polynucleotides into families, each family amplified from the same tagged parent polynucleotide; and (ii) determining a consensus sequence based on sequence reads in a family.

[00125] The disclosure also provides for a system comprising a computer readable medium for performing the following steps: a) accepting at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides; b) amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides; c) sequencing a subset (including a proper subset) of the set of amplified progeny polynucleotides, to produce a set of sequencing reads; d) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides and, optionally, e) analyzing the set of consensus sequences for each set of tagged parent molecules.

[00126] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 10% of the individual's genome is sequenced.

[00127] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 20% of the individual's genome is sequenced.

[00128] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 30% of the individual's genome is sequenced.

[00129] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 40% of the individual's genome is sequenced.

[00130] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 50% of the individual's genome is sequenced.

IDT Exhibit 1020.4032

[00131] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 60% of the individual's genome is sequenced.

[00132] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 70% of the individual's genome is sequenced.

[00133] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 80% of the individual's genome is sequenced.

[00134] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 90% of the individual's genome is sequenced.

[00135] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 10% of the individual's genome is sequenced.

[00136] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 20% of the individual's genome is sequenced.

[00137] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 30% of the individual's genome is sequenced.

[00138] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 40% of the individual's genome is sequenced.

IDT Exhibit 1020.4033

[00139] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 50% of the individual's genome is sequenced.

[00140] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 60% of the individual's genome is sequenced.

[00141] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 70% of the individual's genome is sequenced.

[00142] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 80% of the individual's genome is sequenced.

[00143] The disclosure also provides for a method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 90% of the individual's genome is sequenced.

[00144] In some embodiments, the genetic alteration is copy number variation or one or more rare mutations. In some embodiments, the genetic variation comprises one or more causal variants and one or more polymorphisms. In some embodiments, the genetic alteration and/or amount of genetic variation in the individual may be compared to a genetic alteration and/or amount of genetic variation in one or more individuals with a known disease. In some embodiments, the genetic alteration and/or amount of genetic variation in the individual may be compared to a genetic alteration and/or amount of genetic variation in one or more individuals, without a disease. In some embodiments, the cell-free nucleic acid is DNA. In some embodiments, the cell-free nucleic acid is RNA. In some embodiments, the cell-free nucleic acid is DNA and RNA. In some embodiments, the disease is cancer or pre-cancer. In some embodiments, the method further comprising diagnosis or treatment of a disease.

IDT Exhibit 1020.4034

[00145] The disclosure also provides for a method comprising: a) providing at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides; b) amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides; c) sequencing a subset (including a proper subset) of the set of amplified progeny polynucleotides, to produce a set of sequencing reads; d) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides; and e) filtering out from among the consensus sequences those that fail to meet a quality threshold.

[00146] In some embodiments, the quality threshold considers a number of sequence reads from amplified progeny polynucleotides collapsed into a consensus sequence. In some embodiments, the quality threshold considers a number of sequence reads from amplified progeny polynucleotides collapsed into a consensus sequence.

[00147] The disclosure also provides for a system comprising a computer readable medium for performing the methods described herein.

[00148] The disclosure also provides for a method comprising: a) providing at least one set of tagged parent polynucleotides, wherein each set maps to a different mappable position in a reference sequence in one or more genomes, and, for each set of tagged parent polynucleotides; i) amplifying the first polynucleotides to produce a set of amplified polynucleotides; ii) sequencing a subset of the set of amplified polynucleotides, to produce a set of sequencing reads; and iii) collapsing the sequence reads by: (1)grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide.

[00149] In some embodiments, collapsing further comprises determining a quantitative measure of sequence reads in each family. In some embodiments, the method further comprises: a) determining a quantitative measure of unique families; and b) based on (1) the quantitative measure of unique families and (2) the quantitative measure of sequence reads in each group, inferring a measure of unique tagged parent polynucleotides in the set. In some embodiments, inferring is performed using statistical or probabilistic models. In some embodiments, the at least one set is a plurality of sets. In some embodiments, the method further comprises correcting for amplification or representational bias between the two sets. In some embodiments, the method further comprises using a control or set of control samples to correct for amplification or representational biases between the two sets. In some embodiments, the method further comprises determining copy number variation between the sets.

IDT Exhibit 1020.4035

[00150] In some embodiments, the method further comprises: d) determining a quantitative measure of polymorphic forms among the families; and e) based on the determined quantitative measure of polymorphic forms, inferring a quantitative measure of polymorphic forms in the number of inferred unique tagged parent polynucleotides. In some embodiments, polymorphic forms include but are not limited to: substitutions, insertions, deletions, inversions, microsatellite changes, transversions, translocations, fusions, methylation, hypermethylation, hyrdroxymethylation, acetylation, epigenetic variants, regulatory-associated variants or protein binding sites.

[00151] In some embodiments, the sets derive from a common sample, and the method further comprises: d) inferring copy number variation for the plurality of sets based on a comparison of the inferred number of tagged parent polynucleotides in each set mapping to each of a plurality of mappable positions in a reference sequence. In some embodiments, the original number of polynucleotides in each set is further inferred. In some embodiments, at least a subset of the tagged parent polynucleotides in each set are non-uniquely tagged.

[00152] The disclosure also provides for a method of determining copy number variation in a sample that includes polynucleotides, the method comprising: a) providing at least two sets of first polynucleotides, wherein each set maps to a different mappable position in a reference sequence in a genome, and, for each set of first polynucleotides; (i) amplifying the polynucleotides to produce a set of amplified polynucleotides; (ii) sequencing a subset of the set of amplified polynucleotides, to produce a set of sequencing reads; (iii) grouping sequences reads sequenced from amplified polynucleotides into families, each family amplified from the same first polynucleotide in the set; (iv) inferring a quantitative measure of families in the set; and b) determining copy number variation by comparing the quantitative measure of families in each set.

[00153] The disclosure also provides for a method of inferring frequency of sequence calls in a sample of polynucleotides comprising: a) providing at least one set of first polynucleotides, wherein each set maps to a different mappable position in a reference sequence in one or more genomes, and, for each set of first polynucleotides; (i) amplifying the first polynucleotides to produce a set of amplified polynucleotides; (ii) sequencing a subset of the set of amplified polynucleotides, to produce a set of sequencing reads; (iii) grouping the sequence reads into families, each family comprising sequence reads of amplified polynucleotides amplified from the same first polynucleotide; b) inferring, for each set of first polynucleotides, a call frequency for one or more bases in the set of first polynucleotides, wherein inferring comprises: (i) assigning,

IDT Exhibit 1020.4036

for each family, confidence score for each of a plurality of calls, the confidence score taking into consideration a frequency of the call among members of the family; and (ii) estimating a frequency of the one or more calls taking into consideration the confidence scores of the one or more calls assigned to each family.

[00154] The disclosure also provides for a method of communicating sequence information about at least one individual polynucleotide molecule, comprising: a) providing at least one individual polynucleotide molecule; b) encoding sequence information in the at least one individual polynucleotide molecule to produce a signal; c) passing at least part of the signal through a channel to produce a received signal comprising nucleotide sequence information about the at least one individual polynucleotide molecule, wherein the received signal comprises noise and/or distortion; d) decoding the received signal to produce a message comprising sequence information about the at least one individual polynucleotide molecule, wherein decoding reduces noise and/or distortion about each individual polynucleotide in the message; and e) providing the message comprising sequence information about the at least one individual polynucleotide molecule to a recipient.

[00155] In some embodiments, the noise comprises incorrect nucleotide calls. In some embodiments, distortion comprises uneven amplification of the individual polynucleotide molecule compared with other individual polynucleotide molecules. In some embodiments, distortion results from amplification or sequencing bias. In some embodiments, the at least one individual polynucleotide molecule is a plurality of individual polynucleotide molecules, and decoding produces a message about each molecule in the plurality. In some embodiments, encoding comprises amplifying the at least one individual polynucleotide molecule, which has optionally been tagged, wherein the signal comprises a collection of amplified molecules. In some embodiments, the channel comprises a polynucleotide sequencer and the received signal comprises sequence reads of a plurality of polynucleotides amplified from the at least one individual polynucleotide molecule. In some embodiments, decoding comprises grouping sequence reads of amplified molecules amplified from each of the at least one individual polynucleotide molecules. In some embodiments, the decoding consists of a probabilistic or statistical method of filtering the generated sequence signal.

[00156] In some embodiments, the polynucleotides are derived from tumor genomic DNA or RNA. In some embodiments, the polynucleotides are derived from cell-free polynucleotides, exosomal polynucleotides, bacterial polynucleotides or viral polynucleotides. In some

IDT Exhibit 1020.4037

embodiments of any of the methods herein, the method further comprises the detection and/or association of affected molecular pathways. In some embodiments of any of the methods herein, the method further comprises serial monitoring of the health or disease state of an individual. In some embodiments the phylogeny of a genome associated with a disease within an individual is inferred. In some embodiments, any of the methods described herein further comprise diagnosis, monitoring or treatment of a disease. In some embodiments, the treatment regimen is selected or modified based on detected polymorphic forms or CNVs or associated pathways. In some embodiments, the treatment comprises of a combination therapy. In some embodiments, the diagnosis further comprises localizing the disease using a radiographic technique, such as, a CT-Scan, PET-CT, MRI, Ultrasound, Ultraound with microbubbles, etc.

[00157] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: selecting predefined regions in a genome; accessing sequence reads and enumerating number of sequence reads in the predefined regions; normalizing the number of sequence reads across the predefined regions; and determining percent of copy number variation in the predefined regions.

[00158] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: accessing a data file comprising a plurality of sequencing reads; filtering out reads that fail to meet a set threshold; mapping sequence reads derived from the sequencing onto a reference sequence; identifying a subset of mapped sequence reads that align with a variant of the reference sequence at each mappable base position; for each mappable base position, calculating a ratio of (a) a number of mapped sequence reads that include a variant as compared to the reference sequence, to (b) a number of total sequence reads for each mappable base position; normalizing the ratios or frequency of variance for each mappable base position and determining potential rare variant(s) or other genetic alteration(s); and comparing the resulting number for each of the regions with potential rare variant(s) or mutation(s) to similarly derived numbers from a reference sample.

[00159] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at

IDT Exhibit 1020.4038

least one set of tagged parent polynucleotides; and b) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides.

[00160] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; b) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides; and c) filtering out from among the consensus sequences those that fail to meet a quality threshold.

[00161] A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; and i) collapsing the sequence reads by: (1) grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide and, optionally, (2) determining a quantitative measure of sequence reads in each family.

[00162] In some embodiments, the executable code, upon execution by a computer processor, further performs the steps of: b) determining a quantitative measure of unique families; and c) based on (1) the quantitative measure of unique families and (2) the quantitative measure of sequence reads in each group, inferring a measure of unique tagged parent polynucleotides in the set.

[00163] In some embodiments, the executable code, upon execution by a computer processor, further performs the steps of: d) determining a quantitative measure of polymorphic forms among the families; and e) based on the determined quantitative measure of polymorphic forms, inferring a quantitative measure of polymorphic forms in the number of inferred unique tagged parent polynucleotides.

[00164] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing

IDT Exhibit 1020.4039

reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides grouping sequences reads sequenced from amplified polynucleotides into families, each family amplified from the same first polynucleotide in the set; b) inferring a quantitative measure of families in the set; and c) determining copy number variation by comparing the quantitative measure of families in each set.

[00165] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides grouping the sequence reads into families, each family comprising sequence reads of amplified polynucleotides amplified from the same first polynucleotide; b) inferring, for each set of first polynucleotides, a call frequency for one or more bases in the set of first polynucleotides, wherein inferring comprises: c) assigning, for each family, confidence score for each of a plurality of calls, the confidence score taking into consideration a frequency of the call among members of the family; and d) estimating a frequency of the one or more calls taking into consideration the confidence scores of the one or more calls assigned to each family.

[00166] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a received signal that comprises endoded sequence information from at least one individual polynucleotide molecule wherein the received signal comprises noise and/or distortion; b) decoding the received signal to produce a message comprising sequence information about the at least one individual polynucleotide molecule, wherein decoding reduces noise and/or distortion about each individual polynucleotide in the message; and c) writing the message comprising sequence information about the at least one individual polynucleotide molecule to a computer file.

[00167] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; b) collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique

IDT Exhibit 1020.4040

polynucleotide among the set of tagged parent polynucleotides; and c) filtering out from among the consensus sequences those that fail to meet a quality threshold.

[00168] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; and b) collapsing the sequence reads by: (i) grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide; and (ii) optionally, determining a quantitative measure of sequence reads in each family.

[00169] In some embodiments, the executable code, upon execution by a computer processor, further performs the steps of: d) determining a quantitative measure of unique families;e) based on (1) the quantitative measure of unique families and (2) the quantitative measure of sequence reads in each group, inferring a measure of unique tagged parent polynucleotides in the set.

[00170] In some embodiments, the executable code, upon execution by a computer processor, further performs the steps of: e) determining a quantitative measure of polymorphic forms among the families; and f) based on the determined quantitative measure of polymorphic forms, inferring a quantitative measure of polymorphic forms in the number of inferred unique tagged parent polynucleotides.

[00171] In some embodiments, the the executable code, upon execution by a computer processor, further performs the steps of: e) inferring copy number variation for the plurality of sets based on a comparison of the inferred number of tagged parent polynucleotides in each set mapping to each of a plurality of reference sequences.

[00172] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: a) accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; b) grouping sequences reads sequenced from amplified polynucleotides into families, each family amplified from the same first polynucleotide in the set; c) inferring a quantitative measure of families in the set; d) determining copy number variation by comparing the quantitative measure of families in each set.

IDT Exhibit 1020.4041

[00173] The disclosure also provides for a computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising: accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides grouping the sequence reads into families, each family comprising sequence reads of amplified polynucleotides amplified from the same first polynucleotide; and inferring, for each set of first polynucleotides, a call frequency for one or more bases in the set of first polynucleotides, wherein inferring comprises: (i) assigning, for each family, confidence score for each of a plurality of calls, the confidence score taking into consideration a frequency of the call among members of the family; and (ii) estimating a frequency of the one or more calls taking into consideration the confidence scores of the one or more calls assigned to each family.

[00174] The disclosure also provides for a composition comprising between 100 and 100,000 human haploid genome equivalents of cfDNA polynucleotides, wherein the polynucleotides are tagged with between 2 and 1,000,000 unique identifiers.

[00175] In some embodiments, the composition comprises between 1000 and 50,000 haploid human genome equivalents of cfDNA polynucleotides, wherein the polynucleotides are tagged with between 2 and 1,000 unique identifiers. In some embodiments, the unique identifiers comprise nucleotide barcodes. The disclosure also provides for a method comprising: a) providing a sample comprising between 100 and 100,000 haploid human genome equivalents of cfDNA polynucleotides; and b) tagging the polynucleotides with between 2 and 1,000,000 unique identifiers.

[00176] The disclosure also provides for a method comprising: a) providing a sample comprising a plurality of human haploid genome equivalents of fragmented polynucleotides; b) determining z, wherein z is a measure of central tendency (e.g., mean, median or mode) of expected number of duplicate polynucleotides starting at any position in the genome, wherein duplicate polynucleotides have the same start and stop positions; and c) tagging polynucleotides in sample with n unique identifiers, wherein n is between 2 and 100,000*z, 2 and 10,000*z, 2 and 1,000*z or 2 and 100*z. The disclosure also provides for a method comprising: a) providing at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides; b) producing a plurality of sequence reads for each tagged parent polynucleotide in the set to produce a set of sequencing reads; and c) collapsing the set of sequencing reads to generate a set

IDT Exhibit 1020.4042

of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides.

[00177] The disclosure also provides for a system comprising a computer readable medium comprising machine-executable code as described herein. The disclosure also provides for a system comprising a computer readable medium comprising machine-executable code that, upon execution by a computer processor, implements a method as described herein.

[00178] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[00179] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[00180] The novel features of a system and methods of this disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of this disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of a systems and methods of this disclosure are utilized, and the accompanying drawings of which:

[00181] **FIG. 1** is a flow chart representation of a method of detection of copy number variation using a single sample.

[00182] **FIG. 2** is a flow chart representation of a method of detection of copy number variation using paired samples.

[00183] **FIG. 3** is a flow chart representation of a method of detection of rare mutations (e.g., single nucleotide variants).

[00184] **FIG. 4A** is graphical copy number variation detection report generated from a normal, non cancerous subject.

IDT Exhibit 1020.4043

[00185] **FIG. 4B** is a graphical copy number variation detection report generated from a subject with prostate cancer.

[00186] **FIG. 4C** is schematic representation of internet enabled access of reports generated from copy number variation analysis of a subject with prostate cancer.

[00187] **FIG. 5A** is a graphical copy number variation detection report generated from a subject with prostate cancer remission.

[00188] **FIG. 5B** is a graphical copy number variation detection report generated from a subject with prostate recurrence cancer.

[00189] **FIG. 6A** is graphical detection report (e.g., for single nucleotide variants) generated from various mixing experiments using DNA samples containing both wildtype and mutant copies of MET and TP53.

[00190] **FIG. 6B** is logarithmic graphical representation of (e.g., single nucleotide variant) detection results. Observed vs. expected percent cancer measurements are shown for various mixing experiments using DNAs samples containing both wildtype and mutant copies of MET, HRAS and TP53.

[00191] **FIG. 7A** is graphical report of percentage of two (e.g., single nucleotide variants) in two genes, PIK3CA and TP53, in a subject with prostate cancer as compared to a reference (control).

[00192] **FIG. 7B** is schematic representation of internet enabled access of reports generated from (e.g., single nucleotide variant) analysis of a subject with prostate cancer.

[00193] **FIG. 8** is a flow chart representation of a method of analyzing genetic material.

[00194] **FIG. 9** is a flow chart representation of a method of decoding information in a set of sequence reads to produce, with reduced noise and/or distortion, a representation of information in a set of tagged parent polynucleotides.

[00195] **FIG. 10** is a flow chart representation of a method of reducing distortion in the determination of CNV from a set of sequence reads.

[00196] **FIG. 11** is a flow chart representation of a method of estimating frequency of a base or sequence of bases at a locus in a tagged parent polynucleotide population from a set of sequence reads.

[00197] **FIG. 12** shows a method of communicating sequence information.

[00198] **FIG. 13** shows detected minor allele frequencies across an entire 70kb panel in 0.3% LNCaP cfDNA titration using standard sequencing and Digital Sequencing workflows.

IDT Exhibit 1020.4044

Standard "analog" sequencing (Fig. 13A) masks all true-positive rare variants in tremendous noise due to PCR and sequencing errors despite Q30 filtering. Digital Sequencing (Fig. 13B) eliminates all PCR and sequencing noise, revealing true mutations with no false positives: green circles are SNP points in normal cfDNA and red circles are detected LNCaP mutations.

[00199] **FIG. 14:** Shows titration of LNCap cfDNA.

[00200] **FIG. 15** shows a computer system that is programmed or otherwise configured to implement various methods of the present disclosure.

DETAILED DESCRIPTION OF THE INVENTION

I. General Overview

[00201] The present disclosure provides a system and method for the detection of rare mutations (e.g., single or multiple nucleotide variations) and copy number variations in cell free polynucleotides. Generally, the systems and methods comprise sample preparation, or the extraction and isolation of cell free polynucleotide sequences from a bodily fluid; subsequent sequencing of cell free polynucleotides by techniques known in the art; and application of bioinformatics tools to detect rare mutations and copy number variations as compared to a reference. The systems and methods also may contain a database or collection of different rare mutations or copy number variation profiles of different diseases, to be used as additional references in aiding detection of rare mutations (e.g., single nucleotide variation profiling), copy number variation profiling or general genetic profiling of a disease.

[00202] The systems and methods may be particularly useful in the analysis of cell free DNAs. In some cases, cell free DNA are extracted and isolated from a readily accessible bodily fluid such as blood. For example, cell free DNA can be extracted using a variety of methods known in the art, including but not limited to isopropanol precipitation and/or silica based purification. Cell free DNA may be extracted from any number of subjects, such as subjects without cancer, subjects at risk for cancer, or subjects known to have cancer (e.g. through other means).

[00203] Following the isolation/extraction step, any of a number of different sequencing operations may be performed on the cell free polynucleotide sample. Samples may be processed before sequencing with one or more reagents (e.g., enzymes, unique identifiers (e.g., barcodes), probes, etc.). In some cases if the sample is processed with a unique identifier such as a barcode, the samples or fragments of samples may be tagged individually or in subgroups with the unique

IDT Exhibit 1020.4045

identifier. The tagged sample may then be used in a downstream application such as a sequencing reaction by which individual molecules may be tracked to parent molecules.

[00204] After sequencing data of cell free polynucleotide sequences is collected, one or more bioinformatics processes may be applied to the sequence data to detect genetic features or aberrations such as copy number variation, rare mutations (e.g., single or multiple nucleotide variations) or changes in epigenetic markers, including but not limited to methylation profiles. In some cases, in which copy number variation analysis is desired, sequence data may be: 1) aligned with a reference genome; 2) filtered and mapped; 3) partitioned into windows or bins of sequence; 4) coverage reads counted for each window; 5) coverage reads can then be normalized using a stochastic or statistical modeling algorithm; 6) and an output file can be generated reflecting discrete copy number states at various positions in the genome. In other cases, in which rare mutation analysis is desired, sequence data may be 1) aligned with a reference genome; 2) filtered and mapped; 3) frequency of variant bases calculated based on coverage reads for that specific base; 4) variant base frequency normalized using a stochastic, statistical or probabilistic modeling algorithm; 5) and an output file can be generated reflecting mutation states at various positions in the genome.

