGCATTCTGAGAAACCATTTTGTGCTG
5 TGTGCATTCACCTCACAGAGTTCAACTTTATTTGATACAGCAGTTTTGAAACACTCTTCTTGTGGA
ATCTGCAAGTGGAAATTGGGAAATATTTAGGCATATGGTGGAAAAGGAAACATCCGCACATAAAAA
CTACACAGACACATTCTGTGAAACTTCTTTGTGCTGTGTGCATTCAAACCACAGAGTTGAACCTAT
CTTTTGAATGAGCAGTTTTGAAACTCTCTTTTCATAGTATCTGCAAGTGGATATTTGGAGCCTTTT
GTGGCCTACGGTGGGAAAGGAAATATCTTCATATAAAAACTACACAGAAGCATTCTGAGAAACTTC
10 TCAGTGATGTGAGCATTCTTCTCACAGAGTTGAACTATCTTTTGATTGAGCAGTTTTGAAACACTG
TTTTTTTAGAATCTGCAAGTGAATATTTGGAGCCTTTTGGGTCTATTGTGGAAAAGGAAATATCTT
CATATAAAAACTACACAGAAGCATTCTGAGACACTTCTTTGTGGTATGTGCATTCATCTCACAGAG
GTGAAGCTTTCTTTTGATGGAGCAGTTTTGGAAAACCTCTTTTGTAGAATCTGCAAGTGGATATTT
GGAGCACTTTGAGGCCCATTTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAATAATTC
15 GAGAAACATCTTTTTGTTGTGTGCATTCAACTCACAGAGTTGAACTATCTTTTGATTGTGCAGTTT
TGAAACACTCTTTTTGTAGAATGTGAAAGTGGATATTTTGAGCGCTTTGAGGCCTATTGTGGAAAA
AGAAATATCTTCAATTA AAAACTACACAGAAGCATTCTGAGAAACTTCTTTGTGCATGTGTGGATTC
ATCTCACAGAGTTAAATCTTTCTTTTGATTGAGCAGTTTGAAACACTCTTTTTGTGGTATCTCCAG
GAGGATATTTGGAGTGCTTTGAGGCCTATGTTGGAAAAGGAAGTATCTTCCCTTAAAGCTATGCA
20 GAAGCATTCTGAGAACTTCCTTCTGATGTGTGCATTCATCTCACCTAGTTGAACCTTTCTTTTGG
TTGTGCACTTTTGAAACACTCTTTTTGTGGAATCTGCAAGTGGATATCTGGATCACTTTGACGTCT
ATTGTGGAAAAGGAAATATCTTCACATAAAAACTACACAGAAGAATCCGACATAGTTCTTTGTGA
TGTGTGCATTCAACTCACATAGTTGAAACCATCTCTTGATCGAGTAGTTTGAACCTCTCTTGTG
TAGAATCTGAAAGTGGATATTTGTGTCCCTGGCGGTCTATGGTGGAAAAGAAATATCTTCACAAA
25 AATACTACACAGAAGCATTCTGAGAAACTTCTTTGTGATGTGTCCATTTCATCCACAGAGTTGAAC
CTTTCTTTTGATTGAGCAGTTTTGAAATACTCCTTTTGTAGAATCTGCAAGTGGATATTTTGAGTG
CTTTGAGAACTATTGTGGAAAAGGAATTATCTTCTCATAAAACCTACACTGAAGGATTCTGAGAAA
TTTCTTGTGATGTGTGCATTCATCTCACAGAGTTGAACATTTCCATGATTGAGCAGTTTGAAATA
TTCTTTTCATAGAATCTGGAAGTGGATATTTGGAGCCCTTTGAGGCCTATTGTGGAAAAGGAAATA
30 TCTTCACATAAAAACTACAGAGAAGCATTCTGAGAAACTTCTTTGTGATGTGTGCATTTCATCAAC
AGAATTGAACATTTCTTTTTTTGTGCAGTTTTGAAACAATCTTCTTGTAGTATCTGCAAGTGGATA
TTTGGAGCGTTTTAAGACCTAAGGTGGGAAAGGAAATATCTTCACATAAAAAATTACACAGAGAGAT
TCTGAGAACTTCTTTGTGATGTGTGCATTCATCTCATATATTTGAACCTTTCTTTTCATTGTGCA
GTTTCAAGCAATCTTTTTCTAGA

35 >p82H h3 (SEQ ID NO: 151)

AAGCTTTGAGACTATGGTGGAAAAGGAAATACCTTAGATAAAAAACCAGACAGAAGCATTCTCAGAA
ACTTCTTTGTGATGTTTGCATTCAACTCACAGAGTTGAATGTAACTTTTCAAGAGCATGTTTTGA
AACACTCATTTTCGTAGAATCTGCAACTGGATATTTGGACTGTGTTGAGGCCTACGGTGGAAAGGAG
ATACCTTAGATAAAAAACCAGACAGAAGCATTCTCAGAAACTTCTTTGTGATGTTTGCATTCAACTC
40 ACAGAGTTGAATGTAACTTTTCAAGAGCATGTTTTGAAACACTCATTTTCGTAGAATCTGCAACTG

FIG. 9B Continued

GATATTTGGACTGTGTTGAGGC/CTTCGTTGGAAACGGGA/ATACGGTGGAAAGGAGATATACCTT
AGATAAAAACCAGACAGAAGCATTCTCAGAACTTCTTTGTGATGTTTGCATTCAACTCACAGAGT
TGAATGTAACTTTTCAAGAGCATGTTTTGAAACACTCATTTCGTAGAATCTGCAACTGGATATTT
GGACTGTGTTGA