[00205] A variety of different reactions and/operations may occur within the systems and methods disclosed herein, including but not limited to: nucleic acid sequencing, nucleic acid quantification, sequencing optimization, detecting gene expression, quantifying gene expression, genomic profiling, cancer profiling, or analysis of expressed markers. Moreover, the systems and methods have numerous medical applications. For example, it may be used for the identification, detection, diagnosis, treatment, staging of, or risk prediction of various genetic and non-genetic diseases and disorders including cancer. It may be used to assess subject response to different treatments of said genetic and non-genetic diseases, or provide information regarding disease progression and prognosis.

[00206] Polynucleotide sequencing can be compared with a problem in communication theory. An initial individual polynucleotide or ensemble of polynucleotides is thought of as an original message. Tagging and/or amplifying can be thought of as encoding the original message into a signal. Sequencing can be thought of as communication channel. The output of a sequencer, e.g., sequence reads, can be thought of as a received signal. Bioinformatic processing can be thought of as a receiver that decodes the received signal to produce a transmitted message, e.g., a nucleotide sequence or sequences. The received signal can include artifacts, such as noise

IDT Exhibit 1020.4046

and distortion. Noise can be thought of as an unwanted random addition to a signal. Distortion can be thought of as an alteration in the amplitude of a signal or portion of a signal.

[00207] Noise can be introduced through errors in copying and/or reading a polynucleotide. For example, in a sequencing process a single polynucleotide can first be subject to amplification. Amplification can introduce errors, so that a subset of the amplified polynucleotides may contain, at a particular locus, a base that is not the same as the original base at that locus. Furthermore, in the reading process a base at any particular locus may be read incorrectly. As a consequence, the collection of sequence reads can include a certain percentage of base calls at a locus that are not the same as the original base. In typical sequencing technologies this error rate can be in the single digits, e.g., 2%-3%. When a collection of molecules that are all presumed to have the same sequence are sequenced, this noise is sufficiently small that one can identify the original base with high reliability.

[00208] However, if a collection of parent polynucleotides includes a subset of polynucleotides having sequence variants at a particular locus, noise can be a significant problem. This can be the case, for example, when cell free DNA includes not only germline DNA, but DNA from another source, such as fetal DNA or DNA from a cancer cell. In this case, if the frequency of molecules with sequence variants is in the same range as the frequency of errors introduced by the sequencing process, then true sequence variants may not be distinguishable from noise. This could interfere, for example, with detecting sequence variants in a sample.

[00209] Distortion can be manifested in the sequencing process as a difference in signal strength, e.g., total number of sequence reads, produced by molecules in a parent population at the same frequency. Distortion can be introduced, for example, through amplification bias, GC bias, or sequencing bias. This could interfere with detecting copy number variation in a sample. GC bias results in the uneven representation of areas rich or poor in GC content in the sequence reading.

[00210] This invention provides methods of reducing sequencing artifacts, such as noise and/or distortion, in a polynucleotide sequencing process. Grouping sequence reads into families derived from original individual molecules can reduce noise and/or distortion from a single individual molecule or from an ensemble of molecules. With respect to a single molecule, grouping reads into a family reduces distortion by, for example, indicating that many sequence reads actually represent a single molecule rather than many different molecules. Collapsing sequence reads into a consensus sequence is one way to reduce noise in the received message

IDT Exhibit 1020.4047

from one molecule. Using probabilistic functions that convert received frequencies is another way. With respect to an ensemble of molecules, grouping reads into families and determining a quantitative measure of the families reduces distortion, for example, in the quantity of molecules at each of a plurality of different loci. Again, collapsing sequence reads of different families into consensus sequences eliminate errors introduced by amplification and/or sequencing error. Furthermore, determining frequencies of base calls based on probabilities derived from family information also reduces noise in the received message from an ensemble of molecules.

[00211] Methods of reducing noise and/or distortion from a sequencing process are known. These include, for example, filtering sequences, e.g., requiring them to meet a quality threshold, or reducing GC bias. Such methods typically are performed on the collection of sequence reads that are the output of a sequencer, and can be performed sequence read-by-sequence read, without regard for family structure (sub-collections of sequences derived from a single original parent molecule). Certain methods of this invention reduce noise and distortion by reducing noise and/or distortion within families of sequence reads, that is, operating on sequence reads grouped into families derived from a single parent polynucleotide molecule. Signal artifact reduction at the family level can produce significantly less noise and distortion in the ultimate message that is provided than artifact reduction performed at a sequence read-by-sequence read level or on sequencer output as a whole.

[00212] The present disclosure further provides methods and systems for detecting with high sensitivity genetic variation in a sample of initial genetic material. The methods involve using one or both of the following tools: First, the efficient conversion of individual polynucleotides in a sample of initial genetic material into sequence-ready tagged parent polynucleotides, so as to increase the probability that individual polynucleotides in a sample of initial genetic material will be represented in a sequence-ready sample. This can produce sequence information about more polynucleotides in the initial sample. Second, high yield generation of consensus sequences for tagged parent polynucleotides by high rate sampling of progeny polynucleotides amplified from the tagged parent polynucleotides, and collapsing of generated sequence reads into consensus sequences representing sequences of parent tagged polynucleotides. This can reduce noise introduced by amplification bias and/or sequencing errors, and can increase sensitivity of detection. Collapsing is performed on a plurality of sequence reads, generated either from reads of amplified molecules, or multiple reads of a single molecule.

IDT Exhibit 1020.4048

[00213] Sequencing methods typically involve sample preparation, sequencing of polynucleotides in the prepared sample to produce sequence reads and bioinformatic manipulation of the sequence reads to produce quantitative and/or qualitative genetic information about the sample. Sample preparation typically involves converting polynucleotides in a sample into a form compatible with the sequencing platform used. This conversion can involve tagging polynucleotides. In certain embodiments of this invention the tags comprise polynucleotide sequence tags. Conversion methodologies used in sequencing may not be 100% efficient. For example, it is not uncommon to convert polynucleotides in a sample with a conversion efficiency of about 1-5%, that is, about 1-5% of the polynucleotides in a sample are converted into tagged polynucleotides. Polynucleotides that are not converted into tagged molecules are not represented in a tagged library for sequencing. Accordingly, polynucleotides having genetic variants represented at low frequency in the initial genetic material may not be represented in the tagged library and, therefore may not be sequenced or detected. By increasing conversion efficiency, the probability that a rare polynucleotide in the initial genetic material will be represented in the tagged library and, consequently, detected by sequencing is increased. Furthermore, rather than directly address the low conversion efficiency issue of library preparation, most protocols to date call for greater than 1 microgram of DNA as input material. However, when input sample material is limited or detection of polynucleotides with low representation is desired, high conversion efficiency can efficiently sequence the sample and/or to adequately detect such polynucleotides.

[00214] This disclosure provides methods of converting initial polynucleotides into tagged polynucleotides with a conversion efficiency of at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 80% or at least 90%. The methods involve, for example, using any of blunt-end ligation, sticky end ligation, molecular inversion probes, PCR, ligation-based PCR, multiplex PCR, single strand ligation and single strand circularization. The methods can also involve limiting the amount of initial genetic material. For example, the amount of initial genetic material can be less than 1 ug, less than 100 ng or less than 10 ng. These methods are described in more detail herein.

[00215] Obtaining accurate quantitative and qualitative information about polynucleotides in a tagged library can result in a more sensitive characterization of the initial genetic material. Typically, polynucleotides in a tagged library are amplified and the resulting amplified molecules are sequenced. Depending on the throughput of the sequencing platform used, only a subset of the molecules in the amplified library produce sequence reads. So, for example, the number of

IDT Exhibit 1020.4049

amplified molecules sampled for sequencing may be about only 50% of the unique polynucleotides in the tagged library. Furthermore, amplification may be biased in favor of or against certain sequences or certain members of the tagged library. This may distort quantitative measurement of sequences in the tagged library. Also, sequencing platforms can introduce errors in sequencing. For example, sequences can have a per-base error rate of 0.5-1%. Amplification bias and sequencing errors introduce noise into the final sequencing product. This noise can diminish sensitivity of detection. For example, sequence variants whose frequency in the tagged population is less than the sequencing error rate can be mistaken for noise. Also, by providing reads of sequences in greater or less amounts than their actual number in a population, amplification bias can distort measurements of copy number variation. Alternatively, a plurality of sequence reads from a single polynucleotide can be produced without amplification. This can be done, for example, with nanopore methods.

[00216] This disclosure provides methods of accurately detecting and reading unique polynucleotides in a tagged pool. In certain embodiments this disclosure provides sequence-tagged polynucleotides that, when amplified and sequenced, or when sequenced a plurality of times to produce a plurality of sequence reads, provide information that allowed the tracing back, or collapsing, of progeny polynucleotides to the unique tag parent polynucleotide molecule. Collapsing families of amplified progeny polynucleotides reduces amplification bias by providing information about original unique parent molecules. Collapsing also reduces sequencing errors by eliminating from sequencing data mutant sequences of progeny molecules.

[00217] Detecting and reading unique polynucleotides in the tagged library can involve two strategies. In one strategy a sufficiently large subset of the amplified progeny polynucleotide pool is a sequenced such that, for a large percentage of unique tagged parent polynucleotides in the set of tagged parent polynucleotides, there is a sequence read that is produced for at least one amplified progeny polynucleotide in a family produced from a unique tagged parent polynucleotide. In a second strategy, the amplified progeny polynucleotide set is sampled for sequencing at a level to produce sequence reads from multiple progeny members of a family derived from a unique parent polynucleotide. Generation of sequence reads from multiple progeny members of a family allows collapsing of sequences into consensus parent sequences.

[00218] So, for example, sampling a number of amplified progeny polynucleotides from the set of amplified progeny polynucleotides that is equal to the number of unique tagged parent polynucleotides in the set of tagged parent polynucleotides (particularly when the number is at

IDT Exhibit 1020.4050

least 10,000) will produce, statistically, a sequence read for at least one of progeny of about 68% of the tagged parent polynucleotides in the set, and about 40% of the unique tagged parent polynucleotides in the original set will be represented by at least two progeny sequence reads. In certain embodiments the amplified progeny polynucleotide set is sampled sufficiently so as to produce an average of five to ten sequence reads for each family. Sampling from the amplified progeny set of 10-times as many molecules as the number of unique tagged parent polynucleotides will produce, statistically, sequence information about 99.995% of the families, of which 99.95% of the total families will be covered by a plurality of sequence reads. A consensus sequence can be built from the progeny polynucleotides in each family so as to dramatically reduce the error rate from the nominal per-base sequencing error rate to a rate possibly many orders of magnitude lower. For example, if the sequencer has a random per-base error rate of 1% and the chosen family has 10 reads, a consensus sequence built from these 10 reads would possess an error rate of below 0.0001%. Accordingly, the sampling size of the amplified progeny to be sequenced can be chosen so as to ensure a sequence having a frequency in the sample that is no greater than the nominal per-base sequencing error rate to a rate of the sequencing platform used, has at least 99% chance being represented by at least one read.

[00219] In another embodiment the set of amplified progeny polynucleotides is sampled to a level to produce a high probability e.g., at least 90%, that a sequence represented in the set of tagged parent polynucleotides at a frequency that is about the same as the per base sequencing error rate of the sequencing platform used is covered by at least one sequence read and preferably a plurality of sequence reads. So, for example, if the sequencing platform has a per base error rate of 0.2% in a sequence or set of sequences is represented in the set of tagged parent polynucleotides at a frequency of about 0.2%, then the number of polynucleotides in the amplified progeny pool that are sequenced can be about X times the number of unique molecules in the set of tagged parent polynucleotides.

[00220] These methods can be combined with any of the noise reduction methods described. Including, for example, qualifying sequence reads for inclusion in the pool of sequences used to generate consensus sequences.

[00221] This information can now be used for both qualitative and quantitative analysis. For example, for quantitative analysis, a measure, e.g., a count, of the amount of tagged parent molecules mapping to a reference sequence is determined. This measure can be compared with a measure of tagged parent molecules mapping to a different genomic region. That is, the amount

IDT Exhibit 1020.4051

of tagged parent molecules mapping to a first location or mappable position in a reference sequence, such as the human genome, can be compared with a measure of tagged parent molecules mapping to a second location or mappable position in a reference sequence. This comparison can reveal, for example, the relative amounts of parent molecules mapping to each region. This, in turn, provides an indication of copy number variation for molecules mapping to a particular region. For example, if the measure of polynucleotides mapping to a first reference sequence is greater than the measure of polynucleotides mapping to a second reference sequence, this may indicate that the parent population, and by extension the original sample, included polynucleotides from cells exhibiting aneuploidy. The measures can be normalized against a control sample to eliminate various biases. Quantitative measures can include, for example, number, count, frequency (whether relative, inferred or absolute).

[00222] A reference genome can include the genome of any species of interest. Human genome sequences useful as references can include the hg19 assembly or any previous or available hg assembly. Such sequences can be interrogated using the genome brower available at genome.ucsc.edu/index.html. Other species genomes include, for example PanTro2 (chimp) and mm9 (mouse).

[00223] For qualitative analysis, sequences from a set of tagged polynucleotides mapping to a reference sequence can be analyzed for variant sequences and their frequency in the population of tagged parent polynucleotides can be measured.

II. Sample Preparation

A. Polynucleotide Isolation and Extraction

[00224] The systems and methods of this disclosure may have a wide variety of uses in the manipulation, preparation, identification and/or quantification of cell free polynucleotides. Examples of polynucleotides include but are not limited to: DNA, RNA, amplicons, cDNA, dsDNA, ssDNA, plasmid DNA, cosmid DNA, high Molecular Weight (MW) DNA, chromosomal DNA, genomic DNA, viral DNA, bacterial DNA, mtDNA (mitochondrial DNA), mRNA, rRNA, tRNA, nRNA, siRNA, snRNA, snoRNA, scaRNA, microRNA, dsRNA, ribozyme, riboswitch and viral RNA (e.g., retroviral RNA).

[00225] Cell free polynucleotides may be derived from a variety of sources including human, mammal, non-human mammal, ape, monkey, chimpanzee, reptilian, amphibian, or avian, sources. Further, samples may be extracted from variety of animal fluids containing cell free

IDT Exhibit 1020.4052

sequences, including but not limited to blood, serum, plasma, vitreous, sputum, urine, tears, perspiration, saliva, semen, mucosal excretions, mucus, spinal fluid, amniotic fluid, lymph fluid and the like. Cell free polynucleotides may be fetal in origin (via fluid taken from a pregnant subject), or may be derived from tissue of the subject itself.

[00226] Isolation and extraction of cell free polynucleotides may be performed through collection of bodily fluids using a variety of techniques. In some cases, collection may comprise aspiration of a bodily fluid from a subject using a syringe. In other cases collection may comprise pipetting or direct collection of fluid into a collecting vessel.

[00227] After collection of bodily fluid, cell free polynucleotides may be isolated and extracted using a variety of techniques known in the art. In some cases, cell free DNA may be isolated, extracted and prepared using commercially available kits such as the Qiagen Qiamp® Circulating Nucleic Acid Kit protocol. In other examples, Qiagen Qubit ™ dsDNA HS Assay kit protocol, Agilent™ DNA 1000 kit, or TruSeq™ Sequencing Library Preparation; Low-Throughput (LT) protocol may be used.

[00228] Generally, cell free polynucleotides are extracted and isolated by from bodily fluids through a partitioning step in which cell free DNAs, as found in solution, are separated from cells and other non soluble components of the bodily fluid. Partitioning may include, but is not limited to, techniques such as centrifugation or filtration. In other cases, cells are not partitioned from cell free DNA first, but rather lysed. In this example, the genomic DNA of intact cells is partitioned through selective precipitation. Cell free polynucleotides, including DNA, may remain soluble and may be separated from insoluble genomic DNA and extracted. Generally, after addition of buffers and other wash steps specific to different kits, DNA may be precipitated using isopropanol precipitation. Further clean up steps may be used such as silica based columns to remove contaminants or salts. General steps may be optimized for specific applications. Non specific bulk carrier polynucleotides, for example, may be added throughout the reaction to optimize certain aspects of the procedure such as yield.

[00229] Isolation and purification of cell free DNA may be accomplished using any means, including, but not limited to, the use of commercial kits and protocols provided by companies such as Sigma Aldrich, Life Technologies, Promega, Affymetrix, IBI or the like. Kits and protocols may also be non-commercially available.

IDT Exhibit 1020.4053

[00230] After isolation, in some cases, the cell free polynucleotides are pre-mixed with one or more additional materials, such as one or more reagents (e.g., ligase, protease, polymerase) prior to sequencing.

[00231] One method of increasing conversion efficiency involves using a ligase engineered for optimal reactivity on single-stranded DNA, such as a ThermoPhage ssDNA ligase derivative. Such ligases bypass traditional steps in library preparation of end-repair and A-tailing that can have poor efficiencies and/or accumulated losses due to intermediate cleanup steps, and allows for twice the probability that either the sense or anti-sense starting polynucleotide will be converted into an appropriately tagged polynucleotide. It also converts double-stranded polynucleotides that may possess overhangs that may not be sufficiently blunt-ended by the typical end-repair reaction. Optimal reactions conditions for this ssDNA reaction are: 1 x reaction buffer (50 mM MOPS (pH 7.5), 1 mM DTT, 5 mM MgCl2, 10 mM KCl). With 50 mM ATP, 25 mg/ml BSA, 2.5 mM MnCl2 , 200 pmol 85 nt ssDNA oligomer and 5 U ssDNA ligase incubated at 65°C for 1 hour. Subsequent amplification using PCR can further convert the tagged single-stranded library to a double-stranded library and yield an overall conversion efficiency of well above 20%. Other methods of increasing conversion rate, e.g., to above 10%, include, for example, any of the following, alone or in combination: Annealing-optimized molecular-inversion probes, blunt-end ligation with a well-controlled polynucleotide size range, sticky-end ligation or an upfront multiplex amplification step with or without the use of fusion primers.

B. Molecular Bar Coding of Cell Free Polynucleotides

[00232] The systems and methods of this disclosure may also enable the cell free polynucleotides to be tagged or tracked in order to permit subsequent identification and origin of the particular polynucleotide. This feature is in contrast with other methods that use pooled or multiplex reactions and that only provide measurements or analyses as an average of multiple samples. Here, the assignment of an identifier to individual or subgroups of polynucleotides may allow for a unique identity to be assigned to individual sequences or fragments of sequences. This may allow acquisition of data from individual samples and is not limited to averages of samples.

[00233] In some examples, nucleic acids or other molecules derived from a single strand may share a common tag or identifier and therefore may be later identified as being derived from that strand. Similarly, all of the fragments from a single strand of nucleic acid may be tagged with the same identifier or tag, thereby permitting subsequent identification of fragments from the parent strand. In other cases, gene expression products (e.g., mRNA) may be tagged in order to

IDT Exhibit 1020.4054

quantify expression, by which the barcode, or the barcode in combination with sequence to which it is attached can be counted. In still other cases, the systems and methods can be used as a PCR amplification control. In such cases, multiple amplification products from a PCR reaction can be tagged with the same tag or identifier. If the products are later sequenced and demonstrate sequence differences, differences among products with the same identifier can then be attributed to PCR error.

[00234] Additionally, individual sequences may be identified based upon characteristics of sequence data for the read themselves. For example, the detection of unique sequence data at the beginning (start) and end (stop) portions of individual sequencing reads may be used, alone or in combination, with the length, or number of base pairs of each sequence read unique sequence to assign unique identities to individual molecules. Fragments from a single strand of nucleic acid, having been assigned a unique identity, may thereby permit subsequent identification of fragments from the parent strand. This can be used in conjunction with bottlenecking the initial starting genetic material to limit diversity.

[00235] Further, using unique sequence data at the beginning (start) and end (stop) portions of individual sequencing reads and sequencing read length may be used, alone or combination, with the use of barcodes. In some cases, the barcodes may be unique as described herein. In other cases, the barcodes themselves may not be unique. In this case, the use of non unique barcodes, in combination with sequence data at the beginning (start) and end (stop) portions of individual sequencing reads and sequencing read length may allow for the assignment of a unique identity to individual sequences. Similarly, fragments from a single strand of nucleic acid having been assigned a unique identity, may thereby permit subsequent identification of fragments from the parent strand.

[00236] Generally, the methods and systems provided herein are useful for preparation of cell free polynucleotide sequences to a down-stream application sequencing reaction. Often, a sequencing method is classic Sanger sequencing. Sequencing methods may include, but are not limited to: high-throughput sequencing, pyrosequencing, sequencing-by-synthesis, single-molecule sequencing, nanopore sequencing, semiconductor sequencing, sequencing-by-ligation, sequencing-by-hybridization, RNA-Seq (Illumina), Digital Gene Expression (Helicos), Next generation sequencing, Single Molecule Sequencing by Synthesis (SMSS)(Helicos), massively-parallel sequencing, Clonal Single Molecule Array (Solexa), shotgun sequencing, Maxim-Gilbert sequencing, primer walking, and any other sequencing methods known in the art.

IDT Exhibit 1020.4055

C. Assignment of Barcodes to Cell Free Polynucleotide Sequences

[00237] The systems and methods disclosed herein may be used in applications that involve the assignment of unique or non-unique identifiers, or molecular barcodes, to cell free polynucleotides. Often, the identifier is a bar-code oligonucleotide that is used to tag the polynucleotide; but, in some cases, different unique identifiers are used. For example, in some cases, the unique identifier is a hybridization probe. In other cases, the unique identifier is a dye, in which case the attachment may comprise intercalation of the dye into the analyte molecule (such as intercalation into DNA or RNA) or binding to a probe labeled with the dye. In still other cases, the unique identifier may be a nucleic acid oligonucleotide, in which case the attachment to the polynucleotide sequences may comprise a ligation reaction between the oligonucleotide and the sequences or incorporation through PCR. In other cases, the reaction may comprise addition of a metal isotope, either directly to the analyte or by a probe labeled with the isotope. Generally, assignment of unique or non-unique identifiers, or molecular barcodes in reactions of this disclosure may follow methods and systems described by, for example, US patent applications 20010053519, 20030152490, 20110160078 and US patent US 6,582,908.

[00238] Often, the method comprises attaching oligonucleotide barcodes to nucleic acid analytes through an enzymatic reaction including but not limited to a ligation reaction. For example, the ligase enzyme may covalently attach a DNA barcode to fragmented DNA (e.g., high molecular-weight DNA). Following the attachment of the barcodes, the molecules may be subjected to a sequencing reaction.

[00239] However, other reactions may be used as well. For example, oligonucleotide primers containing barcode sequences may be used in amplification reactions (e.g., PCR, qPCR, reverse-transcriptase PCR, digital PCR, etc.) of the DNA template analytes, thereby producing tagged analytes. After assignment of barcodes to individual cell free polynucleotide sequences, the pool of molecules may be sequenced.

[00240] In some cases, PCR may be used for global amplification of cell free polynucleotide sequences. This may comprise using adapter sequences that may be first ligated to different molecules followed by PCR amplification using universal primers. PCR for sequencing may be performed using any means, including but not limited to use of commercial kits provided by Nugen (WGA kit), Life Technologies, Affymetrix, Promega, Qiagen and the like. In other cases, only certain target molecules within a population of cell free polynucleotide molecules may

IDT Exhibit 1020.4056

be amplified. Specific primers, may in conjunction with adapter ligation, may be used to selectively amplify certain targets for downstream sequencing.

[00241] The unique identifiers (e.g., oligonucleotide bar-codes, antibodies, probes, etc.) may be introduced to cell free polynucleotide sequences randomly or non-randomly. In some cases, they are introduced at an expected ratio of unique identifiers to microwells. For example, the unique identifiers may be loaded so that more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10000, 50,000, 100,000, 500,000, 1,000,000, 10,000,000, 50,000,000 or 1,000,000,000 unique identifiers are loaded per genome sample. In some cases, the unique identifiers may be loaded so that less than about 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10000, 50,000, 100,000, 500,000, 1,000,000, 10,000,000, 50,000,000 or 1,000,000,000 unique identifiers are loaded per genome sample. In some cases, the average number of unique identifiers loaded per sample genome is less than, or greater than, about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000, 5000, 10000, 50,000, 100,000, 500,000, 1,000,000, 10,000,000, 50,000,000 or 1,000,000,000 unique identifiers per genome sample.

[00242] In some cases, the unique identifiers may be a variety of lengths such that each barcode is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000 base pairs. In other cases, the barcodes may comprise less than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, 100, 500, 1000 base pairs.

[00243] In some cases, unique identifiers may be predetermined or random or semi-random sequence oligonucleotides. In other cases, a plurality of barcodes may be used such that barcodes are not necessarily unique to one another in the plurality. In this example, barcodes may be ligated to individual molecules such that the combination of the bar code and the sequence it may be ligated to creates a unique sequence that may be individually tracked. As described herein, detection of non unique barcodes in combination with sequence data of beginning (start) and end (stop) portions of sequence reads may allow assignment of a unique identity to a particular molecule. The length, or number of base pairs, of an individual sequence read may also be used to assign a unique identity to such a molecule. As described herein, fragments from a single strand of nucleic acid having been assigned a unique identity, may thereby permit subsequent identification of fragments from the parent strand. In this way the polynucleotides in the sample can be uniquely or substantially uniquely tagged.

[00244] The unique identifiers may be used to tag a wide range of analytes, including but not limited to RNA or DNA molecules. For example, unique identifiers (e.g., barcode

IDT Exhibit 1020.4057

oligonucleotides) may be attached to whole strands of nucleic acids or to fragments of nucleic acids (e.g., fragmented genomic DNA, fragmented RNA). The unique identifiers (e.g., oligonucleotides) may also bind to gene expression products, genomic DNA, mitochondrial DNA, RNA, mRNA, and the like.

[00245] In many applications, it may be important to determine whether individual cell free polynucleotide sequences each receive a different unique identifier (e.g., oligonucleotide barcode). If the population of unique identifiers introduced into the systems and methods is not significantly diverse, different analytes may possibly be tagged with identical identifiers. The systems and methods disclosed herein may enable detection of cell free polynucleotide sequences tagged with the same identifier. In some cases, a reference sequences may be included with the population of cell free polynucleotide sequences to be analyzed. The reference sequence may be, for example, a nucleic acid with a known sequence and a known quantity. If the unique identifiers are oligonucleotide barcodes and the analytes are nucleic acids, the tagged analytes may subsequently be sequenced and quantified. These methods may indicate if one or more fragments and/or analytes may have been assigned an identical barcode.

[00246] A method disclosed herein may comprise utilizing reagents necessary for the assignment of barcodes to the analytes. In the case of ligation reactions, reagents including, but not limited to, ligase enzyme, buffer, adapter oligonucleotides, a plurality of unique identifier DNA barcodes and the like may be loaded into the systems and methods. In the case of enrichment, reagents including but not limited to a plurality of PCR primers, oligonucleotides containing unique identifying sequence, or barcode sequence, DNA polymerase, DNTPs, and buffer and the like may be used in preparation for sequencing.

[00247] Generally, the method and system of this disclosure may utilize the methods of US patent US 7,537,897 in using molecular barcodes to count molecules or analytes.

[00248] In a sample comprising fragmented genomic DNA, e.g., cell-free DNA (cfDNA), from a plurality of genomes, there is some likelihood that more than one polynucleotide from different genomes will have the same start and stop positions ("duplicates" or "cognates"). The probable number of duplicates beginning at any position is a function of the number of haploid genome equivalents in a sample and the distribution of fragment sizes. For example, cfDNA has a peak of fragments at about 160 nucleotides, and most of the fragments in this peak range from about 140 nucleotides to 180 nucleotides. Accordingly, cfDNA from a genome of about 3 billion bases (e.g., the human genome) may be comprised of almost 20 million ($2x10^7$) polynucleotide

IDT Exhibit 1020.4058

fragments. A sample of about 30 ng DNA can contain about 10,000 haploid human genome equivalents. (Similarly, a sample of about 100 ng of DNA can contain about 30,000 haploid human genome equivalents.) A sample containing about 10,000 (10^4) haploid genome equivalents of such DNA can have about 200 billion ($2x10^{11}$) individual polynucleotide molecules. It has been empirically determined that in a sample of about 10,000 haploid genome equivalents of human DNA, there are about 3 duplicate polynucleotides beginning at any given position. Thus, such a collection can contain a diversity of about $6x10^{10}$-$8x10^{10}$ (about 60 billion-80 billion e.g., about 70 billion ($7x10^{10}$)) differently sequenced polynucleotide molecules.

[00249] The probability of correctly identifying molecules is dependent on initial number of genome equivalents, the length distribution of sequenced molecules, sequence uniformity and number of tags. When the tag count is equal to one, that is, equivalent to having no unique tags or not tagging. The table below lists the probability of correctly identifying a molecule as unique assuming a typical cell-free size distribution as above.

Tag Count	Tag %Correctly uniquely identified
1000 human haploid genome equivalents	
1	96.9643
4	99.2290
9	99.6539
16	99.8064
25	99.8741
100	99.9685
3000 human haploid genome equivalents	
1	91.7233
4	97.8178
9	99.0198
16	99.4424

IDT Exhibit 1020.4059

25	99.6412
100	99.9107

[00250] In this case, upon sequencing the genomic DNA, it may not be possible to determine which sequence reads are derived from which parent molecules. This problem can be diminished by tagging parent molecules with a sufficient number of unique identifiers (e.g., the tag count) such that there is a likelihood that two duplicate molecules, i.e., molecules having the same start and stop positions, bear different unique identifiers so that sequence reads are traceable back to particular parent molecules. One approach to this problem is to uniquely tag every, or nearly every, different parent molecule in the sample. However, depending on the number of haploid gene equivalents and distribution of fragment sizes in the sample, this may require billions of different unique identifiers.

[00251] This method can be cumbersome and expensive. This invention provides methods and compositions in which a population of polynucleotides in a sample of fragmented genomic DNA is tagged with n different unique identifiers, wherein n is at least 2 and no more than 100,000*z, wherein z is a measure of central tendency (e.g., mean, median, mode) of an expected number of duplicate molecules having the same start and stop positions. In certain embodiments, n is at least any of 2*z, 3*z, 4*z, 5*z, 6*z, 7*z, 8*z, 9*z, 10*z, 11*z, 12*z, 13*z, 14*z, 15*z, 16*z, 17*z, 18*z, 19*z, or 20*z (e.g., lower limit). In other embodiments, n is no greater than100,000*z, 10,000*z, 1000*z or 100*z (e.g., upper limit). Thus, n can range between any combination of these lower and upper limits. In certain embodiments, n is between 5*z and 15*z, between 8*z and 12*z, or about 10*z. For example, a haploid human genome equivalent has about 3 picograms of DNA. A sample of about 1 microgram of DNA contains about 300,000 haploid human genome equivalents. The number n can be between 15 and 45, between 24 and 36 or about 30. Improvements in sequencing can be achieved as long as at least some of the duplicate or cognate polynucleotides bear unique identifiers, that is, bear different tags. However, in certain embodiments, the number of tags used is selected so that there is at least a 95% chance that all duplicate molecules starting at any one position bear unique identifiers. For example, a sample comprising about 10,000 haploid human genome equivalents of cfDNA can be tagged with about 36 unique identifiers. The unique identifiers can comprise six unique DNA barcodes. Attached to both ends of a polynucleotide, 36 possible unique identifiers are produced. Samples

IDT Exhibit 1020.4060

tagged in such a way can be those with a range of about 10 ng to any of about 100 ng, about 1 μg, about 10 μg of fragmented polynucleotides, e.g., genomic DNA, e.g. cfDNA.

[00252] Accordingly, this invention also provides compositions of tagged polynucleotides. The polynucleotides can comprise fragmented DNA, e.g. cfDNA. A set of polynucleotides in the composition that map to a mappable base position in a genome can be non-uniquely tagged, that is, the number of different identifiers can be at least at least 2 and fewer than the number of polynucleotides that map to the mappable base position. A composition of between about 10 ng to about 10 μg (e.g., any of about 10 ng-1 μg, about 10 ng-100 ng, about 100 ng-10 μg, about 100 ng-1 μg, about 1 μg-10 μg) can bear between any of 2, 5, 10, 50 or 100 to any of 100, 1000, 10,000 or 100,000 different identifiers. For example, between 5 and 100 different identifiers can be used to tag the polynucleotides in such a composition.

III. Nucleic Acid Sequencing Platforms

[00253] After extraction and isolation of cell free polynucleotides from bodily fluids, cell free sequences may be sequenced. Often, a sequencing method is classic Sanger sequencing. Sequencing methods may include, but are not limited to: high-throughput sequencing, pyrosequencing, sequencing-by-synthesis, single-molecule sequencing, nanopore sequencing, semiconductor sequencing, sequencing-by-ligation, sequencing-by-hybridization, RNA-Seq (Illumina), Digital Gene Expression (Helicos), Next generation sequencing, Single Molecule Sequencing by Synthesis (SMSS)(Helicos), massively-parallel sequencing, Clonal Single Molecule Array (Solexa), shotgun sequencing, Maxim-Gilbert sequencing, primer walking, sequencing using PacBio, SOLiD, Ion Torrent, or Nanopore platforms and any other sequencing methods known in the art.

[00254] In some cases, sequencing reactions various types, as described herein, may comprise a variety of sample processing units. Sample processing units may include but are not limited to multiple lanes, multiple channels, multiple wells, or other mean of processing multiple sample sets substantially simultaneously. Additionally, the sample processing unit may include multiple sample chambers to enable processing of multiple runs simultaneously.

[00255] In some examples, simultaneous sequencing reactions may be performed using multiplex sequencing. In some cases, cell free polynucleotides may be sequenced with at least 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000 sequencing reactions. In other cases cell free poly nucleotides may be sequenced with less than 1000, 2000,

IDT Exhibit 1020.4061

3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000 sequencing reactions. Sequencing reactions may be performed sequentially or simultaneously. Subsequent data analysis may be performed on all or part of the sequencing reactions. In some cases, data analysis may be performed on at least 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000 sequencing reactions. In other cases data analysis may be performed on less than 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 50000, 100,000 sequencing reactions.

[00256] In other examples, the number of sequence reactions may provide coverage for different amounts of the genome. In some cases, sequence coverage of the genome may be at least 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 99.9% or 100%. In other cases, sequence coverage of the genome may be less than 5%, 10%, 15%, 20%, 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 99%, 99.9% or 100%.

[00257] In some examples, sequencing can be performed on cell free polynucleotides that may comprise a variety of different types of nucleic acids. Nucleic acids may be polynucleotides or oligonucleotides. Nucleic acids included, but are not limited to DNA or RNA, single stranded or double stranded or a RNA/cDNA pair.

IV. Polynucleotide Analysis Strategy

[00258] Fig 8. is a diagram, **800**, showing a strategy for analyzing polynucleotides in a sample of initial genetic material. In step **802,** a sample containing initial genetic material is provided. The sample can include target nucleic acid in low abundance. For example, nucleic acid from a normal or wild-type genome (e.g., a germline genome) can predominate in a sample that also includes no more than 20%, no more than 10%, no more than 5%, no more than 1%, no more than 0.5% or no more than 0.1% nucleic acid from at least one other genome containing genetic variation, e.g., a cancer genome or a fetal genome, or a genome from another species. The sample can include, for example, cell free nucleic acid or cells comprising nucleic acid. The initial genetic material can constitute no more than 100 ng nucleic acid. This can contribute to proper oversampling of the original polynucleotides by the sequencing or genetic analysis process. Alternatively, the sample can be artificially capped or bottlenecked to reduce the amount of nucleic acid to no more than 100 ng or selectively enriched to analyze only sequences of interest. The sample can be modified to selectively produce sequence reads of molecules mapping to each of one or more selected locations in a reference sequence. A sample of 100 ng of nucleic acid can

IDT Exhibit 1020.4062

contain about 30,000 human haploid genome equivalents, that is, molecules that, together, provide 30,000-fold coverage of a human genome.

[00259] In step **804** the initial genetic material is converted into a set of tagged parent polynucleotides. Tagging can include attaching sequenced tags to molecules in the initial genetic material. Sequenced tags can be selected so that all unique polynucleotides mapping to the same location in a reference sequence had a unique identifying tag. Conversion can be performed at high efficiency, for example at least 50%.

[00260] In step **806**, the set of tagged parent polynucleotides is amplified to produce a set of amplified progeny polynucleotides. Amplification may be, for example, 1,000-fold.

[00261] In step **808**, the set of amplified progeny polynucleotides are sampled for sequencing. The sampling rate is chosen so that the sequence reads produced both (1) cover a target number of unique molecules in the set of tagged parent polynucleotides and (2) cover unique molecules in the set of tagged parent polynucleotides at a target coverage fold (e.g., 5- to 10-fold coverage of parent polynucleotides.

[00262] In step **810**, the set of sequence reads is collapsed to produce a set of consensus sequences corresponding to unique tagged parent polynucleotides. Sequence reads can be qualified for inclusion in the analysis. For example, sequence reads that fail to meet a quality control scores can be removed from the pool. Sequence reads can be sorted into families representing reads of progeny molecules derived from a particular unique parent molecule. For example, a family of amplified progeny polynucleotides can constitute those amplified molecules derived from a single parent polynucleotide. By comparing sequences of progeny in a family, a consensus sequence of the original parent polynucleotide can be deduced. This produces a set of consensus sequences representing unique parent polynucleotides in the tagged pool.

[00263] In step **812**, the set of consensus sequences is analyzed using any of the analytical methods described herein. For example, consensus sequences mapping to a particular reference sequence location can be analyzed to detect instances of genetic variation. Consensus sequences mapping to particular reference sequences can be measured and normalized against control samples. Measures of molecules mapping to reference sequences can be compared across a genome to identify areas in the genome in which copy number varies, or heterozygosity is lost.

[00264] Fig. 9 is a diagram presenting a more generic method of extracting information from a signal represented by a collection of sequence reads. In this method, after sequencing amplified progeny polynucleotides, the sequence reads are grouped into families of molecules

IDT Exhibit 1020.4063

amplified from a molecule of unique identity (910). This grouping can be a jumping off point for methods of interpreting the information in the sequence to determine the contents of the tagged parent polynucleotides with higher fidelity, e.g., less noise and/or distortion.

[00265] Analysis of the collection of sequence reads allows one to make inferences about the parent polynucleotide population from which the sequence reads were generated. Such inferences may be useful because sequencing typically involves reading only a partial subset of the global total amplified polynucleotides. Therefore, one cannot be certain that every parent polynucleotide will be represented by at least one sequence read in the collection of sequence reads.

[00266] One such inference is the number of unique parent polynucleotides in the original pool. Such an inference can be made based on the number of unique families into which the sequence reads can be grouped and the number of sequence reads in each family. In this case, a family refers to a collection of sequence reads traceable back to an original parent polynucleotide. The inference can be made using well-known statistical methods. For example, if grouping produces many families, each represented by one or a few progeny, then one can infer that the original population included more unique parent polynucleotides that were not sequenced. On the other hand, if grouping produces only a few families, each family represented by many progeny, then one can infer that most of the unique polynucleotides in the parent population are represented by at least one sequence read group into that family.

[00267] Another such inference is the frequency of a base or sequence of bases at a particular locus in an original pool of polynucleotides. Such an inference can be made based on the number of unique families into which the sequence reads can be grouped and the number of sequence reads in each family. Analyzing the base calls at a locus in a family of sequence reads, a confidence score is assigned to each particular base call or sequence. Then, taking into consideration the confidence score for each base call in a plurality of the families, the frequency of each base or sequence at the locus is determined.

V. Copy Number Variation Detection

A. Copy Number Variation Detection Using Single Sample

[00268] Fig 1. is a diagram, **100**, showing a strategy for detection of copy number variation in a single subject. As shown herein, copy number variation detection methods can be implemented as follows. After extraction and isolation of cell free polynucleotides in step **102**, a

IDT Exhibit 1020.4064

single unique sample can be sequenced by a nucleic acid sequencing platform known in the art in step **104**. This step generates a plurality of genomic fragment sequence reads. In some cases, these sequences reads may contain barcode information. In other examples, barcodes are not utilized. After sequencing, reads are assigned a quality score. A quality score may be a representation of reads that indicates whether those reads may be useful in subsequent analysis based on a threshold. In some cases, some reads are not of sufficient quality or length to perform the subsequent mapping step. Sequencing reads with a quality score at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data. In other cases, sequencing reads assigned a quality scored less than 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In step **106**, the genomic fragment reads that meet a specified quality score threshold are mapped to a reference genome, or a template sequence that is known not to contain copy number variations. After mapping alignment, sequence reads are assigned a mapping score. A mapping score may be a representation or reads mapped back to the reference sequence indicating whether each position is or is not uniquely mappable. In instances, reads may be sequences unrelated to copy number variation analysis. For example, some sequence reads may originate from contaminant polynucleotides. Sequencing reads with a mapping score at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In other cases, sequencing reads assigned a mapping scored less than 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set.

[00269] After data filtering and mapping, the plurality of sequence reads generates a chromosomal region of coverage. In step **108** these chromosomal regions may be divided into variable length windows or bins. A window or bin may be at least 5 kb, 10, kb, 25 kb, 30 kb, 35, kb, 40 kb, 50 kb, 60 kb, 75 kb, 100 kb, 150 kb, 200 kb, 500 kb, or 1000 kb. A window or bin may also have bases up to 5 kb, 10, kb, 25 kb, 30 kb, 35, kb, 40 kb, 50 kb, 60 kb, 75 kb, 100 kb, 150 kb, 200 kb, 500 kb, or 1000 kb. A window or bin may also be about 5 kb, 10, kb, 25 kb, 30 kb, 35, kb, 40 kb, 50 kb, 60 kb, 75 kb, 100 kb, 150 kb, 200 kb, 500 kb, or 1000 kb.

[00270] For coverage normalization in step **110**, each window or bin is selected to contain about the same number of mappable bases. In some cases, each window or bin in a chromosomal region may contain the exact number of mappable bases. In other cases, each window or bin may contain a different number of mappable bases. Additionally, each window or bin may be non-overlapping with an adjacent window or bin. In other cases, a window or bin may overlap with another adjacent window or bin. In some cases a window or bin may overlap by at least 1 bp, 2,

bp, 3 bp, 4 bp, 5, bp, 10 bp, 20 bp, 25 bp, 50 bp, 100 bp, 200 bp, 250 bp, 500 bp, or 1000 bp. In other cases, a window or bin may overlap by up to 1 bp, 2, bp, 3 bp, 4 bp, 5, bp, 10 bp, 20 bp, 25 bp, 50 bp, 100 bp, 200 bp, 250 bp, 500 bp, or 1000 bp. In some cases a window or bin may overlap by about 1 bp, 2, bp, 3 bp, 4 bp, 5, bp, 10 bp, 20 bp, 25 bp, 50 bp, 100 bp, 200 bp, 250 bp, 500 bp, or 1000 bp.