FIG. 10A

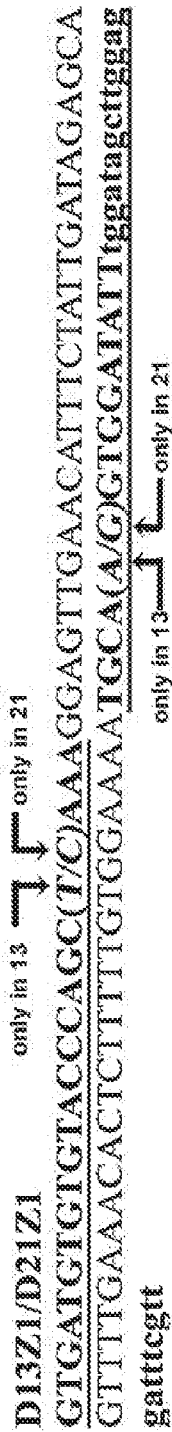


FIG. 10B

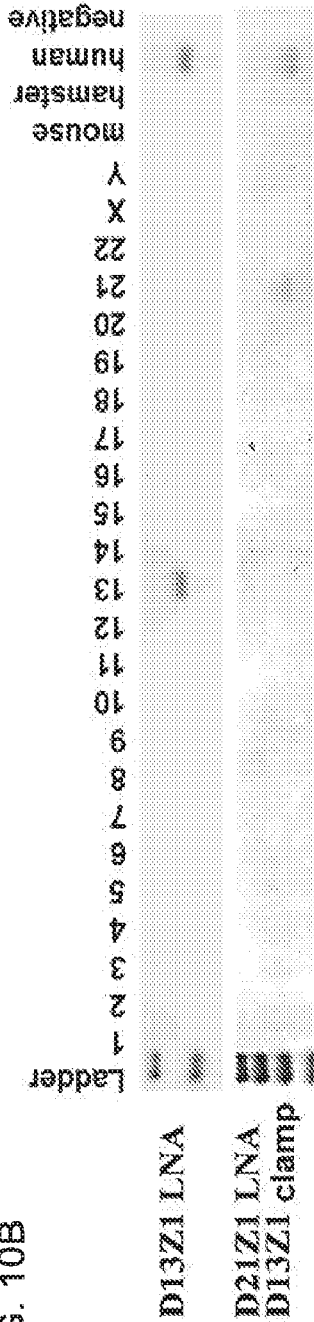


FIG. 10C

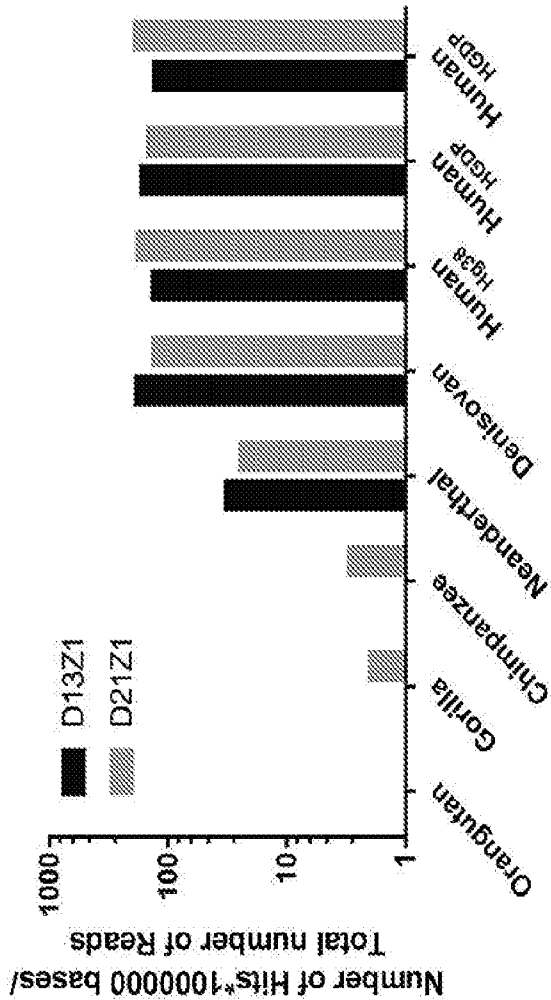
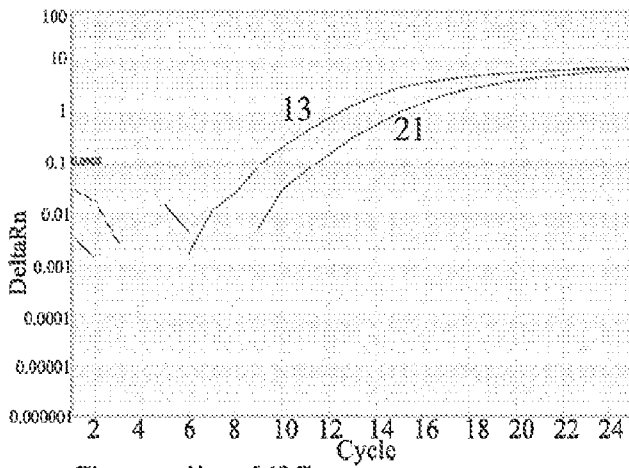
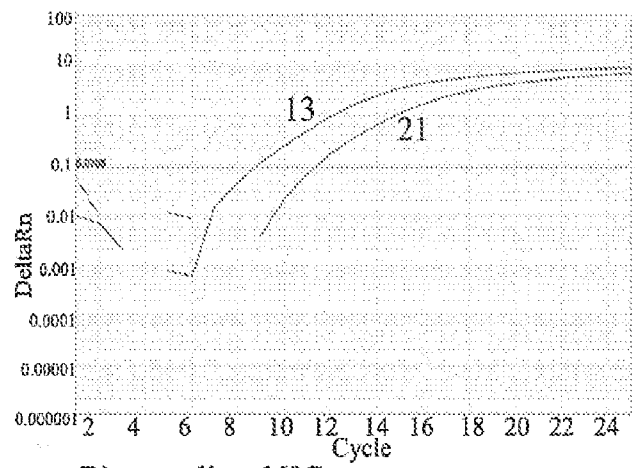