[00271] In some cases, each of the window regions may be sized so they contain about the same number of uniquely mappable bases. The mappability of each base that comprise a window region is determined and used to generate a mappability file which contains a representation of reads from the references that are mapped back to the reference for each file. The mappability file contains one row per every position, indicating whether each position is or is not uniquely mappable.

[00272] Additionally, predefined windows, known throughout the genome to be hard to sequence, or contain a substantially high GC bias, may be filtered from the data set. For example, regions known to fall near the centromere of chromosomes (i.e., centromeric DNA) are known to contain highly repetitive sequences that may produce false positive results. These regions may be filtered out. Other regions of the genome, such as regions that contain an unusually high concentration of other highly repetitive sequences such as microsatellite DNA, may be filtered from the data set.

[00273] The number of windows analyzed may also vary. In some cases, at least 10, 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 5,000, 10,000, 20,000, 50,000 or 100,000 windows are analyzed. In other cases, the number of widows analyzed is up to 10, 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 5,000, 10,000, 20,000, 50,000 or 100,000 windows are analyzed.

[00274] For an exemplary genome derived from cell free polynucleotide sequences, the next step comprises determining read coverage for each window region. This may be performed using either reads with barcodes, or without barcodes. In cases without barcodes, the previous mapping steps will provide coverage of different base positions. Sequence reads that have sufficient mapping and quality scores and fall within chromosome windows that are not filtered, may be counted. The number of coverage reads may be assigned a score per each mappable position. In cases involving barcodes, all sequences with the same barcode, physical properties or combination of the two may be collapsed into one read, as they are all derived from the sample parent molecule. This step reduces biases which may have been introduced during any of the preceding steps, such as steps involving amplification. For example, if one molecule is amplified

IDT Exhibit 1020.4066

10 times but another is amplified 1000 times, each molecule is only represented once after collapse thereby negating the effect of uneven amplification. Only reads with unique barcodes may be counted for each mappable position and influence the assigned score.

[00275] Consensus sequences can be generated from families of sequence reads by any method known in the art. Such methods include, for example, linear or non-linear methods of building consensus sequences (such as voting, averaging, statistical, maximum a posteriori or maximum likelihood detection, dynamic programming, Bayesian, hidden Markov or support vector machine methods, etc.) derived from digital communication theory, information theory, or bioinformatics.

[00276] After the sequence read coverage has been determined, a stochastic modeling algorithm is applied to convert the normalized nucleic acid sequence read coverage for each window region to the discrete copy number states. In some cases, this algorithm may comprise one or more of the following: Hidden Markov Model, dynamic programming, support vector machine, Bayesian network, trellis decoding, Viterbi decoding, expectation maximization, Kalman filtering methodologies and neural networks.

[00277] In step **112**, the discrete copy number states of each window region can be utilized to identify copy number variation in the chromosomal regions. In some cases, all adjacent window regions with the same copy number can be merged into a segment to report the presence or absence of copy number variation state. In some cases, various windows can be filtered before they are merged with other segments.

[00278] In step **114**, the copy number variation may be reported as graph, indicating various positions in the genome and a corresponding increase or decrease or maintenance of copy number variation at each respective position. Additionally, copy number variation may be used to report a percentage score indicating how much disease material (or nucleic acids having a copy number variation) exists in the cell free polynucleotide sample.

[00279] One method of determining copy number variation is shown in Fig. 10. In that method, after grouping sequence reads into families generated from a single parent polynucleotide (1010), the families are quantified, for example, by determining the number of families mapping to each of a plurality of different reference sequence locations. CNVs can be determined directly by comparing a quantitative measure of families at each of a plurality of different loci (1016b). Alternatively, one can infer a quantitative measure of families in the population of tagged parent polynucleotides using both a quantitative measure of families and a quantitative measure of

IDT Exhibit 1020.4067

family members in each family, e.g., as discussed above. Then, CNV can be determined by comparing the inferred measure of quantity at the plurality of loci. In other embodiments, a hybrid approach can be taken whereby a similar inference of original quantity can be made following normalization for representational bias during the sequencing process, such as GC bias, etc

B. Copy Number Variation Detection Using Paired Sample

[00280] Paired sample copy number variation detection shares many of the steps and parameters as the single sample approach described herein. However, as depicted in **200** of Fig. 2 of copy number variation detection using paired samples requires comparison of sequence coverage to a control sample rather than comparing it the predicted mappability of the genome. This approach may aid in normalization across windows.

[00281] Fig 2. is a diagram, **200** showing a strategy for detection of copy number variation in paired subject. As shown herein, copy number variation detection methods can be implemented as follows. In step **204**, a single unique sample can be sequenced by a nucleic acid sequencing platform known in the art after extraction and isolation of the sample in step **202**. This step generates a plurality of genomic fragment sequence reads. Additionally, a sample or control sample is taken from another subject. In some cases, the control subject may be a subject not known to have disease, whereas the other subject may have or be at risk for a particular disease. In some cases, these sequences reads may contain barcode information. In other examples, barcodes are not utilized. After sequencing, reads are assigned a quality score. In some cases, some reads are not of sufficient quality or length to perform the subsequent mapping step. Sequencing reads with a quality score at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In other cases, sequencing reads assigned a quality scored less than 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In step **206**, the genomic fragment reads that meet a specified quality score threshold are mapped to a reference genome, or a template sequence that is known not to contain copy number variations. After mapping alignment, sequence reads are assigned a mapping score. In instances, reads may be sequences unrelated to copy number variation analysis. For example, some sequence reads may originate from contaminant polynucleotides. Sequencing reads with a mapping score at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In other cases, sequencing reads assigned a mapping scored less than 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set.

IDT Exhibit 1020.4068

[00282] After data filtering and mapping, the plurality of sequence reads generates a chromosomal region of coverage for each of the test and control subjects. In step **208** these chromosomal regions may be divided into variable length windows or bins. A window or bin may be at least 5 kb, 10, kb, 25 kb, 30 kb, 35, kb, 40 kb, 50 kb, 60 kb, 75 kb, 100 kb, 150 kb, 200 kb, 500 kb, or 1000 kb. A window or bin may also be less than 5 kb, 10, kb, 25 kb, 30 kb, 35, kb, 40 kb, 50 kb, 60 kb, 75 kb, 100 kb, 150 kb, 200 kb, 500 kb, or 1000 kb.

[00283] For coverage normalization in step **210**, each window or bin is selected to contain about the same number of mappable bases for each of the test and control subjects. In some cases, each window or bin in a chromosomal region may contain the exact number of mappable bases. In other cases, each window or bin may contain a different number of mappable bases. Additionally, each window or bin may be non-overlapping with an adjacent window or bin. In other cases, a window or bin may overlap with another adjacent window or bin. In some cases a window or bin may overlap by at least 1 bp, 2, bp, 3 bp, 4 bp, 5, bp, 10 bp, 20 bp, 25 bp, 50 bp, 100 bp, 200 bp, 250 bp, 500 bp, or 1000 bp. In other cases, a window or bin may overlap by less than 1 bp, 2, bp, 3 bp, 4 bp, 5, bp, 10 bp, 20 bp, 25 bp, 50 bp, 100 bp, 200 bp, 250 bp, 500 bp, or 1000 bp.

[00284] In some cases, each of the window regions is sized so they contain about the same number of uniquely mappable bases for each of the test and control subjects. The mappability of each base that comprise a window region is determined and used to generate a mappability file which contains a representation of reads from the references that are mapped back to the reference for each file. The mappability file contains one row per every position, indicating whether each position is or is not uniquely mappable.

[00285] Additionally, predefined windows, known throughout the genome to be hard to sequence, or contain a substantially high GC bias, are filtered from the data set. For example, regions known to fall near the centromere of chromosomes (i.e., centromeric DNA) are known to contain highly repetitive sequences that may produce false positive results. These regions may be filtered. Other regions of the genome, such as regions that contain an unusually high concentration of other highly repetitive sequences such as microsatellite DNA, may be filtered from the data set.

[00286] The number of windows analyzed may also vary. In some cases, at least 10, 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 5,000, 10,000, 20,000, 50,000 or 100,000 windows are

IDT Exhibit 1020.4069

analyzed. In other cases, less than 10, 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 5,000, 10,000, 20,000, 50,000 or 100,000 windows are analyzed.

[00287] For an exemplary genome derived from cell free polynucleotide sequences, the next step comprises determining read coverage for each window region for each of the test and control subjects. This may be performed using either reads with barcodes, or without barcodes. In cases without barcodes, the pervious mapping steps will provide coverage of different base positions. Sequence reads that have sufficient mapping and quality scores and fall within chromosome windows that are not filtered, may be counted. The number of coverage reads may be assigned a score per each mappable position. In cases involving barcodes, all sequences with the same barcode may be collapsed into one read, as they are all derived from the sample parent molecule. This step reduces biases which may have been introduced during any of the preceding steps, such as steps involving amplification. Only reads with unique barcodes may be counted for each mappable position and influence the assigned score. For this reason, it is important that the barcode ligation step be performed in a manner optimized for producing the lowest amount of bias.

[00288] In determining the nucleic acid read coverage for each window, the coverage of each window can be normalized by the mean coverage of that sample. Using such an approach, it may be desirable to sequence both the test subject and the control under similar conditions. The read coverage for each window may be then expressed as a ratio across similar windows

[00289] Nucleic acid read coverage ratios for each window of the test subject can be determined by dividing the read coverage of each window region of the test sample with read coverage of a corresponding window region of the control ample.

[00290] After the sequence read coverage ratios have been determined, a stochastic modeling algorithm is applied to convert the normalized ratios for each window region into discrete copy number states. In some cases, this algorithm may comprise a Hidden Markov Model. In other cases, the stochastic model may comprise dynamic programming, support vector machine, Bayesian modeling, probabilistic modeling, trellis decoding, Viterbi decoding, expectation maximization, Kalman filtering methodologies, or neural networks.

[00291] In step **212**, the discrete copy number states of each window region can be utilized to identify copy number variation in the chromosomal regions. In some cases, all adjacent window regions with the same copy number can be merged into a segment to report the presence

IDT Exhibit 1020.4070

or absence of copy number variation state. In some cases, various windows can be filtered before they are merged with other segments.

[00292] In step **214**, the copy number variation may be reported as graph, indicating various positions in the genome and a corresponding increase or decrease or maintenance of copy number variation at each respective position. Additionally, copy number variation may be used to report a percentage score indicating how much disease material exists in the cell free poly nucleotide sample.

VI. Rare mutation Detection

[00293] Rare mutation detection shares similar features as both copy number variation approaches. However, as depicted in Fig. 3, **300**, rare mutation detection uses comparison of sequence coverage to a control sample or reference sequence rather than comparing it the relative mappability of the genome. This approach may aid in normalization across windows.

[00294] Generally, rare mutation detection may be performed on selectively enriched regions of the genome or transcriptome purified and isolated in step **302**. As described herein, specific regions, which may include but are not limited to genes, oncogenes, tumor suppressor genes, promoters, regulatory sequence elements, non-coding regions, miRNAs, snRNAs and the like may be selectively amplified from a total population of cell free polynucleotides. This may be performed as herein described. In one example, multiplex sequencing may be used, with or without barcode labels for individual polynucleotide sequences. In other examples, sequencing may be performed using any nucleic acid sequencing platforms known in the art. This step generates a plurality of genomic fragment sequence reads as in step **304**. Additionally, a reference sequence is obtained from a control sample, taken from another subject. In some cases, the control subject may be a subject known to not have known genetic aberrations or disease. In some cases, these sequence reads may contain barcode information. In other examples, barcodes are not utilized. After sequencing, reads are assigned a quality score. A quality score may be a representation of reads that indicates whether those reads may be useful in subsequent analysis based on a threshold. In some cases, some reads are not of sufficient quality or length to perform the subsequent mapping step. Sequencing reads with a quality score at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In other cases, sequencing reads assigned a quality scored at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In step **306**, the genomic fragment reads that meet a specified quality score

IDT Exhibit 1020.4071

threshold are mapped to a reference genome, or a reference sequence that is known not to contain rare mutations. After mapping alignment, sequence reads are assigned a mapping score. A mapping score may be a representation or reads mapped back to the reference sequence indicating whether each position is or is not uniquely mappable. In instances, reads may be sequences unrelated to rare mutation analysis. For example, some sequence reads may originate from contaminant polynucleotides. Sequencing reads with a mapping score at least 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set. In other cases, sequencing reads assigned a mapping scored less than 90%, 95%, 99%, 99.9%, 99.99% or 99.999% may be filtered out of the data set.

[00295] For each mappable base, bases that do not meet the minimum threshold for mappability, or low quality bases, may be replaced by the corresponding bases as found in the reference sequence.

[00296] After data filtering and mapping, variant bases found between the sequence reads obtained from the subject and the reference sequence are analyzed.

[00297] For an exemplary genome derived from cell free polynucleotide sequences, the next step comprises determining read coverage for each mappable base position. This may be performed using either reads with barcodes, or without barcodes. In cases without barcodes, the previous mapping steps will provide coverage of different base positions. Sequence reads that have sufficient mapping and quality scores may be counted. The number of coverage reads may be assigned a score per each mappable position. In cases involving barcodes, all sequences with the same barcode may be collapsed into one consensus read, as they are all derived from the sample parent molecule. The sequence for each base is aligned as the most dominant nucleotide read for that specific location. Further, the number of unique molecules can be counted at each position to derive simultaneous quantification at each position. This step reduces biases which may have been introduced during any of the preceding steps, such as steps involving amplification. Only reads with unique barcodes may be counted for each mappable position and influence the assigned score.

[00298] Once read coverage may be ascertained and variant bases relative to the control sequence in each read are identified, the frequency of variant bases may be calculated as the number of reads containing the variant divided by the total number of reads. This may be expressed as a ratio for each mappable position in the genome.

IDT Exhibit 1020.4072

[00299] For each base position, the frequencies of all four nucleotides, cytosine, guanine, thymine, adenine are analyzed in comparison to the reference sequence. A stochastic or statistical modeling algorithm is applied to convert the normalized ratios for each mappable position to reflect frequency states for each base variant. In some cases, this algorithm may comprise one or more of the following: Hidden Markov Model, dynamic programming, support vector machine, Bayesian or probabilistic modeling, trellis decoding, Viterbi decoding, expectation maximization, Kalman filtering methodologies, and neural networks.

[00300] In step **312**, the discrete rare mutation states of each base position can be utilized to identify a base variant with high frequency of variance as compared to the baseline of the reference sequence. In some cases, the baseline might represent a frequency of at least 0.0001%, 0.001%, 0.01%, 0.1%, 1.0%, 2.0%, 3.0%, 4.0% 5.0%, 10%, or 25%. In other cases the baseline might represent a frequency of at least 0.0001%, 0.001%, 0.01%, 0.1%, 1.0%, 2.0%, 3.0%, 4.0% 5.0%. 10%, or 25%. In some cases, all adjacent base positions with the base variant or mutation can be merged into a segment to report the presence or absence of a rare mutation. In some cases, various positions can be filtered before they are merged with other segments.

[00301] After calculation of frequencies of variance for each base position, the variant with largest deviation for a specific position in the sequence derived from the subject as compared to the reference sequence is identified as a rare mutation. In some cases, a rare mutation may be a cancer mutation. In other cases, a rare mutation might be correlated with a disease state.

[00302] A rare mutation or variant may comprise a genetic aberration that includes, but is not limited to a single base substitution, or small indels, transversions, translocations, inversion, deletions, truncations or gene truncations. In some cases, a rare mutation may be at most 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nucleotides in length. On other cases a rare mutation may be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nucleotides in length.

[00303] In step **314**, the presence or absence of a mutation may be reflected in graphical form, indicating various positions in the genome and a corresponding increase or decrease or maintenance of a frequency of mutation at each respective position. Additionally, rare mutations may be used to report a percentage score indicating how much disease material exists in the cell free polynucleotide sample. A confidence score may accompany each detected mutation, given known statistics of typical variances at reported positions in non-disease reference sequences. Mutations may also be ranked in order of abundance in the subject or ranked by clinically actionable importance.

IDT Exhibit 1020.4073

[00304] Fig. 11 shows a method of inferring the frequency of a base or sequence of bases at a particular locus in a population polynucleotides. Sequence reads are grouped into families generated from an original tagged polynucleotide (1110). For each family, one or more bases at the locus is each assigned a confidence score. The confidence score can be assigned by any of a number of known statistical methods is assigned and can be based, at least in part, on the frequency at which a base appears among the sequence reads belonging to the family (1112). For example, the confidence score can be the frequency at which the base appears among the sequence reads. As another example, for each family, a hidden Markov model can be built, such that a maximum likelihood or maximum a posteriori decision can be made based on the frequency of occurrence of a particular base in a single family. As part of this model, the probability of error and resultant confidence score for a particular decision can be output as well. A frequency of the base in the original population can then be assigned based on the confidence scores among the families (1114).

VII. Applications

A. Early Detection of Cancer

[00305] Numerous cancers may be detected using the methods and systems described herein. Cancers cells, as most cells, can be characterized by a rate of turnover, in which old cells die and replaced by newer cells. Generally dead cells, in contact with vasculature in a given subject, may release DNA or fragments of DNA into the blood stream. This is also true of cancer cells during various stages of the disease. Cancer cells may also be characterized, dependent on the stage of the disease, by various genetic aberrations such as copy number variation as well as rare mutations. This phenomenon may be used to detect the presence or absence of cancers individuals using the methods and systems described herein.

[00306] For example, blood from subjects at risk for cancer may be drawn and prepared as described herein to generate a population of cell free polynucleotides. In one example, this might be cell free DNA. The systems and methods of the disclosure may be employed to detect rare mutations or copy number variations that may exist in certain cancers present. The method may help detect the presence of cancerous cells in the body, despite the absence of symptoms or other hallmarks of disease.

[00307] The types and number of cancers that may be detected may include but are not limited to blood cancers, brain cancers, lung cancers, skin cancers, nose cancers, throat cancers,

IDT Exhibit 1020.4074

liver cancers, bone cancers, lymphomas, pancreatic cancers, skin cancers, bowel cancers, rectal cancers, thyroid cancers, bladder cancers, kidney cancers, mouth cancers, stomach cancers, solid state tumors, heterogeneous tumors, homogenous tumors and the like.

[00308] In the early detection of cancers, any of the systems or methods herein described, including rare mutation detection or copy number variation detection may be utilized to detect cancers. These system and methods may be used to detect any number of genetic aberrations that may cause or result from cancers. These may include but are not limited to mutations, rare mutations, indels, copy number variations, transversions, translocations, inversion, deletions, aneuploidy, partial aneuploidy, polyploidy, chromosomal instability, chromosomal structure alterations, gene fusions, chromosome fusions, gene truncations, gene amplification, gene duplications, chromosomal lesions, DNA lesions, abnormal changes in nucleic acid chemical modifications, abnormal changes in epigenetic patterns, abnormal changes in nucleic acid methylation infection and cancer..

[00309] Additionally, the systems and methods described herein may also be used to help characterize certain cancers. Genetic data produced from the system and methods of this disclosure may allow practitioners to help better characterize a specific form of cancer. Often times, cancers are heterogeneous in both composition and staging. Genetic profile data may allow characterization of specific sub-types of cancer that may be important in the diagnosis or treatment of that specific sub-type. This information may also provide a subject or practitioner clues regarding the prognosis of a specific type of cancer.

B. Cancer Monitoring and Prognosis

[00310] The systems and methods provided herein may be used to monitor already known cancers, or other diseases in a particular subject. This may allow either a subject or practitioner to adapt treatment options in accord with the progress of the disease. In this example, the systems and methods described herein may be used to construct genetic profiles of a particular subject of the course of the disease. In some instances, cancers can progress, becoming more aggressive and genetically unstable. In other examples, cancers may remain benign, inactive, dormant or in remission. The system and methods of this disclosure may be useful in determining disease progression, remission or recurrence.

[00311] Further, the systems and methods described herein may be useful in determining the efficacy of a particular treatment option. In one example, successful treatment options may actually increase the amount of copy number variation or rare mutations detected in subject's

blood if the treatment is successful as more cancers may die and shed DNA. In other examples, this may not occur. In another example, perhaps certain treatment options may be correlated with genetic profiles of cancers over time. This correlation may be useful in selecting a therapy. Additionally, if a cancer is observed to be in remission after treatment, the systems and methods described herein may be useful in monitoring residual disease or recurrence of disease.

[00312] For example, mutations occurring within a range of frequency beginning at threshold level can be determined from DNA in a sample from a subject, e.g., a patient. The mutations can be, e.g., cancer related mutations. The frequency can range from, for example, at least 0.1%, at least 1%, or at least 5% to 100%. The sample can be, e.g., cell free DNA or a tumor sample. A course of treatment can be prescribed based on any or all of mutations occurring within the frequency range including, e.g., their frequencies. A sample can be taken from the subject at any subsequent time. Mutations occurring within the original range of frequency or a different range of frequency can be determined. The course of treatment can be adjusted based on the subsequent measurements.

C. Early Detection and Monitoring of Other Diseases or Disease States

[00313] The methods and systems described herein may not be limited to detection of rare mutations and copy number variations associated with only cancers. Various other diseases and infections may result in other types of conditions that may be suitable for early detection and monitoring. For example, in certain cases, genetic disorders or infectious diseases may cause a certain genetic mosaicism within a subject. This genetic mosaicism may cause copy number variation and rare mutations that could be observed. In another example, the system and methods of the disclosure may also be used to monitor the genomes of immune cells within the body. Immune cells, such as B cells, may undergo rapid clonal expansion upon the presence certain diseases. Clonal expansions may be monitored using copy number variation detection and certain immune states may be monitored. In this example, copy number variation analysis may be performed over time to produce a profile of how a particular disease may be progressing.

[00314] Further, the systems and methods of this disclosure may also be used to monitor systemic infections themselves, as may be caused by a pathogen such as a bacteria or virus. Copy number variation or even rare mutation detection may be used to determine how a population of pathogens are changing during the course of infection. This may be particularly important during chronic infections, such as HIV/AIDs or Hepatitis infections, whereby viruses may change life cycle state and/or mutate into more virulent forms during the course of infection.

IDT Exhibit 1020.4076

[00315] Yet another example that the system and methods of this disclosure may be used for is the monitoring of transplant subjects. Generally, transplanted tissue undergoes a certain degree of rejection by the body upon transplantation. The methods of this disclosure may be used to determine or profile rejection activities of the host body, as immune cells attempt to destroy transplanted tissue. This may be useful in monitoring the status of transplanted tissue as well as altering the course of treatment or prevention of rejection.

[00316] Further, the methods of the disclosure may be used to characterize the heterogeneity of an abnormal condition in a subject, the method comprising generating a genetic profile of extracellular polynucleotides in the subject, wherein the genetic profile comprises a plurality of data resulting from copy number variation and rare mutation analyses. In some cases, including but not limited to cancer, a disease may be heterogeneous. Disease cells may not be identical. In the example of cancer, some tumors are known to comprise different types of tumor cells, some cells in different stages of the cancer. In other examples, heterogeneity may comprise multiple foci of disease. Again, in the example of cancer, there may be multiple tumor foci, perhaps where one or more foci are the result of metastases that have spread from a primary site.

[00317] The methods of this disclosure may be used to generate or profile, fingerprint or set of data that is a summation of genetic information derived from different cells in a heterogeneous disease. This set of data may comprise copy number variation and rare mutation analyses alone or in combination.

D. Early Detection and Monitoring of Other Diseases or Disease States of Fetal Origin

[00318] Additionally, the systems and methods of the disclosure may be used to diagnose, prognose, monitor or observe cancers or other diseases of fetal origin. That is, these methodologies may be employed in a pregnant subject to diagnose, prognose, monitor or observe cancers or other diseases in a unborn subject whose DNA and other polynucleotides may co-circulate with maternal molecules.

VIII. Terminology

[00319] The terminology used therein is for the purpose of describing particular embodiments only and is not intended to be limiting of a systems and methods of this disclosure. As used herein, the singular forms “a”, “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. Furthermore, to the extent that the terms “including”, “includes”, “having”, “has”, “with”, or variants thereof are used in either the detailed

IDT Exhibit 1020.4077

description and/or the claims, such terms are intended to be inclusive in a manner similar to the term "comprising".

[00320] Several aspects of a systems and methods of this disclosure are described above with reference to example applications for illustration. It should be understood that numerous specific details, relationships, and methods are set forth to provide a full understanding of a systems and methods. One having ordinary skill in the relevant art, however, will readily recognize that a systems and methods can be practiced without one or more of the specific details or with other methods. This disclosure is not limited by the illustrated ordering of acts or events, as some acts may occur in different orders and/or concurrently with other acts or events. Furthermore, not all illustrated acts or events are required to implement a methodology in accordance with this disclosure.

[00321] Ranges can be expressed herein as from "about" one particular value, and/or to "about" another particular value. When such a range is expressed, another embodiment includes from the one particular value and/or to the other particular value. Similarly, when values are expressed as approximations, by use of the antecedent "about," it will be understood that the particular value forms another embodiment. It will be further understood that the endpoints of each of the ranges are significant both in relation to the other endpoint, and independently of the other endpoint. The term "about" as used herein refers to a range that is 15% plus or minus from a stated numerical value within the context of the particular usage. For example, about 10 would include a range from 8.5 to 11.5.

Computer systems

[00322] Methods of the present disclosure can be implemented using, or with the aid of, computer systems. FIG. 15 shows a computer system 1501 that is programmed or otherwise configured to implement the methods of the present disclosure. The computer system 1501 can regulate various aspects sample preparation, sequencing and/or analysis. In some examples, the computer system 1501 is configured to perform sample preparation and sample analysis, including nucleic acid sequencing.

[00323] The computer system 1501 includes a central processing unit (CPU, also "processor" and "computer processor" herein) 1505, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system 1501 also includes memory or memory location 1510 (e.g., random-access memory, read-only memory, flash memory), electronic storage unit 1515 (e.g., hard disk), communication interface 1520 (e.g.,

network adapter) for communicating with one or more other systems, and peripheral devices 1525, such as cache, other memory, data storage and/or electronic display adapters. The memory 1510, storage unit 1515, interface 1520 and peripheral devices 1525 are in communication with the CPU 1505 through a communication bus (solid lines), such as a motherboard. The storage unit 1515 can be a data storage unit (or data repository) for storing data. The computer system 1501 can be operatively coupled to a computer network ("network") 1530 with the aid of the communication interface 1520. The network 1530 can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network 1530 in some cases is a telecommunication and/or data network. The network 1530 can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network 1530, in some cases with the aid of the computer system 1501, can implement a peer-to-peer network, which may enable devices coupled to the computer system 1501 to behave as a client or a server.

[00324] The CPU 1505 can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions may be stored in a memory location, such as the memory 1510. Examples of operations performed by the CPU 1505 can include fetch, decode, execute, and writeback.

[00325] The storage unit 1515 can store files, such as drivers, libraries and saved programs. The storage unit 1515 can store programs generated by users and recorded sessions, as well as output(s) associated with the programs. The storage unit 1515 can store user data, e.g., user preferences and user programs. The computer system 1501 in some cases can include one or more additional data storage units that are external to the computer system 1501, such as located on a remote server that is in communication with the computer system 1501 through an intranet or the Internet.

[00326] The computer system 1501 can communicate with one or more remote computer systems through the network 1530. For instance, the computer system 1501 can communicate with a remote computer system of a user (e.g., operator). Examples of remote computer systems include personal computers (e.g., portable PC), slate or tablet PC's (e.g., Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (e.g., Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system 1501 via the network 1530.

[00327] Methods as described herein can be implemented by way of machine (e.g., computer processor) executable code stored on an electronic storage location of the computer system 1501, such as, for example, on the memory 1510 or electronic storage unit 1515. The machine executable or machine readable code can be provided in the form of software. During use, the code can be executed by the processor 1505. In some cases, the code can be retrieved from the storage unit 1515 and stored on the memory 1510 for ready access by the processor 1505. In some situations, the electronic storage unit 1515 can be precluded, and machine-executable instructions are stored on memory 1510.

[00328] The code can be pre-compiled and configured for use with a machine have a processer adapted to execute the code, or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

[00329] Aspects of the systems and methods provided herein, such as the computer system 1501, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. "Storage" type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" refer to any medium that participates in providing instructions to a processor for execution.

[00330] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[00331] The computer system 1501 can include or be in communication with an electronic display that comprises a user interface (UI) for providing, for example, one or more results of sample analysis. Examples of UI's include, without limitation, a graphical user interface (GUI) and web-based user interface.

EXAMPLES

Example 1 – Prostate Cancer Prognosis and Treatment

[00332] A blood sample is taken from a prostate cancer subject. Previously, an oncologist determines that the subject has stage II prostate cancer and recommends a treatment. Cell free DNA is extracted, isolated, sequenced and analyzed every 6 months after the initial diagnosis.

[00333] Cell free DNA is extracted and isolated from blood using the Qiagen Qubit kit protocol. A carrier DNA is added to increase yields. DNA is amplified using PCR and universal primers. 10 ng of DNA is sequenced using a massively parallel sequencing approach with an

IDT Exhibit 1020.4081

Illumina MiSeq personal sequencer. 90% of the subject's genome is covered through sequencing of cell free DNA.

[00334] Sequence data is assembled and analyzed for copy number variation. Sequence reads are mapped and compared to a healthy individual (control). Based on the number of sequence reads, chromosomal regions are divided into 50 kb non overlapping regions. Sequence reads are compared to one another and a ratio is determined for each mappable position.

[00335] A Hidden Markov Model is applied to convert copy numbers into discrete states for each window.

[00336] Reports are generated, mapping genome positions and copy number variation show in Fig. 4A (for a healthy individual) and Fig. 4B for the subject with cancer.

[00337] These reports, in comparison to other profiles of subjects with known outcomes, indicate that this particular cancer is aggressive and resistant to treatment. The cell free tumor burden is 21%. The subject is monitored for 18 months. At month 18, the copy number variation profile begins to increase dramatically, from cell free tumor burden of 21% to 30%. A comparison is done with genetic profiles of other prostate subjects. It is determined that this increase in copy number variation indicates that the prostate cancer is advancing from stage II to stage III. The original treatment regiment as prescribed is no longer treating the cancer. A new treatment is prescribed.

[00338] Further, these reports are submitted and accessed electronically via the internet. Analysis of sequence data occurs at a site other than the location of the subject. The report is generated and transmitted to the subject's location. Via an internet enabled computer, the subject accesses the reports reflecting his tumor burden (Fig. 4C).

Example 2 – Prostate Cancer Remission and Recurrence.

[00339] A blood sample is taken from a prostate cancer survivor. The subject had previously undergone numerous rounds of chemotherapy and radiation. The subject at the time of testing did not present symptoms or health issues related to the cancer. Standard scans and assays reveal the subject to be cancer free.

[00340] Cell free DNA is extracted and isolated from blood using the Qiagen TruSeq kit protocol. A carrier DNA is added to increase yields. DNA is amplified using PCR and universal primers. 10 ng of DNA is sequenced using a massively parallel sequencing approach with an

IDT Exhibit 1020.4082

Illumina MiSeq personal sequencer. 12mer barcodes are added to individual molecules using a ligation method.

[00341] Sequence data is assembled and analyzed for copy number variation. Sequence reads are mapped and compared to a healthy individual (control). Based on the number of sequence reads, chromosomal regions are divided into 40 kb non overlapping regions. Sequence reads are compared to one another and a ratio is determined for each mappable position.

[00342] Non unique barcoded sequences are collapsed into a single read to help normalize bias from amplification.

[00343] A Hidden Markov Model is applied to convert copy numbers into discrete states for each window.

[00344] Reports are generated, mapping genome positions and copy number variation shown in Fig. 5A, for a subject with cancer in remission and Fig. 5B for a subject with cancer in recurrence.

[00345] This report in comparison to other profiles of subjects with known outcomes indicates that at month 18, rare mutation analysis for copy number variation is detected at cell free tumor burden of 5%. An oncologist prescribes treatment again.

Example 3 – Thyroid Cancer and Treatment

[00346] A subject is known to have Stage IV thyroid cancer and undergoes standard treatment, including radiation therapy with I-131. CT scans are inconclusive as to whether the radiation therapy is destroying cancerous masses. Blood is drawn before and after the latest radiation session.

[00347] Cell free DNA is extracted and isolated from blood using the Qiagen Qubit kit protocol. A sample of non specific bulk DNA is added to the sample preparation reactions increase yields.

[00348] It is known that the BRAF gene may be mutated at amino acid position 600 in this thyroid cancer. From population of cell free DNA, BRAF DNA is selectively amplified using primers specific to the gene. 20mer barcodes are added to the parent molecule as a control for counting reads.

[00349] 10 ng of DNA is sequenced using massively parallel sequencing approach with an Illumina MiSeq personal sequencer.

IDT Exhibit 1020.4083

[00350] Sequence data is assembled and analyzed for copy number variation detection. Sequence reads are mapped and compared to a healthy individual (control). Based on the number of sequence reads, as determined by counting the barcode sequences, chromosomal regions are divided into 50 kb non overlapping regions. Sequence reads are compared to one another and a ratio is determined for each mappable position.

[00351] A Hidden Markov Model is applied to convert copy numbers into discrete states for each window.

[00352] A report is generated, mapping genome positions and copy number variation.

[00353] The reports generated before and after treatment are compared. The tumor cell burden percentage jumps from 30% to 60% after the radiation session. The jump in tumor burden is determined to be an increase in necrosis of cancer tissue versus normal tissue as a result of treatment. Oncologists recommend the subject continue the prescribed treatment.

Example 4 – Sensitivity of Rare mutation Detection

[00354] In order to determine the detection ranges of rare mutation present in a population of DNA, mixing experiments are performed. Sequences of DNA, some containing wildtype copies of the genes TP53, HRAS and MET and some containing copies with rare mutations in the same genes, are mixed together in distinct ratios. DNA mixtures are prepared such that ratios or percentages of mutant DNA to wildtype DNA range from 100% to 0.01%.

[00355] 10 ng of DNA is sequenced for each mixing experiment using a massively parallel sequencing approach with an Illumina MiSeq personal sequencer.

[00356] Sequence data is assembled and analyzed for rare mutation detection. Sequence reads are mapped and compared to a reference sequence (control). Based on the number of sequence reads, the frequency of variance for each mappable position is determined.

[00357] A Hidden Markov Model is applied to convert frequency of variance for each mappable position into discrete states for base position.

[00358] A report is generated, mapping genome base positions and percentage detection of the rare mutation over baseline as determined by the reference sequence (Fig. 6A).

[00359] The results of various mixing experiments ranging from 0.1% to 100% are represented in a logarithmic scale graph, with measured percentage of DNA with a rare mutation graphed as a function of the actual percentage of DNA with a rare mutation (Fig. 6B). The three genes, TP53, HRAS and MET are represented. A strong linear correlation is found between

IDT Exhibit 1020.4084

measured and expected rare mutation populations. Additionally, a lower sensitivity threshold of about 0.1% of DNA with a rare mutation in a population of non mutated DNA is found with these experiments (Fig. 6B).

Example 5 –Rare mutation Detection in Prostate Cancer Subject

[00360] A subject is thought to have early stage prostate cancer. Other clinical tests provide inconclusive results. Blood is drawn from the subject and cell free DNA is extracted, isolated, prepared and sequenced.

[00361] A panel of various oncogenes and tumor suppressor genes are selected for selective amplification using a TaqMan© PCR kit (Invitrogen) using gene specific primers. DNA regions amplified include DNA containing PIK3CA and TP53 genes.

[00362] 10 ng of DNA is sequenced using a massively parallel sequencing approach with an Illumina MiSeq personal sequencer.

[00363] Sequence data is assembled and analyzed for rare mutation detection. Sequence reads are mapped and compared to a reference sequence (control). Based on the number of sequence reads, the frequency of variance for each mappable position was determined.

[00364] A Hidden Markov Model is applied to convert frequency of variance for each mappable position into discrete states for each base position.

[00365] A report is generated, mapping genomic base positions and percentage detection of the rare mutation over baseline as determined by the reference sequence (Fig. 7A). Rare mutations are found at an incidence of 5% in two genes, PIK3CA and TP53, respectively, indicating that the subject has an early stage cancer. Treatment is initiated.

[00366] Further, these reports are submitted and accessed electronically via the internet. Analysis of sequence data occurs at a site other than the location of the subject. The report is generated and transmitted to the subject's location. Via an internet enabled computer, the subject accesses the reports reflecting his tumor burden (Fig. 7B).

Example 6 –Rare mutation Detection in Colorectal Cancer Subjects

[00367] A subject is thought to have mid-stage colorectal cancer. Other clinical tests provide inconclusive results. Blood is drawn from the subject and cell free DNA is extracted.

[00368] 10 ng of the cell-free genetic material that is extracted from a single tube of plasma is used. The initial genetic material is converted into a set of tagged parent

IDT Exhibit 1020.4085

polynucleotides. The tagging included attaching tags required for sequencing as well as non-unique identifiers for tracking progeny molecules to the parent nucleic acids. The conversion is performed through an optimized ligation reaction as described above and conversion yield is confirmed by looking at the size profile of molecules post-ligation. Conversion yield is measured as the percentage of starting initial molecules that have both ends ligated with tags. Conversion using this approach is performed at high efficiency, for example, at least 50%.

[00369] The tagged library is PCR-amplified and enriched for genes most associated with colorectal cancer, (e.g., KRAS, APC, TP53, etc) and the resulting DNA is sequenced using a massively parallel sequencing approach with an Illumina MiSeq personal sequencer.

[00370] Sequence data is assembled and analyzed for rare mutation detection. Sequence reads are collapsed into familial groups belonging to a parent molecule (as well as error-corrected upon collapse) and mapped using a reference sequence (control). Based on the number of sequence reads, the frequency of rare variations (substitutions, insertions, deletions, etc) and variations in copy number and heterozygosity (when appropriate) for each mappable position is determined.

[00371] A report is generated, mapping genomic base positions and percentage detection of the rare mutation over baseline as determined by the reference sequence. Rare mutations are found at an incidence of 0.3-0.4% in two genes, KRAS and FBXW7, respectively, indicating that the subject has residual cancer. Treatment is initiated.

[00372] Further, these reports are submitted and accessed electronically via the internet. Analysis of sequence data occurs at a site other than the location of the subject. The report is generated and transmitted to the subject's location. Via an internet enabled computer, the subject accesses the reports reflecting his tumor burden.

[00373] **Example 7 – Digital Sequencing Technology**

[00374] The concentrations of tumor-shed nucleic acids are typically so low that current next-generation sequencing technologies can only detect such signals sporadically or in patients with terminally high tumor burden. The main reason being that such technologies are plagued by error rates and bias that can be orders of magnitude higher than what is required to reliably detect de novo genetic alterations associated with cancer in circulating DNA. Shown here is a new sequencing methodology, Digital Sequencing Technology (DST), which increases the sensitivity and specificity of detecting and quantifying rare tumor-derived nucleic acids among germline fragments by at least 1-2 orders of magnitude.

IDT Exhibit 1020.4086

[00375] DST architecture is inspired by state-of-the-art digital communication systems that combat the high noise and distortion caused by modern communication channels and are able to transmit digital information flawlessly at exceedingly high data rates. Similarly, current next-gen workflows are plagued by extremely high noise and distortion (due to sample-prep, PCR-based amplification and sequencing). Digital sequencing is able to eliminate the error and distortion created by these processes and produce near-perfect representation of all rare variants (including CNVs).

[00376] High-Diversity Library preparation

[00377] Unlike conventional sequencing library preparation protocols, whereby the majority of extracted circulating DNA fragments are lost due to inefficient library conversion, our Digital Sequencing Technology workflow enables the vast majority of starting molecules to be converted and sequenced. This is critically important for detection of rare variants as there may only be a handful of somatically mutated molecules in an entire 10mL tube of blood. The efficient molecular biology conversion process developed enables the highest possible sensitivity for detection of rare variants.

[00378] Comprehensive Actionable Oncogene Panel

[00379] The workflow engineered around the DST platform is flexible and highly tunable as targeted regions can be as small as single exons or as broad as whole exomes (or even whole genomes). A standard panel consists of all exonic bases of 15 actionable cancer-related genes and coverage of the "hot" exons of an additional 36 onco-/tumor-suppressor genes (e.g., exons containing at least one or more reported somatic mutations in COSMIC).

[00380] **Example 8: Analytical Studies**

[00381] To study the performance of our technology, its sensitivity in analytical samples was evaluated. We spiked varying amounts of LNCaP cancer cell line DNA into a background of normal cfDNA and were able to successfully detect somatic mutations down to 0.1% sensitivity (see Figure 13).

[00382] Preclinical Studies

[00383] The concordance of circulating DNA with tumor gDNA in human xenograft models in mice was investigated. In seven CTC-negative mice, each with one of two different human breast cancer tumors, all somatic mutations detected in tumor gDNA were also detected in mouse blood cfDNA using DST further validating the utility of cfDNA for non-invasive tumor genetic profiling.

IDT Exhibit 1020.4087

[00384] Pilot Clinical Studies

[00385] Correlation of tumor biopsy vs. circulating DNA somatic mutations

[00386] A pilot study was initiated on human samples across different cancer types. The concordance of tumor mutation profiles derived from circulating cell-free DNA with those derived from matched tumor biopsy samples was investigated. Higher than 93% concordance between tumor and cfDNA somatic mutation profiles in both colorectal and melanoma cancers across 14 patients was found (Table 1).