FIG. 11A

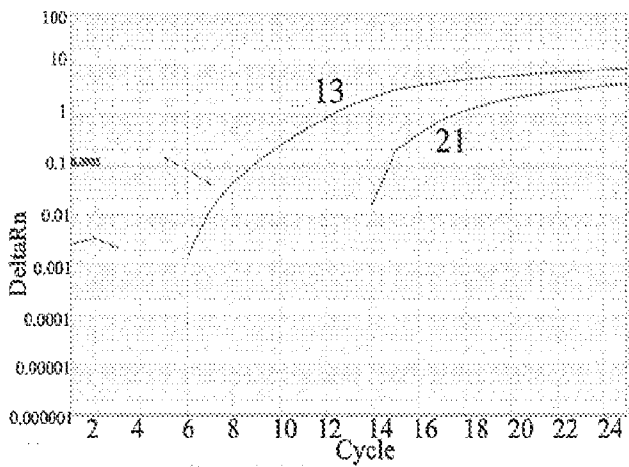
A) annealing 60°C



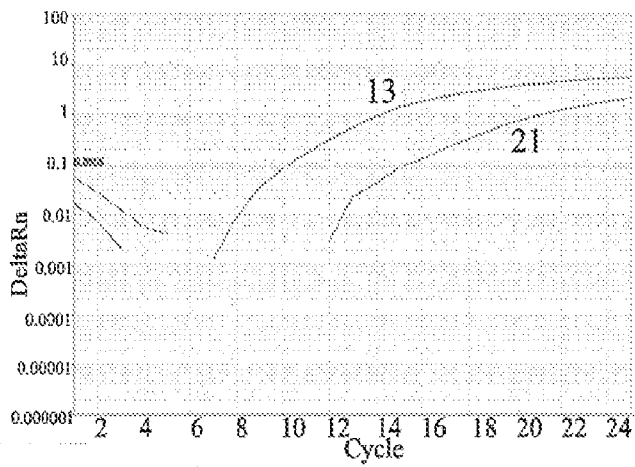
B) annealing 62°C



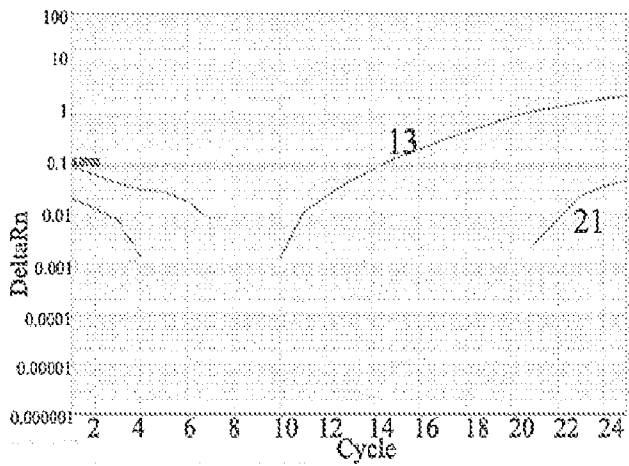
C) annealing 64°C