Table 1

Patient ID	**Stage**	**Mutant Genes in Matched Tumor**	**Percentage of mutant cfDNA**
CRC #1	II-B	TP53	0.2%
CRC #2	II-C	KRAS	0.6%
		SMAD4	1.5%
		GNAS	1.4%
		FBXW7	0.8%
CRC #3	III-B	KRAS	1.1%
		TP53	1.4%
		PIK3CA	1.7%
		APC	0.7%
CRC #4	III-B	KRAS	0.3%
		TP53	0.4%
CRC #5	III-B	KRAS	0.04%
CRC #6	III-C	KRAS	0.03%
CRC #7	IV	PIK3CA	1.3%
		KRAS	0.6%
		TP53	0.8%
CRC #8	IV	APC	0.3%
		SMO	0.6%

IDT Exhibit 1020.4088

		TP53	0.4%
		KRAS	0.0%
CRC #9	IV	APC	47.3%
		APC	40.2%
		KRAS	37.7%
		PTEN	0.0%
		TP53	12.9%
CRC #10	IV	TP53	0.9%
Melanoma #1	IV	BRAF	0.2%
Melanoma #2	IV	APC	0.3%
		EGFR	0.9%
		MYC	10.5%
Melanoma #3	IV	BRAF	3.3%
Melanoma #4	IV	BRAF	0.7%

[00387] It should be understood from the foregoing that, while particular implementations have been illustrated and described, various modifications can be made thereto and are contemplated herein. It is also not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the preferable embodiments herein are not meant to be construed in a limiting sense. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. Various modifications in form and detail of the embodiments of the invention will be apparent to a person skilled in the art. It is therefore contemplated that the invention shall also cover any such modifications, variations and equivalents.

IDT Exhibit 1020.4089

CLAIMS

WHAT IS CLAIMED IS:

1. A method for detecting copy number variation comprising:
   a. sequencing extracellular polynucleotides from a bodily sample from a subject, wherein each of the extracellular polynucleotide generate a plurality of sequencing reads;
   b. filtering out reads that fail to meet a set threshold;
   c. mapping the sequence reads obtained from step (a), after reads are filtered out, to a reference sequence;
   d. quantifying or enumerating mapped reads in two or more predefined regions of the reference sequence; and
   e. determining copy number variation in one or more of the predefined regions by:
      i. normalizing number of reads in the predefined regions to each other and/or the number of unique sequence reads in the predefined regions to one another;
      ii. comparing the normalized numbers obtained in step (i) to normalized numbers obtained from a control sample.

2. A method for detecting a rare mutation in a cell-free or substantially cell free sample obtained from a subject comprising:
   a. sequencing extracellular polynucleotides from a bodily sample from a subject, wherein each of the extracellular polynucleotide generate a plurality of sequencing reads;
   b. performing multiplex sequencing on regions or whole-genome sequencing if enrichment is not performed;
   c. filtering out reads that fail to meet a set threshold;
   d. mapping sequence reads derived from the sequencing onto a reference sequence;
   e. identifying a subset of mapped sequence reads that align with a variant of the reference sequence at each mappable base position;

IDT Exhibit 1020.4090

f. for each mappable base position, calculating a ratio of (a) a number of mapped sequence reads that include a variant as compared to the reference sequence, to (b) a number of total sequence reads for each mappable base position;
g. normalizing the ratios or frequency of variance for each mappable base position and determining potential rare variant(s) or mutation(s); and
h. and comparing the resulting number for each of the regions with potential rare variant(s) or mutation(s) to similarly derived numbers from a reference sample.

3. A method of characterizing the heterogeneity of an abnormal condition in a subject, the method comprising generating a genetic profile of extracellular polynucleotides in the subject, wherein the genetic profile comprises a plurality of data resulting from copy number variation and rare mutation analyses.

4. The method of claim 1, 2, or 3 wherein the prevalence/concentration of each rare variant identified in the subject is reported and quantified simultaneously.

5. The method of claim 1, 2, or 3 wherein a confidence score, regarding the prevalence/concentrations of rare variants in the subject, is reported.

6. The method of claim 1, 2, or 3 wherein the extracellular polynucleotides comprise DNA.

7. The method of claim 1, 2, or 3 wherein the extracellular polynucleotides comprise RNA.

8. The method of claim 1, 2, or 3 further comprising isolating extracellular polynucleotides from the bodily sample.

9. The method claim 1, 2, or 3 wherein the isolating comprises a method for circulating nucleic acid isolation and extraction.

10. The method of claim 1, 2, or 3 further comprising fragmenting said isolated extracellular polynucleotides.

11. The method of claim 8 wherein the bodily sample is selected from the group consisting of blood, plasma, serum, urine, saliva, mucosal excretions, sputum, stool and tears.

IDT Exhibit 1020.4091

12. The method of claim 1, 2, or 3, further comprising the step of determining the percent of sequences having copy number variation or rare mutation or variant in said bodily sample.

13. The method of claim 12 wherein the determining comprises calculating the percentage of predefined regions with an amount of polynucleotides above or below a predetermined threshold.

14. The method of claim 1, 2, or 3 wherein the subject is suspected of having an abnormal condition.

15. The method of claim 14 wherein the abnormal condition is selected from the group consisting of, mutations, rare mutations, indels, copy number variations, transversions, translocations, inversion, deletions, aneuploidy, partial aneuploidy, polyploidy, chromosomal instability, chromosomal structure alterations, gene fusions, chromosome fusions, gene truncations, gene amplification, gene duplications, chromosomal lesions, DNA lesions, abnormal changes in nucleic acid chemical modifications, abnormal changes in epigenetic patterns, abnormal changes in nucleic acid methylation infection and cancer.

16. The method of claim 1, 2, or 3 wherein the subject is a pregnant female.

17. The method of claim 1 or 2 wherein the copy number variation or rare mutation or genetic variant is indicative of a fetal abnormality.

18. The method of claim 17 wherein the fetal abnormality is selected from the group consisting of, mutations, rare mutations, indels, copy number variations, transversions, translocations, inversion, deletions, aneuploidy, partial aneuploidy, polyploidy, chromosomal instability, chromosomal structure alterations, gene fusions, chromosome fusions, gene truncations, gene amplification, gene duplications, chromosomal lesions, DNA lesions, abnormal changes in nucleic acid chemical modifications, abnormal changes in epigenetic patterns, abnormal changes in nucleic acid methylation infection and cancer.

19. The method of claim 1, 2, or 3 further comprising attaching one or more barcodes to the extracellular polynucleotides or fragments thereof prior to sequencing.

IDT Exhibit 1020.4092

20. The method of claim 19 wherein each barcode attached to extracellular polynucleotides or fragments thereof prior to sequencing is unique.

21. The method of claim 19 wherein each barcode attached to extracellular polynucleotides or fragments thereof prior to sequencing is not unique.

22. The method of claim 1, 2, or 3 further comprising selectively enriching regions from the subject's genome or transcriptome prior to sequencing.

23. The method of claim 1, 2, or 3 further comprising non-selectively enriching regions from the subject's genome or transcriptome prior to sequencing.

24. The method of claim 1, 2, or 3 further comprising attaching one or more barcodes to the extracellular polynucleotides or fragments thereof prior to any amplification or enrichment step.

25. The method of claim 19 wherein the barcode is a polynucleotide.

26. The method of claim 19 wherein the barcode comprises random sequence.

27. The method of claim 19 wherein the barcode comprises a fixed or semi-random set of oligonucleotides that in combination with the diversity of molecules sequenced from a select region enables identification of unique molecules.

28. The method of claim 19 wherein the barcodes comprise oligonucleotides is at least a 3, 5, 10, 15, 20 25, 30, 35, 40, 45, or 50mer base pairs in length.

29. The method of claim 1, 2, or 3 further comprises amplifying the extracellular polynucleotides or fragments thereof.

30. The method of claim 29 wherein the amplification comprises global amplification or whole genome amplification.

31. The method of claim 1, 2, or 3, wherein sequence reads of unique identity are detected based on sequence information at the beginning (start) and end (stop) regions of the sequence read and the length of the sequence read.

IDT Exhibit 1020.4093

32. The method of claim 31, wherein sequence molecules of unique identity are detected based on sequence information at the beginning (start) and end (stop) regions of the sequence read, the length of the sequence read and attachment of a barcode.

33. The method of claim 30 wherein the amplification comprises selective amplification.

34. The method of claim 33 wherein the amplification comprises non-selective amplification.

35. The method of claims 1, 2, or 3 wherein suppression amplification or subtractive enrichment is performed.

36. The method of claim 1, 2, or 3 further comprising removing a subset of the reads from further analysis prior to quantifying or enumerating reads.

37. The method of claim 36 wherein removing comprises filtering out reads with an accuracy or quality score of less than a threshold, e.g., 90%, 99%, 99.9%, or 99.99% and/or mapping score less than a threshold, e.g., 90%, 99%, 99.9% or 99.99%.

38. The method of claim 1, 2, or 3 further comprising filtering reads with a quality score lower than a set threshold.

39. The method of claim 1 wherein the predefined regions are uniform or substantially uniform in size.

40. The method of claim 39 wherein the predefined regions are at least about 10kb, 20kb, 30kb 40kb, 50kb, 60kb, 70kb, 80kb, 90kb, or 100kb in size.

41. The method of claim 1, 2, or 3 wherein at least 50, 100, 200, 500, 1000, 2000, 5000, 10,000, 20,000, or 50,000 regions are analyzed.

42. The method of claim 1, 2, or 3 wherein the variant occurs in a region of the genome selected from the group consisting of gene fusions, gene duplications, gene deletions, gene translocations, microsatellite regions, gene fragments or combination thereof.

IDT Exhibit 1020.4094

43. The method of claim 1, 2, or 3 wherein the variant occurs in a region of the genome selected from the group consisting of genes, oncogenes, tumor suppressor genes, promoters, regulatory sequence elements, or combination thereof.

44. The method of claim 2 wherein the variant is a nucleotide variant, single base substitution, small indel, transversion, translocation, inversion, deletion, truncation or gene truncation of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 or 20 nucleotides in length.

45. The method of claim 1, 2, or 3 further comprising correcting/normalizing/adjusting the quantity of mapped reads using the barcodes or unique properties of individual reads.

46. The method of claim 1 or 2 wherein enumerating the reads is performed through enumeration of unique barcodes in each of the predefined regions and normalizing those numbers across at least a subset of predefined regions that were sequenced.

47. The method of claim 1, 2, or 3 wherein samples at succeeding time intervals from the same subject are analyzed and compared to previous sample results.

48. The method of claim 45 wherein the method further comprises amplifying the barcode-attached extracellular polynucleotides.

49. The method of claim 1, 2, or 3 further comprising determining partial copy number variation frequency, determining loss of heterozygosity, performing gene expression analysis, performing epigenetic analysis and/or performing hypermethylation analysis.

50. A method comprising: determining copy number variation or performing rare mutation analysis in a cell-free or substantially cell free sample obtained from a subject using multiplex sequencing.

51. The method of claim 50 wherein the multiplex sequencing comprises performing over 10,000 sequencing reactions.

52. The method of claim 50 wherein the multiplex sequencing comprises simultaneously sequencing at least 10,000 different reads.

IDT Exhibit 1020.4095

53. The method of claim 50 wherein the multiplex sequencing comprising performing data analysis on at least 10,000 different reads across the genome.

54. The method of claim 1 or 2 wherein the normalizing and detection is performed using one or more of hidden markov, dynamic programming, support vector machine, Bayesian or probabilistic modeling, trellis decoding, Viterbi decoding, expectation maximization, Kalman filtering, or neural network methodologies.

55. The method of claim 1, 2, or 3 further comprising monitoring disease progression, monitoring residual disease, monitoring therapy, diagnosing a condition, prognosing a condition, or selecting a therapy based on discovered variants for the subject.

56. The method of claim 55 wherein a therapy is modified based on the most recent sample analysis.

57. The method of claim 1, 2, or 3 wherein the genetic profile of a tumor, infection or other tissue abnormality is inferred.

58. The method of claim 1, 2, or 3 wherein the growth, remission or evolution of a tumor, infection or other tissue abnormality is monitored.

59. The method of claim 1, 2, or 3 wherein sequences related to the subject's immune system are analyzed and monitored at single instances or over time.

60. The method of claim 1, 2, or 3 wherein identification of a variant is followed up through an imaging test (e.g., CT, PET-CT, MRI, X-ray, ultrasound) for localization of the tissue abnormality suspected of causing the identified variant.

61. The method of claim 1, 2, or 3 wherein the analysis further comprises use of genetic data obtained from a tissue or tumor biopsy from the same patient.

62. The method of claim 1, 2, or 3 wherein the phylogenetics of a tumor, infection or other tissue abnormality is inferred.

63. The method of claim 1 or 2, wherein the method further comprises performing population-based no-calling and identification of low-confidence regions.

IDT Exhibit 1020.4096

64. The method of claim 1 or 2 wherein obtaining the measurement data for the sequence coverage comprises measuring sequence coverage depth at every position of the genome.

65. The method of claim 64 wherein correcting the measurement data for the sequence coverage bias comprises calculating window-averaged coverage.

66. The method of claim 64 wherein correcting the measurement data for the sequence coverage bias comprises performing adjustments to account for GC bias in the library construction and sequencing process.

67. The method of claim 64 wherein correcting the measurement data for the sequence coverage bias comprises performing adjustments based on additional weighting factor associated with individual mappings to compensate for bias.

68. The method of claim 1, 2, or 3 wherein extracellular polynucleotide is derived from a diseased cell origin.

69. The method of claim 1, 2, or 3 wherein extracellular polynucleotide is derived from a healthy cell origin.

70. A system comprising a computer readable medium for performing the following steps: selecting predefined regions in a genome; enumerating number of sequence reads in the predefined regions; normalizing the number of sequence reads across the predefined regions; and determining percent of copy number variation in the predefined regions.

71. The system of claim 70 wherein the entirety of the genome or at least 85% of the genome is analyzed.

72. The system of claim 70 wherein the computer readable medium provides data on percent cancer DNA or RNA in plasma or serum to the end user.

73. The method of claim 1 whereby the copy number variants identified are fractional (i.e., non-integer levels) due to heterogeneity in the sample.

74. The method of claim 1 whereby enrichment of selected regions is performed.

IDT Exhibit 1020.4097

75. The method of claim 1 whereby copy number variation information is simultaneously extracted based on the methods described in claims 1, 64, 65, 66 and 67.

76. The method of claim 1 or 2, used with an initial step of polynucleotide bottlenecking to limit the number of starting initial copies or diversity of polynucleotides in the sample.

77. A method for detecting a rare mutation in a cell-free or a substantially cell free sample obtained from a subject comprising:

a. sequencing extracellular polynucleotides from a bodily sample of a subject, wherein each of the extracellular polynucleotides generate a plurality of sequencing reads;
b. filtering out reads that fail to meet a set quality threshold;
c. mapping sequence reads derived from the sequencing onto a reference sequence;
d. identifying a subset of mapped sequence reads that align with a variant of the reference sequence at each mappable base position;
e. for each mappable base position, calculating a ratio of (a) a number of mapped sequence reads that include a variant as compared to the reference sequence, to (b) a number of total sequence reads for each mappable base position;
f. normalizing the ratios or frequency of variance for each mappable base position and determining potential rare variant(s) or other genetic alteration(s); and
g. comparing the resulting number for each of the regions with potential rare variant(s) or mutation(s) to similarly derived numbers from a reference sample.

78. A method comprising:

a. providing at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides;
b. amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides;
c. sequencing a subset (including a proper subset) of the set of amplified progeny polynucleotides, to produce a set of sequencing reads; and

IDT Exhibit 1020.4098

d. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides.

79. The method of claim 78 wherein each polynucleotide in a set is mappable to a reference sequence.

80. The method of claim 78 comprising providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different mappable position in the reference sequence.

81. The method of claim 78 further comprising: e. analyzing the set of consensus sequences for each set of tagged parent molecules separately or in combination.

82. The method of claim 78 further comprising converting initial starting genetic material into the tagged parent polynucleotides.

83. The method of claim 82 wherein the initial starting genetic material comprises no more than 100 ng of polynucleotides.

84. The method of claim 82 comprising bottlenecking the initial starting genetic material prior to converting.

85. The method of claim 82 comprising converting the initial starting genetic material into tagged parent polynucleotides with a conversion efficiency of at least 10%, at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 80% or at least 90%.

86. The method of claim 82 wherein converting comprises any of blunt-end ligation, sticky end ligation, molecular inversion probes, PCR, ligation-based PCR, single strand ligation and single strand circularization.

87. The method of claim 82 wherein the initial starting genetic material is cell-free nucleic acid.

88. The method of claim 79 wherein a plurality of the sets map to different mappable positions in a reference sequence from the same genome.

IDT Exhibit 1020.4099

89. The method of claim 78 wherein each tagged parent polynucleotide in the set is uniquely tagged.

90. The method of claim 78 wherein each set of parent polynucleotides is mappable to a position in a reference sequence, and the polynucleotides in each set are not uniquely tagged.

91. The method of claim 78 wherein the generation of consensus sequences is based on information from the tag and/or at least one of (i) sequence information at the beginning (start) region of the sequence read, (ii) the end (stop) regions of the sequence read and (iii) the length of the sequence read.

92. The method of claim 78 comprising sequencing a subset of the set of amplified progeny polynucleotides sufficient to produce sequence reads for at least one progeny from of each of at least 20%, at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% at least 95%, at least 98%, at least 99%, at least 99.9% or at least 99.99% of unique polynucleotides in the set of tagged parent polynucleotides.

93. The method of claim 92 wherein the at least one progeny is a plurality of progeny, e.g., at least 2, at least 5 or at least 10 progeny.

94. The method of claim 78 wherein the number of sequence reads in the set of sequence reads is greater than the number of unique tagged parent polynucleotides in the set of tagged parent polynucleotides.

95. The method of claim 78 wherein the subset of the set of amplified progeny polynucleotides sequenced is of sufficient size so that any nucleotide sequence represented in the set of tagged parent polynucleotides at a percentage that is the same as the percentage per-base sequencing error rate of the sequencing platform used, has at least a 50%, at least a 60%, at least a 70%, at least a 80%, at least a 90% at least a 95%, at least a 98%, at least a 99%, at least a 99.9% or at least a 99.99% chance of being represented among the set of consensus sequences.

96. The method of claim 78 comprising enriching the set of amplified progeny polynucleotides for polynucleotides mapping to one or more selected mappable positions in a reference sequence by: (i) selective amplification of sequences from initial starting genetic material converted to tagged parent polynucleotides; (ii) selective amplification of tagged parent

IDT Exhibit 1020.4100

polynucleotides; (iii) selective sequence capture of amplified progeny polynucleotides; or (iv) selective sequence capture of initial starting genetic material.

97. The method of claim 81 wherein analyzing comprises normalizing a measure (e.g., number) taken from a set of consensus sequences against a measure taken from a set of consensus sequences from a control sample.

98. The method of claim 81 wherein analyzing comprises detecting mutations, rare mutations, indels, copy number variations, transversions, translocations, inversion, deletions, aneuploidy, partial aneuploidy, polyploidy, chromosomal instability, chromosomal structure alterations, gene fusions, chromosome fusions, gene truncations, gene amplification, gene duplications, chromosomal lesions, DNA lesions, abnormal changes in nucleic acid chemical modifications, abnormal changes in epigenetic patterns, abnormal changes in nucleic acid methylation infection or cancer.

99. The method of claim 78 wherein the polynucleotides comprise DNA, RNA, a combination of the two, or DNA plus RNA-derived cDNA.

100. The method of claim 82 wherein a certain subset of polynucleotides is selected for, or is enriched based on, polynucleotide length in base-pairs from the initial set of polynucleotides or from the amplified polynucleotides.

101. The method of claim 82 wherein analysis further comprises detection and monitoring of an abnormality or disease within an individual, such as, infection and/or cancer.

102. The method of claim 101 performed in combination with immune repertoire profiling.

103. The method of claim 78 wherein the polynucleotides are extracted from a sample selected from the group consisting of blood, plasma, serum, urine, saliva, mucosal excretions, sputum, stool, and tears.

104. The method of claim 78 wherein collapsing comprises detecting and/or correcting errors, nicks or lesions present in the sense or anti-sense strand of the tagged parent polynucleotides or amplified progeny polynucleotides.

IDT Exhibit 1020.4101

105. A method comprising detecting genetic variation in non-uniquely tagged initial starting genetic material with a sensitivity of at least 5%, at least 1%, at least 0.5%, at least 0.1% or at least 0.05%.

106. The method of claim 105 wherein the initial starting genetic material is provided in an amount less than 100 ng of nucleic acid, the genetic variation is copy number/heterozygosity variation and detecting is performed with sub-chromosomal resolution; e.g., at least 100 megabase resolution, at least 10 megabase resolution, at least 1 megabase resolution, at least 100 kilobase resolution, at least 10 kilobase resolution or at least 1 kilobase resolution.

107. The method of claim 81 comprising providing a plurality of sets of tagged parent polynucleotides, wherein each set is mappable to a different mappable position in a reference sequence.

108. The method of claim 107 wherein the mappable position in the reference sequence is the locus of a tumor marker and analyzing comprises detecting the tumor marker in the set of consensus sequences.

109. The method of claim 108 wherein the tumor marker is present in the set of consensus sequences at a frequency less than the error rate introduced at the amplifying step.

110. The method of claim 107 wherein the at least one set is a plurality of sets, and the mappable position of the reference sequence comprise a plurality of mappable positions in the reference sequence, each of which mappable position is the locus of a tumor marker.

111. The method of claim 107 wherein analyzing comprises detecting copy number variation of consensus sequences between at least two sets of parent polynucleotides.

112. The method of claim 107 wherein analyzing comprises detecting the presence of sequence variations compared with the reference sequences.

113. The method of claim 107 wherein analyzing comprises detecting the presence of sequence variations compared with the reference sequences and detecting copy number variation of consensus sequences between at least two sets of parent polynucleotides.

114. The method of claim 78 wherein collapsing comprises:

IDT Exhibit 1020.4102

i. grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide; and

ii. determining a consensus sequence based on sequence reads in a family.

115. A system comprising a computer readable medium for performing the following steps:

a. accepting at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides;
b. amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides;
c. sequencing a subset (including a proper subset) of the set of amplified progeny polynucleotides, to produce a set of sequencing reads;
d. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides and, optionally,
e. analyzing the set of consensus sequences for each set of tagged parent molecules.

116. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 10% of the individual's genome is sequenced.

117. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 20% of the individual's genome is sequenced.

118. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 30% of the individual's genome is sequenced.

IDT Exhibit 1020.4103

119. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 40% of the individual's genome is sequenced.

120. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 50% of the individual's genome is sequenced.

121. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 60% of the individual's genome is sequenced.

122. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 70% of the individual's genome is sequenced.

123. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 80% of the individual's genome is sequenced.

124. A method comprising detecting the presence or absence of genetic alteration or amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 90% of the individual's genome is sequenced.

125. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 10% of the individual's genome is sequenced.

IDT Exhibit 1020.4104

126. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 20% of the individual's genome is sequenced.

127. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 30% of the individual's genome is sequenced.

128. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 40% of the individual's genome is sequenced.

129. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 50% of the individual's genome is sequenced.

130. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 60% of the individual's genome is sequenced.

131. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 70% of the individual's genome is sequenced.

132. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 80% of the individual's genome is sequenced.

IDT Exhibit 1020.4105

133. A method comprising detecting the presence or absence of genetic alteration and amount of genetic variation in an individual, wherein the detecting is performed with the aid of sequencing of cell-free nucleic acid, wherein at least 90% of the individual's genome is sequenced.

134. The method of claims 116-133, wherein the genetic alteration is copy number variation or one or more rare mutations.

135. The method of claims 116-133, wherein the genetic variation comprises one or more causal variants and one or more polymorphisms.

136. The method of claims 116-133, wherein the genetic alteration and/or amount of genetic variation in the individual may be compared to a genetic alteration and/or amount of genetic variation in one or more individuals with a known disease.

137. The method of claims 116-133, wherein the genetic alteration and/or amount of genetic variation in the individual may be compared to a genetic alteration and/or amount of genetic variation in one or more individuals, without a disease.

138. The method of claim 116-133, wherein the cell-free nucleic acid is DNA.

139. The method of claim 116-133, wherein the cell-free nucleic acid is RNA.

140. The method of claim 116-133, wherein the cell-free nucleic acid is DNA and RNA.

141. The method of claim 136, wherein the disease is cancer or pre-cancer.

142. The method of claims 116-133, the method further comprising diagnosis or treatment of a disease.

143. A method comprising:

a. providing at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides;
b. amplifying the tagged parent polynucleotides in the set to produce a corresponding set of amplified progeny polynucleotides;

IDT Exhibit 1020.4106

c. sequencing a subset (including a proper subset) of the set of amplified progeny polynucleotides, to produce a set of sequencing reads;
d. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides; and
e. filtering out from among the consensus sequences those that fail to meet a quality threshold.

144. The method of claim 143 wherein the quality threshold considers a number of sequence reads from amplified progeny polynucleotides collapsed into a consensus sequence.

145. The method of claim 143 wherein the quality threshold considers a number of sequence reads from amplified progeny polynucleotides collapsed into a consensus sequence.

146. A system comprising a computer readable medium for performing the method of any of claims 143-145.

147. A method comprising:
a. providing at least one set of tagged parent polynucleotides, wherein each set maps to a different mappable position in a reference sequence in one or more genomes, and, for each set of tagged parent polynucleotides;
i. amplifying the first polynucleotides to produce a set of amplified polynucleotides;
ii. sequencing a subset of the set of amplified polynucleotides, to produce a set of sequencing reads; and
iii. collapsing the sequence reads by:
1. grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide.

148. The method of claim 147 wherein collapsing further comprises:
2. determining a quantitative measure of sequence reads in each family.

IDT Exhibit 1020.4107

149. The method of claim 148 further comprising:

b. determining a quantitative measure of unique families; and

c. based on (1) the quantitative measure of unique families and (2) the quantitative measure of sequence reads in each group, inferring a measure of unique tagged parent polynucleotides in the set.

150. The method of claim 149 wherein inferring is performed using statistical or probabilistic models.

151. The method of claim 149 wherein the at least one set is a plurality of sets.

152. The method of claim 151 further comprising correcting for amplification or representational bias between the two sets.

153. The method of claim 152 further comprising using a control or set of control samples to correct for amplification or representational biases between the two sets.

154. The method of claim 151 further comprising determining copy number variation between the sets.

155. The method of claim 149 further comprising:

d. determining a quantitative measure of polymorphic forms among the families; and

e. based on the determined quantitative measure of polymorphic forms, inferring a quantitative measure of polymorphic forms in the number of inferred unique tagged parent polynucleotides.

156. The method of claim 155 wherein polymorphic forms include but are not limited to: substitutions, insertions, deletions, inversions, microsatellite changes, transversions, translocations, fusions, methylation, hypermethylation, hydroxymethylation, acetylation, epigenetic variants, regulatory-associated variants or protein binding sites.

157. The method claim 149 wherein the sets derive from a common sample, and the method further comprises:

IDT Exhibit 1020.4108

d. inferring copy number variation for the plurality of sets based on a comparison of the inferred number of tagged parent polynucleotides in each set mapping to each of a plurality of mappable positions in a reference sequence.

158. The method of claim 157 wherein the original number of polynucleotides in each set is further inferred.

159. The method of claim 147, wherein at least a subset of the tagged parent polynucleotides in each set are non-uniquely tagged.

160. A system comprising a computer readable medium comprising machine-executable code that, upon execution by a computer processor, implements a method as in any one of claims 147-158.

161. A method of determining copy number variation in a sample that includes polynucleotides, the method comprising:

a. providing at least two sets of first polynucleotides, wherein each set maps to a different mappable position in a reference sequence in a genome, and, for each set of first polynucleotides;
    i. amplifying the polynucleotides to produce a set of amplified polynucleotides;
    ii. sequencing a subset of the set of amplified polynucleotides, to produce a set of sequencing reads;
    iii. grouping sequences reads sequenced from amplified polynucleotides into families, each family amplified from the same first polynucleotide in the set;
    iv. inferring a quantitative measure of families in the set; and
b. determining copy number variation by comparing the quantitative measure of families in each set.

162. A system comprising a computer readable medium comprising machine-executable code that, upon execution by a computer processor, implements the method of claim 161.

IDT Exhibit 1020.4109

163. A method of inferring frequency of sequence calls in a sample of polynucleotides comprising:

a. providing at least one set of first polynucleotides, wherein each set maps to a different mappable position in a reference sequence in one or more genomes, and, for each set of first polynucleotides;
   i. amplifying the first polynucleotides to produce a set of amplified polynucleotides;
   ii. sequencing a subset of the set of amplified polynucleotides, to produce a set of sequencing reads;
   iii. grouping the sequence reads into families, each family comprising sequence reads of amplified polynucleotides amplified from the same first polynucleotide;
b. inferring, for each set of first polynucleotides, a call frequency for one or more bases in the set of first polynucleotides, wherein inferring comprises:
   i. assigning, for each family, confidence score for each of a plurality of calls, the confidence score taking into consideration a frequency of the call among members of the family; and
   ii. estimating a frequency of the one or more calls taking into consideration the confidence scores of the one or more calls assigned to each family.

164. A system comprising a computer readable medium comprising machine-executable code that, upon execution by a computer processor, implements the method of claim 163.

165. A method of communicating sequence information about at least one individual polynucleotide molecule, comprising:

a. providing at least one individual polynucleotide molecule;
b. encoding sequence information in the at least one individual polynucleotide molecule to produce a signal;
c. passing at least part of the signal through a channel to produce a received signal comprising nucleotide sequence information about the at least one individual polynucleotide molecule, wherein the received signal comprises noise and/or distortion;

IDT Exhibit 1020.4110

d. decoding the received signal to produce a message comprising sequence information about the at least one individual polynucleotide molecule, wherein decoding reduces noise and/or distortion about each individual polynucleotide in the message; and
e. providing the message comprising sequence information about the at least one individual polynucleotide molecule to a recipient.

166. The method of claim 165 wherein the noise comprises incorrect nucleotide calls.

167. The method of claim 165 wherein distortion comprises uneven amplification of the individual polynucleotide molecule compared with other individual polynucleotide molecules.

168. The method of claim 167 wherein distortion results from amplification or sequencing bias.

169. The method of claim 165 wherein the at least one individual polynucleotide molecule is a plurality of individual polynucleotide molecules, and decoding produces a message about each molecule in the plurality.

170. The method of claim 165 wherein encoding comprises amplifying the at least one individual polynucleotide molecule, which has optionally been tagged, wherein the signal comprises a collection of amplified molecules.

171. The method of claim 165 wherein the channel comprises a polynucleotide sequencer and the received signal comprises sequence reads of a plurality of polynucleotides amplified from the at least one individual polynucleotide molecule.

172. The method of claim 165 wherein decoding comprises grouping sequence reads of amplified molecules amplified from each of the at least one individual polynucleotide molecules.

173. The method of claim 169 wherein the decoding consists of a probabilistic or statistical method of filtering the generated sequence signal.

174. A system comprising a computer readable medium comprising machine-executable code that, upon execution by a computer processor, implements a method of any of claims 165-173.

IDT Exhibit 1020.4111

175. The method of any of claims 143-145, 147-159 and 161, wherein the polynucleotides are derived from tumor genomic DNA or RNA.

176. The method of any of claims of 143-175 wherein the polynucleotides are derived from cell-free polynucleotides, exosomal polynucleotides, bacterial polynucleotides or viral polynucleotides.

177. The method of any of claims of 1-3 or 143-175 further comprising the detection and/or association of affected molecular pathways.

178. The method of any of claims of 1-3 or 143-175 further comprising serial monitoring of the health or disease state of an individual.

179. The method of any of claims of 1-3 or 143-175 whereby the phylogeny of a genome associated with a disease within an individual is inferred.

180. The method of any of claims of 1-3 or 143-175 wherein further comprising diagnosis, monitoring or treatment of a disease.

181. The claims of 180 where the treatment regimen is selected or modified based on detected polymorphic forms or CNVs or associated pathways.

182. The claims of 180 or 181 where the treatment comprises of a combination therapy.

183. The claim of 179 where the diagnosis further comprises localizing the disease using a radiographic technique, such as, a CT-Scan, PET-CT, MRI, Ultrasound, Ultraound with microbubbles, etc.

184. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

selecting predefined regions in a genome;

accessing sequence reads and enumerating number of sequence reads in the predefined regions;

normalizing the number of sequence reads across the predefined regions; and

determining percent of copy number variation in the predefined regions.

IDT Exhibit 1020.4112

185. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads;
b. filtering out reads that fail to meet a set threshold;
c. mapping sequence reads derived from the sequencing onto a reference sequence;
d. identifying a subset of mapped sequence reads that align with a variant of the reference sequence at each mappable base position;
e. for each mappable base position, calculating a ratio of (a) a number of mapped sequence reads that include a variant as compared to the reference sequence, to (b) a number of total sequence reads for each mappable base position;
f. normalizing the ratios or frequency of variance for each mappable base position and determining potential rare variant(s) or other genetic alteration(s); and
g. comparing the resulting number for each of the regions with potential rare variant(s) or mutation(s) to similarly derived numbers from a reference sample.

186. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; and
b. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides.

187. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides;

IDT Exhibit 1020.4113

b. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides; and

c. filtering out from among the consensus sequences those that fail to meet a quality threshold.

188. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; and

i. collapsing the sequence reads by:

1. grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide and, optionally,.

2. determining a quantitative measure of sequence reads in each family.

189. The computer readable medium of claim 188 wherein the executable code, upon execution by a computer processor, further performs the steps of:

b. determining a quantitative measure of unique families;

c. based on (1) the quantitative measure of unique families and (2) the quantitative measure of sequence reads in each group, inferring a measure of unique tagged parent polynucleotides in the set.

190. The computer readable medium of claim 189 wherein the executable code, upon execution by a computer processor, further performs the steps of:

d. determining a quantitative measure of polymorphic forms among the families; and

e. based on the determined quantitative measure of polymorphic forms, inferring a quantitative measure of polymorphic forms in the number of inferred unique tagged parent polynucleotides.

IDT Exhibit 1020.4114

191. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides grouping sequences reads sequenced from amplified polynucleotides into families, each family amplified from the same first polynucleotide in the set;
b. inferring a quantitative measure of families in the set;
c. determining copy number variation by comparing the quantitative measure of families in each set.

192. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides grouping the sequence reads into families, each family comprising sequence reads of amplified polynucleotides amplified from the same first polynucleotide;
b. inferring, for each set of first polynucleotides, a call frequency for one or more bases in the set of first polynucleotides, wherein inferring comprises:
c. assigning, for each family, confidence score for each of a plurality of calls, the confidence score taking into consideration a frequency of the call among members of the family; and
d. estimating a frequency of the one or more calls taking into consideration the confidence scores of the one or more calls assigned to each family.

193. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a received signal that comprises endoded sequence information from at least one individual polynucleotide molecule wherein the received signal comprises noise and/or distortion;
b. decoding the received signal to produce a message comprising sequence information about the at least one individual polynucleotide molecule, wherein

IDT Exhibit 1020.4115

decoding reduces noise and/or distortion about each individual polynucleotide in the message; and

c. writing the message comprising sequence information about the at least one individual polynucleotide molecule to a computer file.

194. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides;

b. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides; and

c. filtering out from among the consensus sequences those that fail to meet a quality threshold.

195. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides; and

b. collapsing the sequence reads by:

i. grouping sequences reads sequenced from amplified progeny polynucleotides into families, each family amplified from the same tagged parent polynucleotide; and

ii. optionally, determining a quantitative measure of sequence reads in each family.

196. The computer readable medium of claim 195 wherein the executable code, upon execution by a computer processor, further performs the steps of:

c. determining a quantitative measure of unique families;

IDT Exhibit 1020.4116

d. based on (1) the quantitative measure of unique families and (2) the quantitative measure of sequence reads in each group, inferring a measure of unique tagged parent polynucleotides in the set.

197. The computer readable medium of claim 196 wherein the executable code, upon execution by a computer processor, further performs the steps of:

e. determining a quantitative measure of polymorphic forms among the families; and

f. based on the determined quantitative measure of polymorphic forms, inferring a quantitative measure of polymorphic forms in the number of inferred unique tagged parent polynucleotides.

198. The computer readable medium of claim 196 wherein the executable code, upon execution by a computer processor, further performs the steps of:

e. inferring copy number variation for the plurality of sets based on a comparison of the inferred number of tagged parent polynucleotides in each set mapping to each of a plurality of reference sequences.

199. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides;

b. grouping sequences reads sequenced from amplified polynucleotides into families, each family amplified from the same first polynucleotide in the set;

c. inferring a quantitative measure of families in the set; and

d. determining copy number variation by comparing the quantitative measure of families in each set.

200. A computer readable medium comprising non-transitory machine-executable code that, upon execution by a computer processor, implements a method, the method comprising:

a. accessing a data file comprising a plurality of sequencing reads, wherein the sequence reads derive from a set of progeny polynucleotides amplified from at least one set of tagged parent polynucleotides grouping the sequence reads into

IDT Exhibit 1020.4117

families, each family comprising sequence reads of amplified polynucleotides amplified from the same first polynucleotide; and

b. inferring, for each set of first polynucleotides, a call frequency for one or more bases in the set of first polynucleotides, wherein inferring comprises:

i. assigning, for each family, confidence score for each of a plurality of calls, the confidence score taking into consideration a frequency of the call among members of the family; and

ii. estimating a frequency of the one or more calls taking into consideration the confidence scores of the one or more calls assigned to each family.

201. A composition comprising between 100 and 100,000 human haploid genome equivalents of cfDNA polynucleotides, wherein the polynucleotides are tagged with between 2 and 1,000,000 unique identifiers.

202. The composition of claim 201 comprising between 1000 and 50,000 haploid human genome equivalents of cfDNA polynucleotides, wherein the polynucleotides are tagged with between 2 and 1,000 unique identifiers.

203. The composition of claim 201 wherein the unique identifiers comprise nucleotide barcodes.

204. A method comprising:

a. providing a sample comprising between 100 and 100,000 haploid human genome equivalents of cfDNA polynucleotides; and

b. tagging the polynucleotides with between 2 and 1,000,000 unique identifiers.

205. A method comprising:

a. providing a sample comprising a plurality of human haploid genome equivalents of fragmented polynucleotides;

b. determining z, wherein z is a measure of central tendency (e.g., mean, median or mode) of expected number of duplicate polynucleotides starting at any position in the genome, wherein duplicate polynucleotides have the same start and stop positions; and

IDT Exhibit 1020.4118

c. tagging polynucleotides in sample with n unique identifiers, wherein n is between 2 and 100,000*z, 2 and 10,000*z, 2 and 1,000*z or 2 and 100*z.

206. A method comprising:

a. providing at least one set of tagged parent polynucleotides, and for each set of tagged parent polynucleotides;

b. producing a plurality of sequence reads for each tagged parent polynucleotide in the set to produce a set of sequencing reads; and

c. collapsing the set of sequencing reads to generate a set of consensus sequences, each consensus sequence corresponding to a unique polynucleotide among the set of tagged parent polynucleotides.

IDT Exhibit 1020.4119

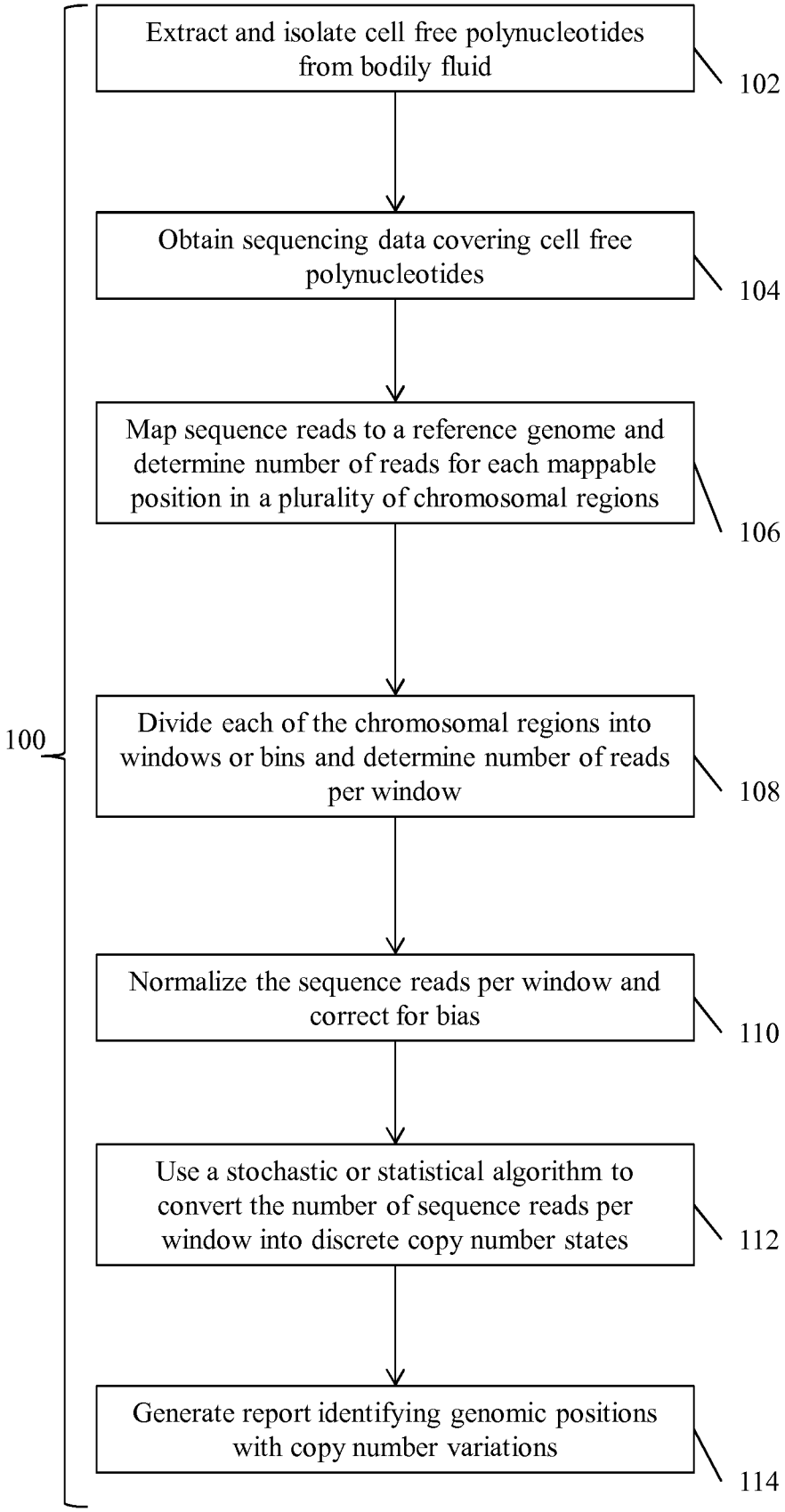

Extract and isolate cell free polynucleotides from bodily fluid
102
Obtain sequencing data covering cell free polynucleotides
104
Map sequence reads to a reference genome and determine number of reads for each mappable position in a plurality of chromosomal regions
106
Divide each of the chromosomal regions into windows or bins and determine number of reads per window
108
Normalize the sequence reads per window and correct for bias
110
Use a stochastic or statistical algorithm to convert the number of sequence reads per window into discrete copy number states
112
Generate report identifying genomic positions with copy number variations
114
100