D) annealing 66°C



E) annealing 68°C



F) annealing 70°C

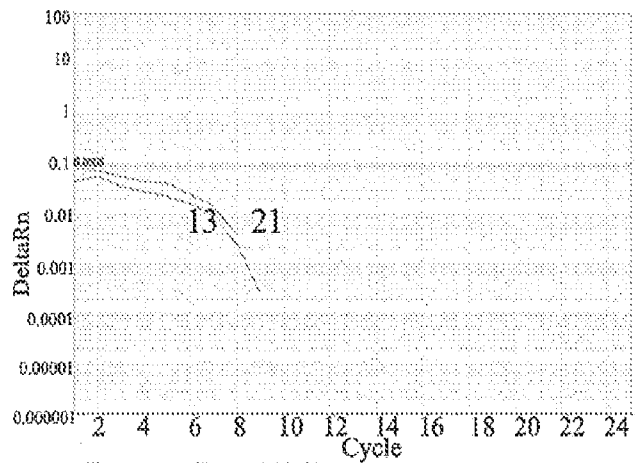
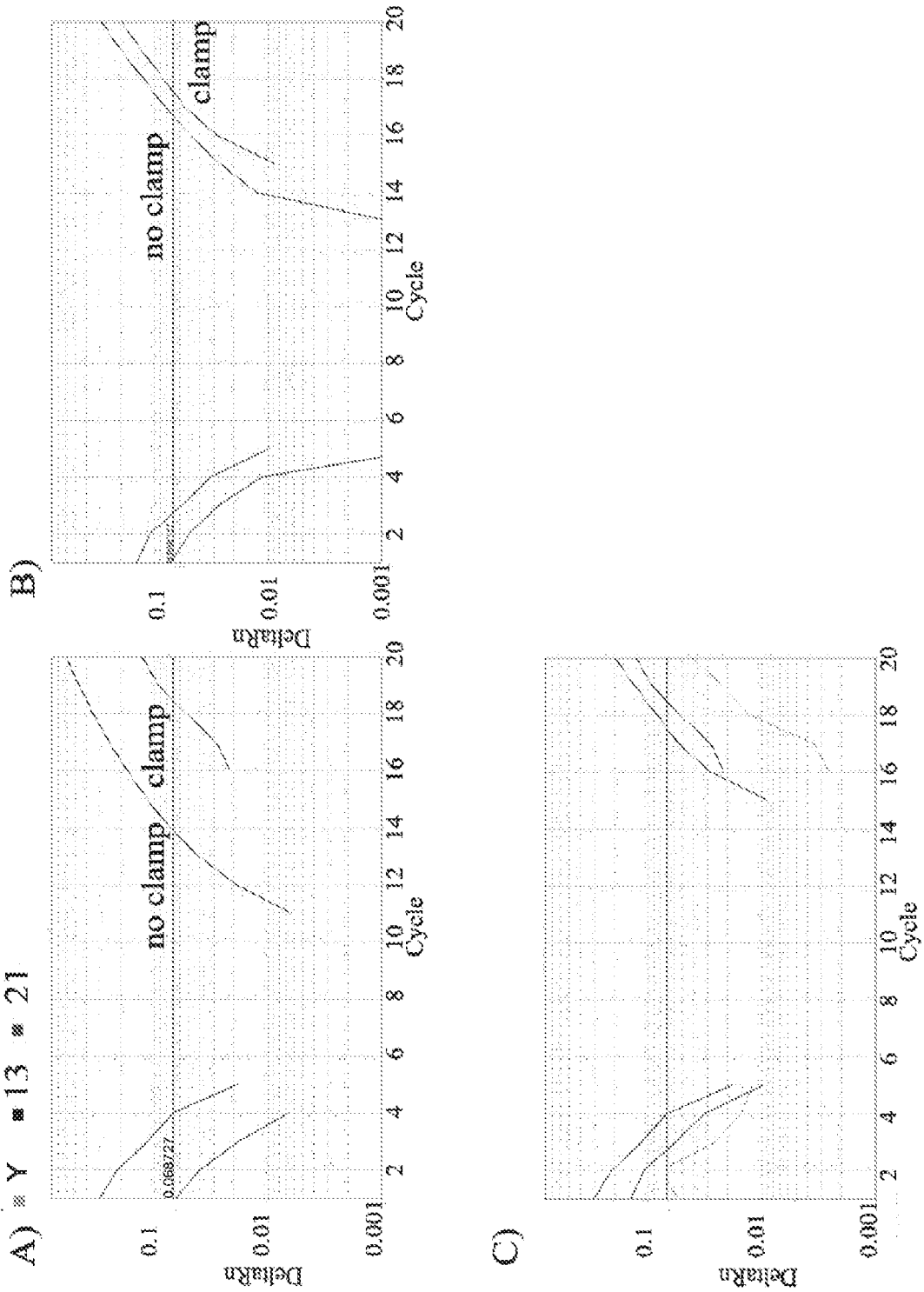
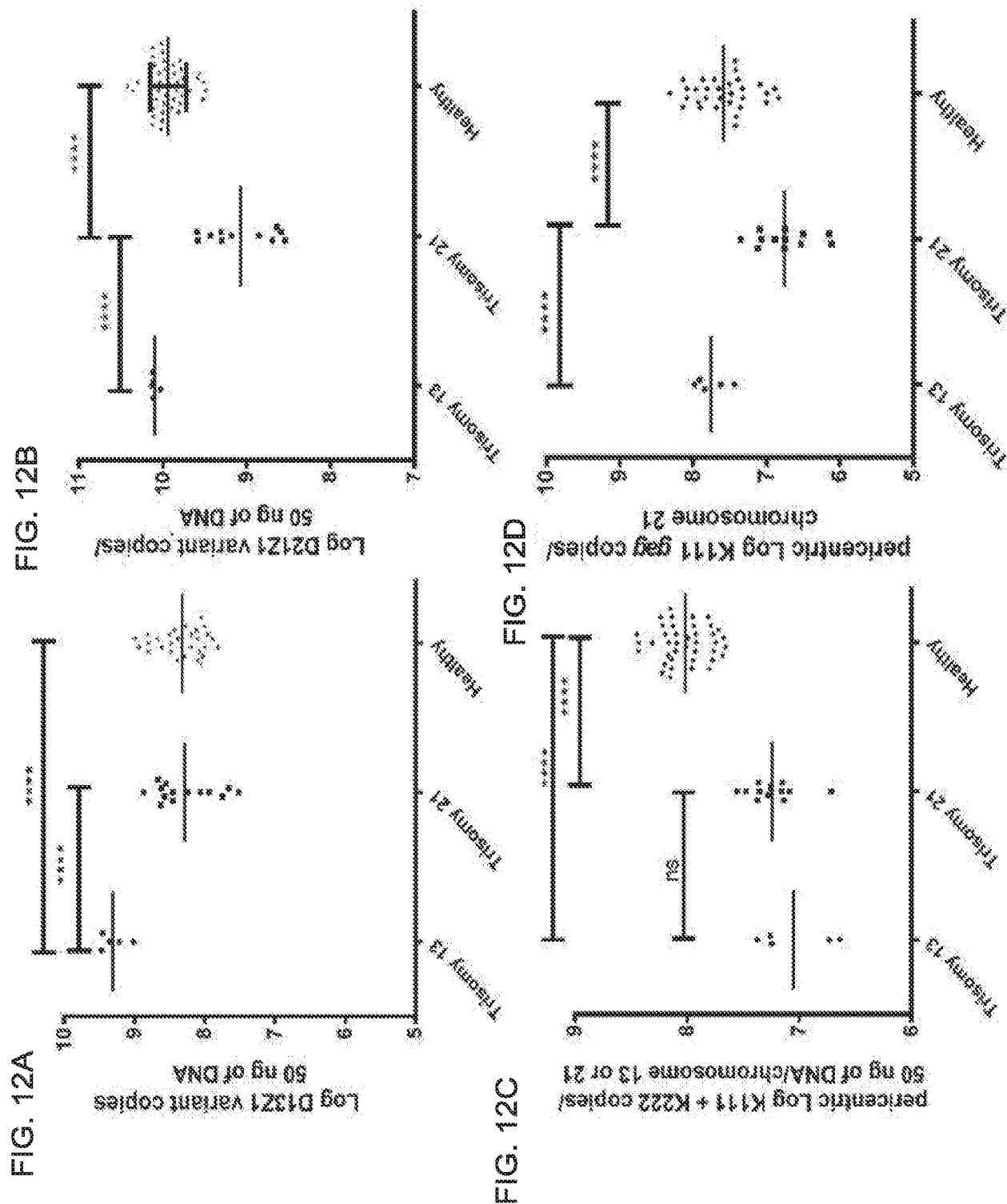


FIG. 11B





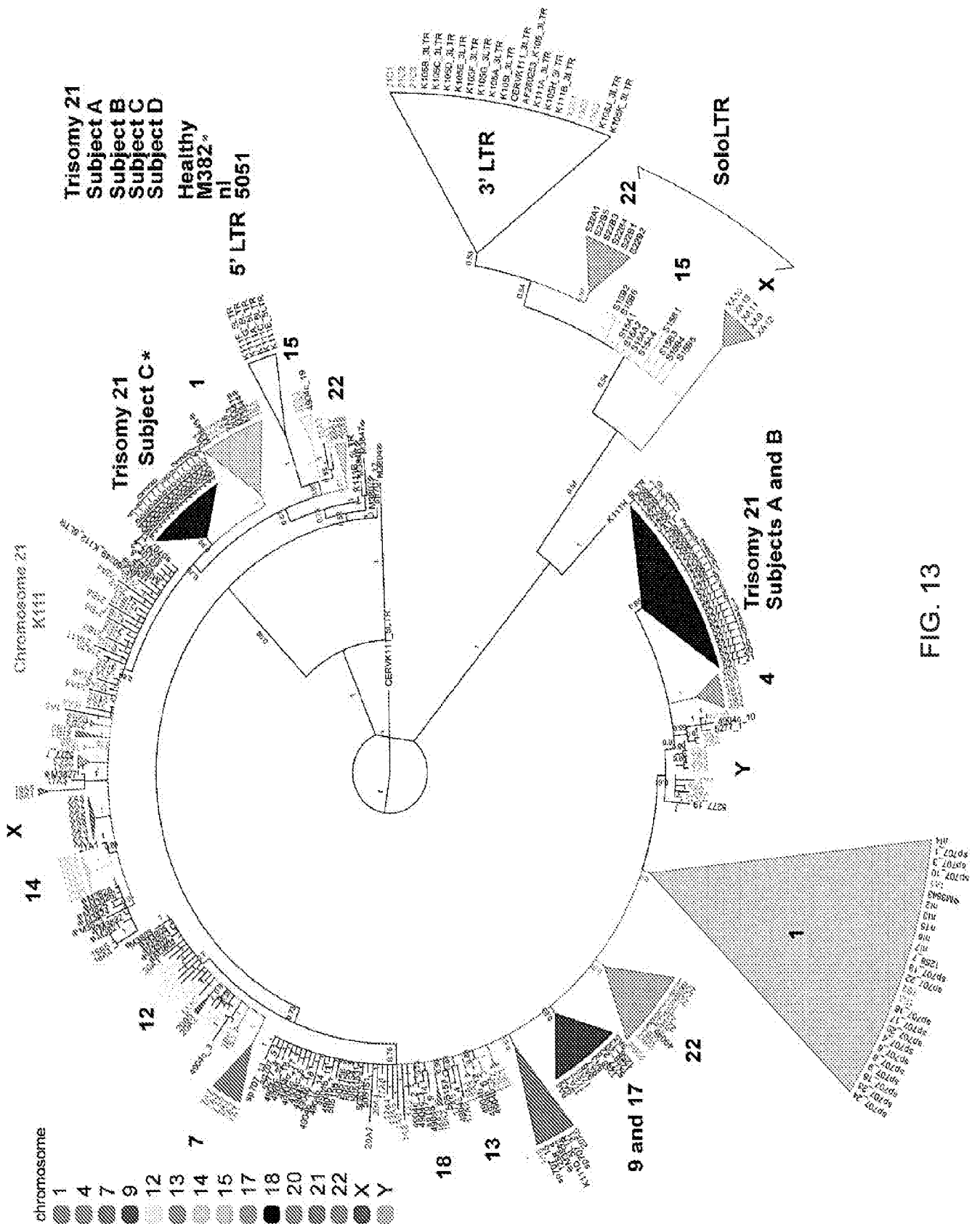


FIG. 13

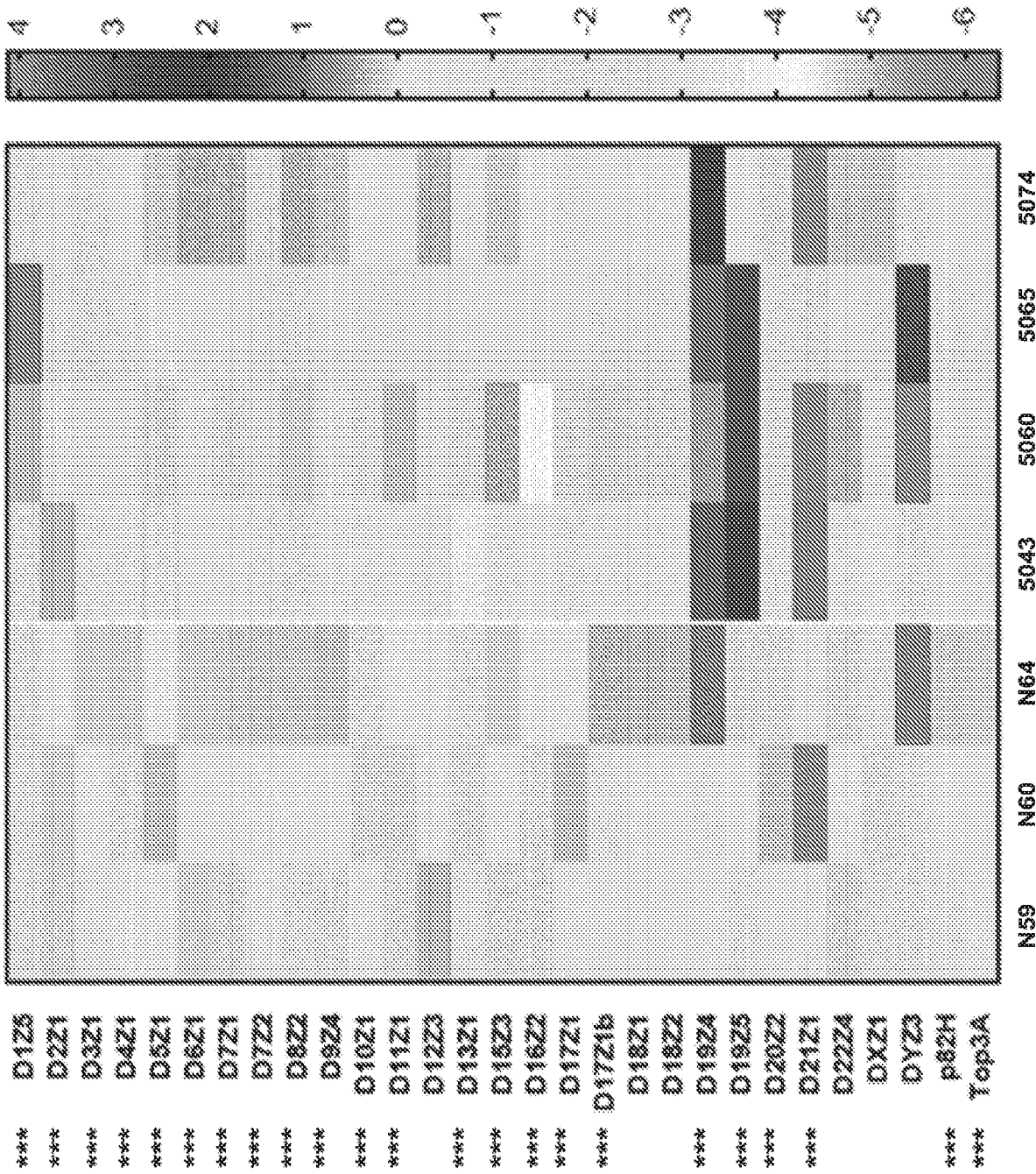


FIG. 14

FIG. 15A

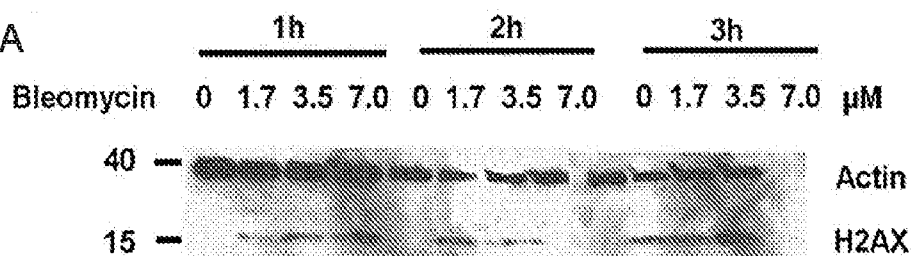
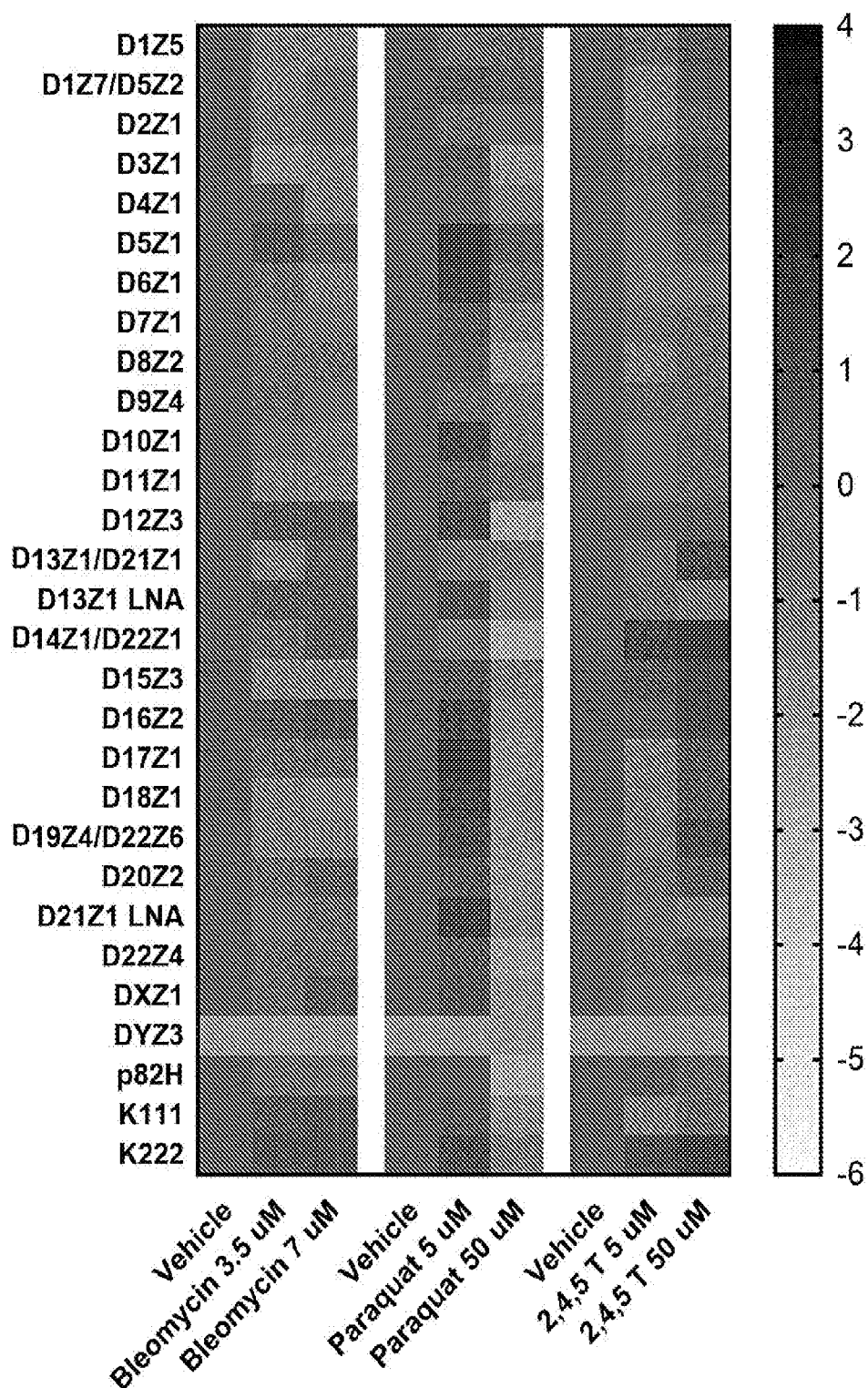


FIG. 15B



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/061543

Box No. 1 Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a. ☒ forming part of the international application as filed:

☒ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*. 1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*. 1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*. 1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/061543

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet(s).

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos. 1-6, 12-17, 20, 22, and 26-40 to the extent that they read on an α -repeat array of D1Z5, SEQ ID NO:112, SEQ ID NO:1, and SEQ ID NO:4.

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/061543

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C07H 21/04; C12N 15/09; C12N 15/10; C12Q 1/68; C12Q 1/6806; C12Q 1/6813 (2018.01)

CPC - C12Q 1/6841; C12Q 1/6851; C12Q 1/6858; C12Q 1/686; C12Q 1/6876; C12Q 1/6879; C12Q 1/6883; C12Q 2600/156 (2018.02)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 435/6; 435/810; 435/6.18; 435/91.2; 436/501; 536/22.1; 536/23.1; 536/24.3; 536/24.31; 536/24.33 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2006/047412 A2 (PROMEGA CORPORATION et al) 04 May 2006 (04.05.2006) entire document	22
X	PIRONON et al. "Molecular and evolutionary characteristics of the fraction of human alpha satellite DNA associated with CENP-A at the centromeres of chromosomes 1, 5, 19, and 21," BMC Genomics, 23 March 2010 (23.03.2010), Vol. 11, No. 195, Pgs. 1-18. entire document	29-32, 34-38
X	US 2011/0281263 A1 (MATTHIESEN et al) 17 November 2011 (17.11.2011) entire document	40
A	US 5,427,932 A (WEIER et al) 27 June 1995 (27.06.1995) entire document	1-6, 12-17, 20, 22, 26-40
A	ENGELSTEIN et al. "A PCR-Based Linkage Map of Human Chromosome 1," Genomics, 01 February 1993 (01.02.1993), Vol. 15, Iss. 2, Pgs. 251-258. entire document	1-6, 12-17, 20, 22, 26-40
A	CA 2104767 C (REYNOLDS et al) 23 June 2009 (23.06.2009) entire document	1-6, 12-17, 20, 22, 26-40
A	AURICHE et al. "Molecular and cytological analysis of a 5.5 Mb minichromosome," EMBO Reports, 01 February 2001 (01.02.2001), Vol. 2, Iss. 2, Pgs. 102-107. entire document	1-6, 12-17, 20, 22, 26-40
A	US 2009/0136924 A1 (LARIONOV et al) 28 May 2009 (28.05.2009) entire document	1-6, 12-17, 20, 22, 26-40
T	CONTRERAS-GALINDO et al. "Rapid molecular assays to study human centromere genomics," Genome Research, 15 November 2017 (15.11.2017), Vol. 27, Pgs. 2040-2049. entire document	1-6, 12-17, 20, 22, 26-40



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

14 February 2018

Date of mailing of the international search report

05 MAR 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I+: claims 1-40 are drawn to human chromosome α -repeat arrays, and methods and reaction mixtures and kits comprising the same.

The first invention of Group I+ is restricted to a human chromosome α -repeat array, and methods and reaction mixtures and kits comprising the same, wherein the α -repeat array is selected to be D1Z5 (SEQ ID NO:112); further comprising a first oligonucleotide and a second oligonucleotide that hybridizes to the α -repeat array, wherein the first oligonucleotide is selected to be SEQ ID NO:1 and the second oligonucleotide is selected to be SEQ ID NO:4. It is believed that claims 1-6, 12-17, 20, 22, and 26-40 read on this first named invention and thus these claims will be searched without fee to the extent that they read on an α -repeat array of D1Z5, SEQ ID NO:112, SEQ ID NO:1, and SEQ ID NO:4.

Applicant is invited to elect additional α -repeat arrays, each with specified SEQ ID NO, and/or pairs of oligonucleotides, each pair with specified SEQ ID NOs, to be searched in a specific combination by paying an additional fee for each set of election. An exemplary election would be a human chromosome α -repeat array, and methods and reaction mixtures and kits comprising the same, wherein the α -repeat array is selected to be D2Z1 (SEQ ID NO:114); further comprising a first oligonucleotide and a second oligonucleotide that hybridizes to the α -repeat array, wherein the first oligonucleotide is selected to be SEQ ID NO:7 and the second oligonucleotide is selected to be SEQ ID NO:8. Additional α -repeat arrays and/or pairs of oligonucleotides will be searched upon the payment of additional fees. Applicants must specify the claims that read on any additional elected inventions. Applicants must further indicate, if applicable, the claims which read on the first named invention if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined.

The inventions listed in Groups I+ do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The Groups I+ formulas do not share a significant structural element responsible for detecting a human chromosome in a sample and/or detecting an aneuploidy of a chromosome and/or determining the number of α -repeats in an α -repeat array of a chromosome, requiring the selection of alternatives for the α -repeat arrays and the first oligonucleotides and the second oligonucleotides, where "a) the human chromosome is chromosome 1 and the target α -repeat array is D1Z5, and/or D1Z7; b) the human chromosome is chromosome 2 and the target α -repeat array is D2Z1; c) the human chromosome is chromosome 3 and the target α -repeat array is D3Z1; d) the human chromosome is chromosome 4 and the target α -repeat array is D4Z1; e) the human chromosome is chromosome 5 and the target α -repeat array is D5Z1; f) the human chromosome is chromosome 6 and the target α -repeat array is D6Z1; g) the human chromosome is chromosome 7 and the target α -repeat array is D7Z1 and/or D7Z2; h) the human chromosome is chromosome 8 and the target α -repeat array is D8Z2; i) the human chromosome is chromosome 9 and the target α -repeat array is D9Z4; j) the human chromosome is chromosome 10 and the target α -repeat array is D10Z1; k) the human chromosome is chromosome 11 and the target α -repeat array is D11Z1; l) the human chromosome is chromosome 12 and the target α -repeat array is D12Z3; m) the human chromosome is chromosome 13 and the target α -repeat array is D13Z1; n) the human chromosome is chromosome 14 and the target α -repeat array is D14Z1 and/or D14Z2 and/or D14Z3; o) the human chromosome is chromosome 15 and the target α -repeat array is D15Z3; p) the human chromosome is chromosome 16 and the target α -repeat array is D16Z2; q) the human chromosome is chromosome 17 and the target α -repeat array is D17Z1 and/or D17b r) the human chromosome is chromosome 18 and the target α -repeat array is D18Z1 and/or D18Z2; s) the human chromosome is chromosome 19 and the target α -repeat array is D19Z4 and/or D19Z5; t) the human chromosome is chromosome 20 and the target α -repeat array is D20Z2; u) the human chromosome is chromosome 21 and the target α -repeat array is D21Z1; v) the human chromosome is chromosome 22 and the target α -repeat array is D22Z4, and/or D22Z5; w) the human chromosome is chromosome X and the target α -repeat array is DXZ1; and/or x) the human chromosome is chromosome Y and the target α -repeat array is DYZ3" and "a) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 1, 3, or 5 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 2, 4, or 6; b) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 7 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 8; c) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 9 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 10; d) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 11 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 12; e) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 13 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 14; f) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 15 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 16; g) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 17 or 19 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 18 or 20; h) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 21 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 22; i) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 23 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 24; j) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 25 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 26; k) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 27 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 28; l) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 29 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 30; m) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 31, 34, 36, 38, 40, 42, 44, or 46 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 32, 35, 37, 39, 41, 43, 45, or 47; n) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 48, 50, or 52 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 49, 51, or 53; o) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 54 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 55; p) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 56 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 57; q) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 58 or 60 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 59 or 61; r) the first oligonucleotide

comprises a nucleotide sequence provided by SEQ ID NO: 62 or 64 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 63 or 65; s) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 66 or 68 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 67 or 69; t) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 70 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 71; u) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 72 or 74 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 73 or 75; v) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 76, 78, or 80 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 77, 79, or 81; w) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 82 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 83; and/or x) the first oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 84 and the second oligonucleotide comprises a nucleotide sequence provided by SEQ ID NO: 85".

Additionally, even if Groups I+ were considered to share the technical features of a method for detecting a human chromosome in a sample, the method comprising: a) producing an amplicon comprising a nucleotide sequence from a target α -repeat array; and b) detecting the human chromosome in the sample by detecting the amplicon comprising the nucleotide sequence from the target α -repeat array; a method for detecting aneuploidy of a chromosome in a subject, the method comprising: a) determining a pericentromere size of a chromosome; b) comparing the pericentromere size of the chromosome to the pericentromere size for the chromosome of a normal subject; and c) identifying the chromosome as aneuploid by determining that the pericentromere size of the chromosome of the subject is smaller than the pericentromere size for the chromosome of a normal subject; a method for determining the number of α -repeats in an α -repeat array of a chromosome, the method comprising: a) hybridizing a first oligonucleotide and a second oligonucleotide to a nucleic acid comprising a target α -repeat array; b) producing an amplicon comprising a nucleotide sequence from the target α -repeat array; c) quantifying the amplicon; and d) calculating the number of α -repeats in the α -repeat array of the chromosome from the quantity of the amplicon; a reaction mixture comprising a nucleotide sequence from a target α -repeat array; a composition comprising a pair of oligonucleotides specific for α -repeat arrays; a kit comprising a first oligonucleotide and a second oligonucleotide; and a kit comprising one or more pairs of oligonucleotides specific for one or more α -repeat arrays; these shared technical features do not represent a contribution over the prior art.

Specifically, US 5,427,932 A to Weier et al. discloses a method for detecting a human chromosome in a sample (this invention can be used to determine the number of chromosomes of a specific type ... in tissue sections, Abstract; human chromosome-specific alphoid DNA, Col. 2, Lns. 30-31; test samples, Col. 18, Ln. 29), the method comprising: a) producing an amplicon comprising a nucleotide sequence from a target α -repeat array ([t]he instant invention provides a primer directed DNA amplification method using the polymerase chain reaction (PCR) ... to isolate chromosome-specific repeated DNA, Col. 2, Lns. 21-25; human chromosome-specific alphoid DNA, Col. 2, Lns. 30-31; amplifying the template DNA by a polymerase chain reaction (PCR) method between and including the first and second primers to produce chromosome-specific repeat sequence, Col. 3, Lns. 61-64; alpha satellite DNA repeat monomers flanked by the PCR primer sequences, Col. 26, Lns. 4-5); and b) detecting the human chromosome in the sample by detecting the amplicon comprising the nucleotide sequence from the target α -repeat array (amplification products ... can be screened by, for example, gel electrophoresis, Col. 4, Lns. 29-33; indicating the presence of an extra copy of chromosome 10 in those cells, Col. 5, Lns. 66-67; amplification products for several samples, Col. 25, Lns. 29-30); a method for determining the number of α -repeats in an α -repeat array of a chromosome ([a] primer directed DNA amplification method to isolate efficiently chromosome-specific repeated DNA, Abstract; [t]he size of the insert can be derived from the product size, Col. 25, Ln. 30; alpha satellite DNA repeat monomers, Col. 26, Lns. 4-5), the method comprising: a) hybridizing a first oligonucleotide and a second oligonucleotide to a nucleic acid comprising a target α -repeat array (binding a first set of oligonucleotide primers to repeat sequence units in template DNA, Col. 3, Lns. 50-51; alpha satellite DNA repeat monomers flanked by the PCR primer sequences, Col. 26, Lns. 4-5); b) producing an amplicon comprising a nucleotide sequence from the target α -repeat array (amplifying the template DNA by a polymerase chain reaction (PCR) method between and including the first and second set of primers to produce chromosome-specific repeat sequence, Col. 3, Lns. 61-64; alpha satellite DNA repeat monomers flanked by the PCR primer sequences, Col. 26, Lns. 4-5); c) quantifying the amplicon ([a]mplification of alpha satellite DNA was confirmed visually by electrophoresis, Col. 20, Lns. 48-49; the discrete size intervals found in gel electrophoretic separation. ... reflect the presence of rather large domains of alphoid DNA, Col. 12, Lns. 30-38); and d) calculating the number of α -repeats in the α -repeat array of the chromosome from the quantity of the amplicon (amplification of longer products multiple 170 bp repeat sequences may occur ... [i]n this example, the primer annealing sites are separated by one or more repeat units. ... the DNA segments amplified contain DNA segments of 175 bp size plus one or more multiples of the 171 bp repeat unit, Col. 11, Lns. 51-60); a reaction mixture comprising a nucleotide sequence from a target α -repeat array (amplification of chromosome 8 alphoid DNA ... chromosome 8 preparation ... were used as DNA template for each PCR amplification. ... reaction buffer ... Taq DNA polymerase mixed with ... the two oligonucleotide primers WA1 and WA2 ... DNA amplification was performed during 40 cycles ... [a]mplification of alpha satellite DNA fragments was confirmed by gel electrophoresis, Col. 17, Lns. 39-51); a composition comprising a pair of oligonucleotides specific for α -repeat arrays (amplification of chromosome 8 alphoid DNA ... chromosome 8 preparation ... were used as DNA template for each PCR amplification. ... reaction buffer ... Taq DNA polymerase mixed with ... the two oligonucleotide primers WA1 and WA2 ... DNA amplification was performed during 40 cycles ... [a]mplification of alpha satellite DNA fragments was confirmed by gel electrophoresis, Col. 17, Lns. 39-51; alpha satellite DNA repeat monomers flanked by the PCR primer sequences, Col. 26, Lns. 4-5).

CA 2104767 C to Reynolds et al. discloses a kit ([t]he present invention also provides kits containing reagents used in the methods of the present invention, Pg. 3, Lns. 10-11) comprising a first oligonucleotide and a second oligonucleotide ([k]its include the labeled oligonucleotide probes, Pg. 7, Lns. 17; biotinylated probes that hybridize to a human alpha satellite locus, Pg. 1, Ln. 3); and a kit ([t]he present invention also provides kits containing reagents used in the methods of the present invention, Pg. 3, Lns. 10-11) comprising one or more pairs of oligonucleotides specific for one or more α -repeat arrays (biotinylated probes that hybridize to a human alpha satellite locus, Pg. 1, Ln. 3, pairs of primers complementary to the conserved regions have been used to amplify the hypervariable regions using the PCR. These regions can be analyzed subsequently by direct DNA sequencing or by hybridization to a collection of sequence-specific oligonucleotide probes, Pg. 6, Lns. 27-30).

WO 2006/047412 A2 to Promega Corporation et al. discloses a method for detecting aneuploidy of a chromosome in a subject (methods of evaluating germ line specific genomic instability.... use in detecting germ line specific genomic instability, Pg. 2, Lns. 4-5 and 8; for determination of germ line aneuploidy, Pg. 18, Ln. 6), the method comprising: a) determining a pericentromere size of a chromosome (the first DNA sample is then contacted with a first primer and a second primer that hybridize to a first DNA sequence ... under conditions that allow amplification ... to form a first amplification product. The size of the first amplification product is determined, Pg. 2, Lns. 16-21; a region ... just below the centromere and proximal, Pg. 33, Lns. 11-16); b) comparing the pericentromere size of the chromosome to the pericentromere size for the chromosome of a normal subject ([t]he size of the first amplification product is

determined and compared to the expected size of the amplification productfrom at least one control cell, Pg. 2, Lns. 20-25; genomic instability was measured by evaluating the sizes of amplification products and deducing the presence of mutant alleles by comparing the size of the amplified product ...to that of somatic control cells ...or the expected size of the amplified product, Pg. 8, Lns. 31-34); and c) identifying the chromosome as aneuploid by determining that the pericentromere size of the chromosome of the subject is smaller than the pericentromere size for the chromosome of a normal subject ([a] difference between the size of the first amplification product and the expected size of the amplification product is indicative of genomic instability, Pg. 2, Lns. 21-23; the result can be a deletion involving a whole region ...just below the centromere and proximal, Pg. 33, Lns. 12-16).

The inventions listed in Groups I+ therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.