Fig. 1

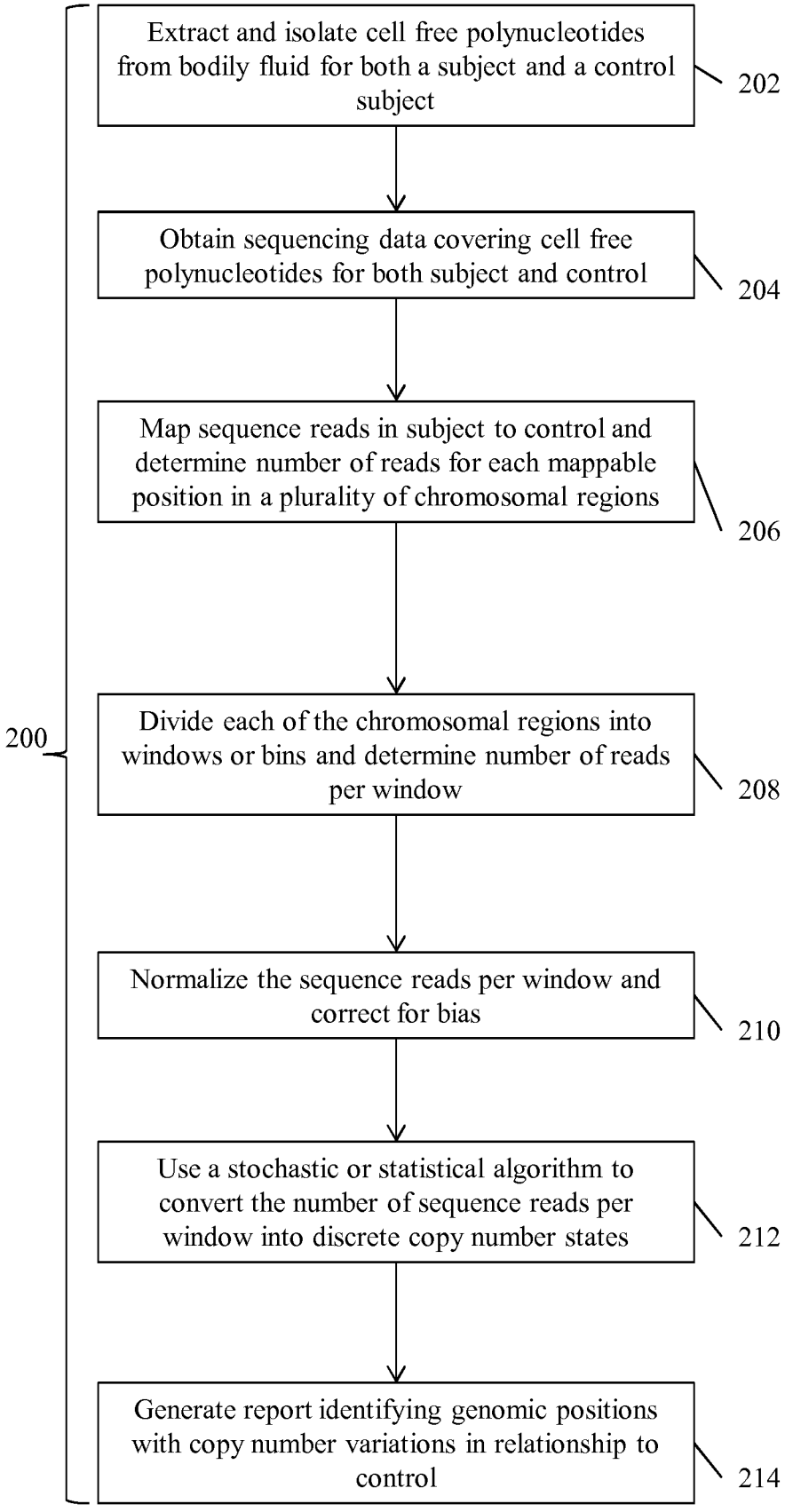

Extract and isolate cell free polynucleotides from bodily fluid for both a subject and a control subject
202
Obtain sequencing data covering cell free polynucleotides for both subject and control
204
Map sequence reads in subject to control and determine number of reads for each mappable position in a plurality of chromosomal regions
206
Divide each of the chromosomal regions into windows or bins and determine number of reads per window
208
Normalize the sequence reads per window and correct for bias
210
Use a stochastic or statistical algorithm to convert the number of sequence reads per window into discrete copy number states
212
Generate report identifying genomic positions with copy number variations in relationship to control
214
200

Fig. 2

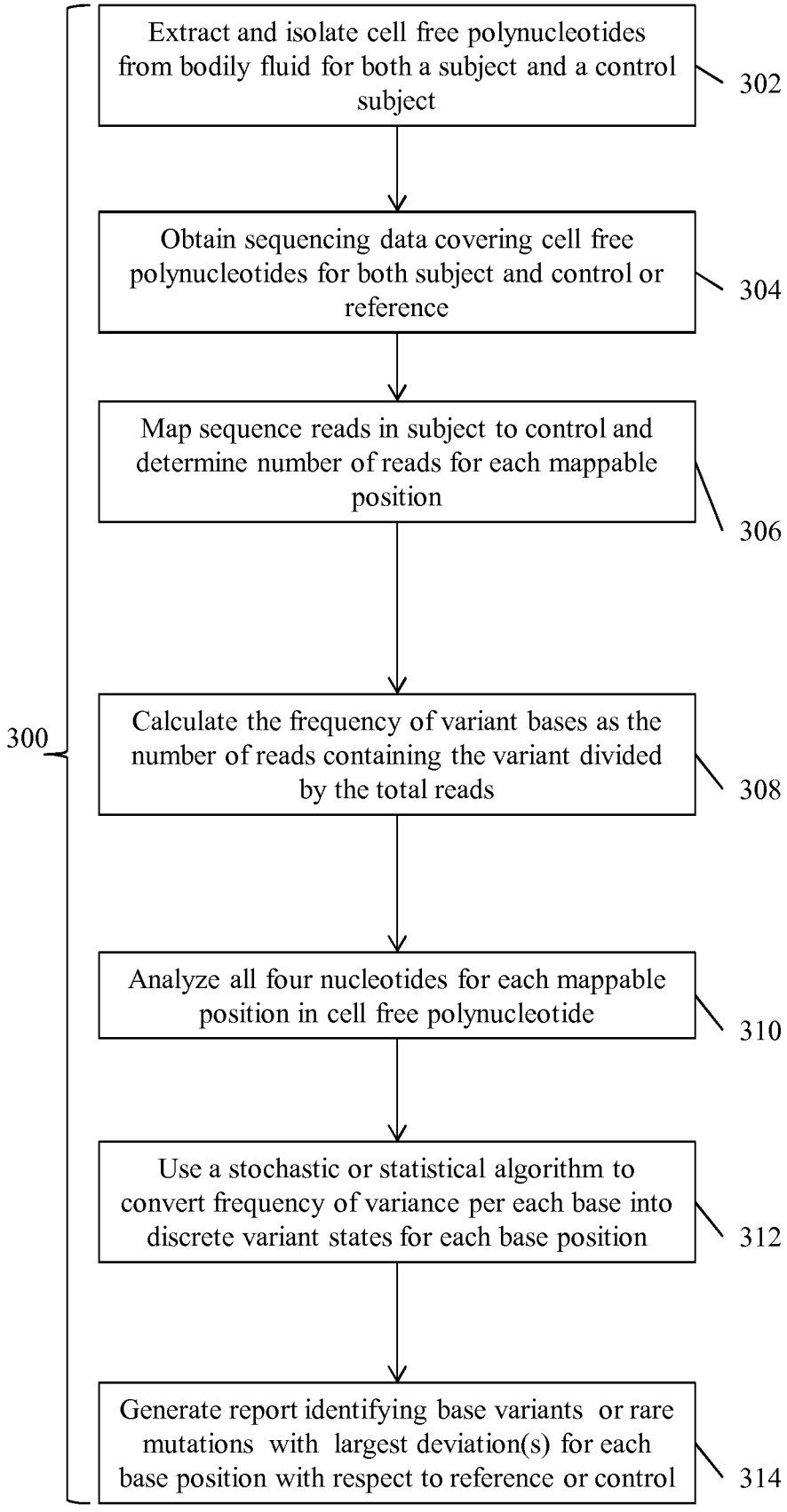
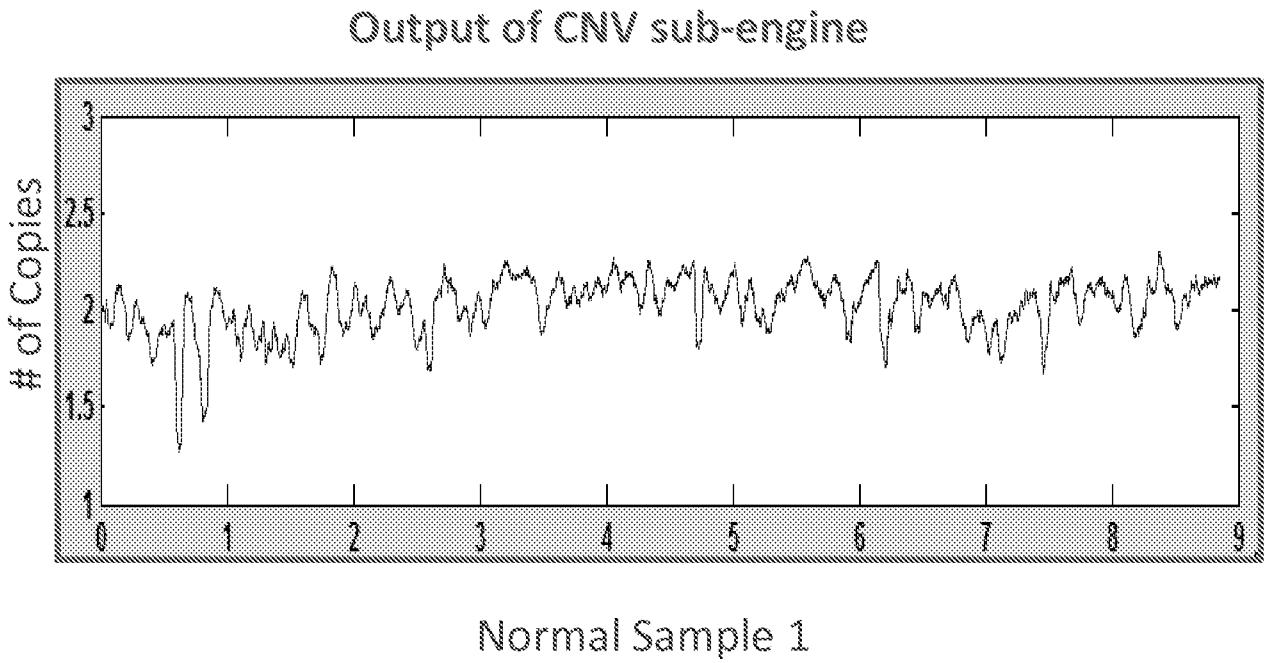
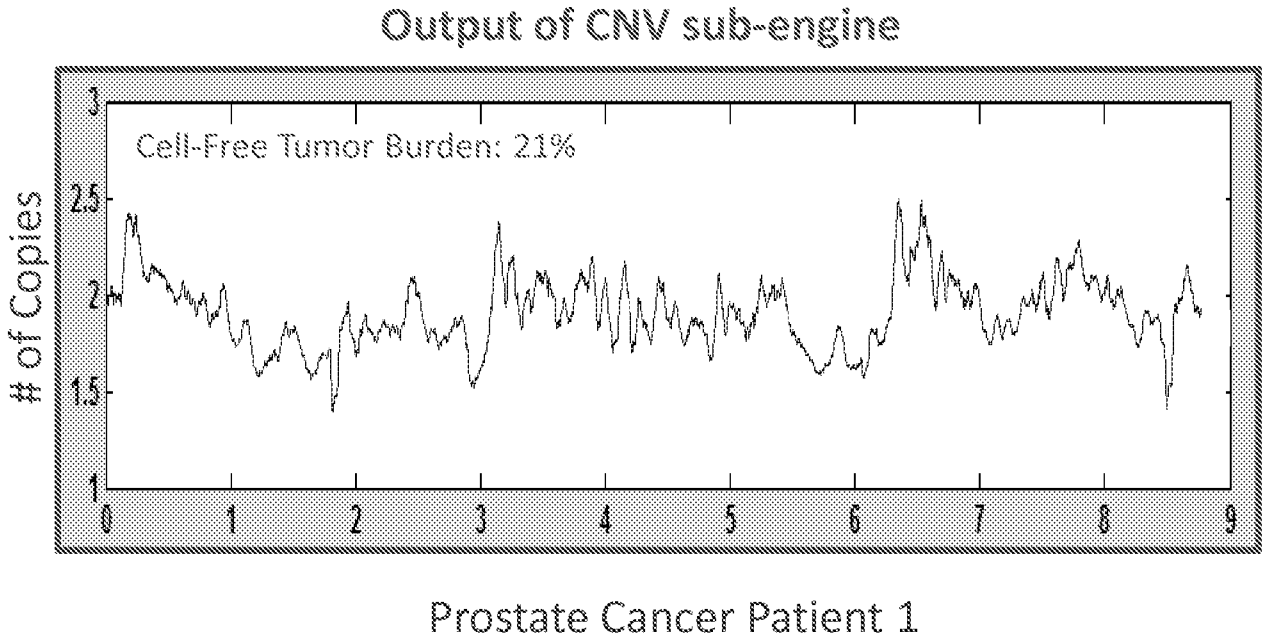

Extract and isolate cell free polynucleotides from bodily fluid for both a subject and a control subject
302
Obtain sequencing data covering cell free polynucleotides for both subject and control or reference
304
Map sequence reads in subject to control and determine number of reads for each mappable position
306
Calculate the frequency of variant bases as the number of reads containing the variant divided by the total reads
308
Analyze all four nucleotides for each mappable position in cell free polynucleotide
310
Use a stochastic or statistical algorithm to convert frequency of variance per each base into discrete variant states for each base position
312
Generate report identifying base variants or rare mutations with largest deviation(s) for each base position with respect to reference or control
314
300

Fig. 3

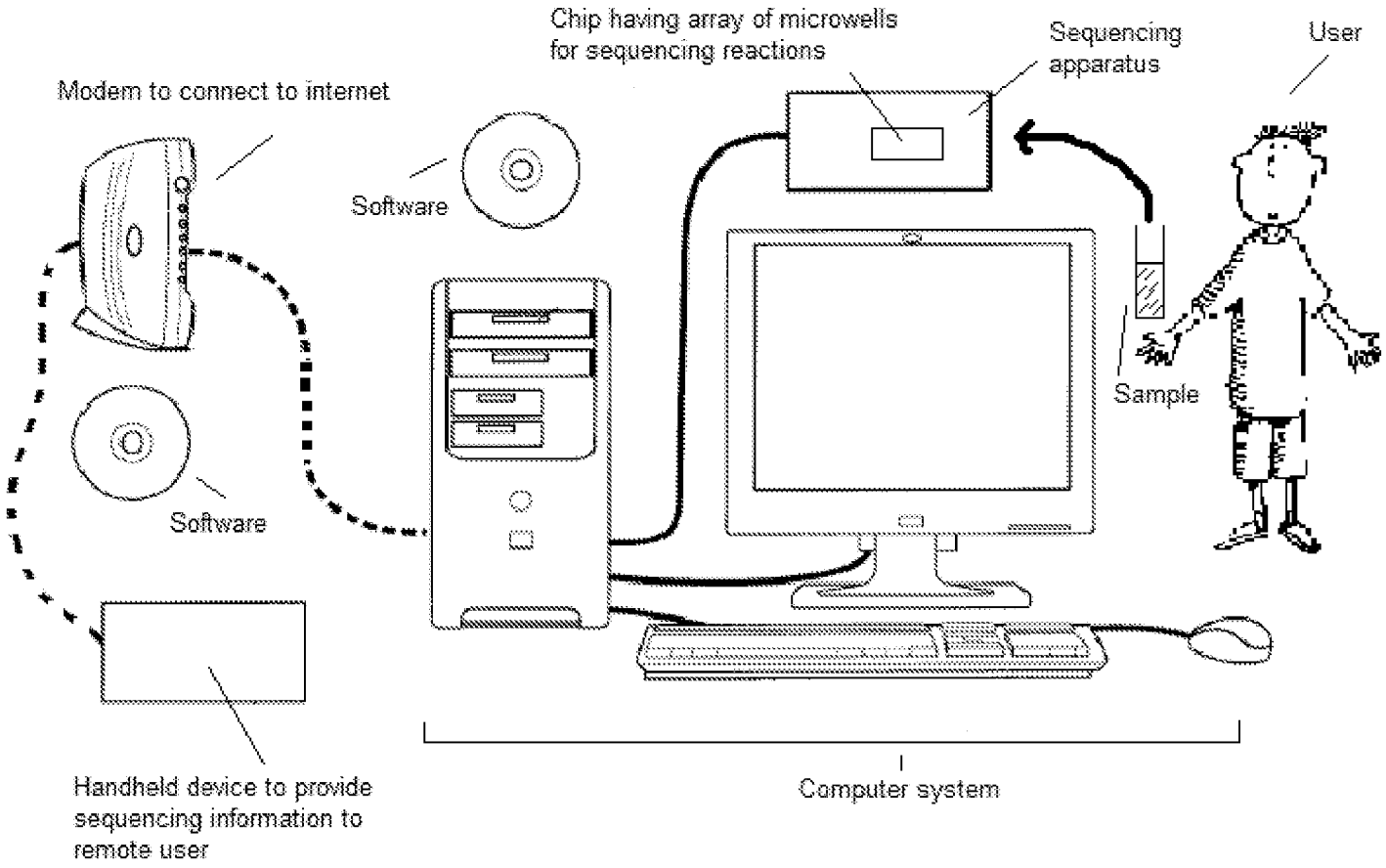

Output of CNV sub-engine
of Copies
3
2.5
2
1.5
1
0
1
2
3
4
5
6
7
8
9
Normal Sample 1

A


Output of CNV sub-engine
Cell-Free Tumor Burden: 21%
of Copies
3
2.5
2
1.5
1
0
1
2
3
4
5
6
7
8
9
Prostate Cancer Patient 1

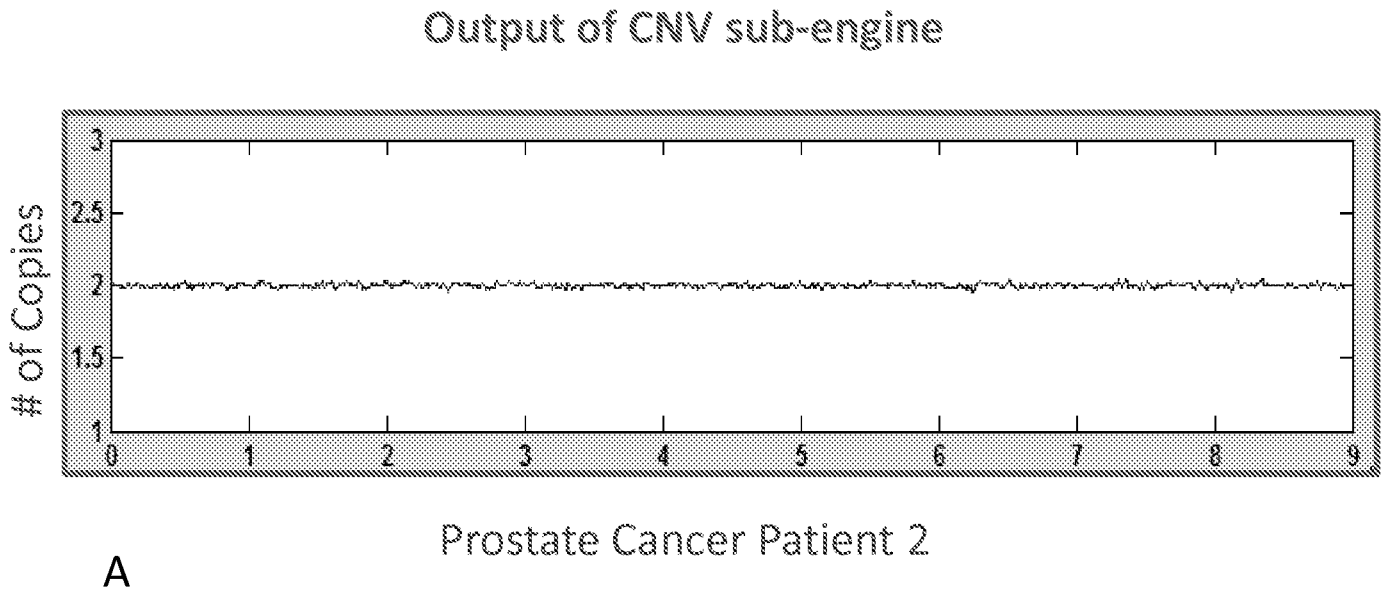
B

Fig. 4

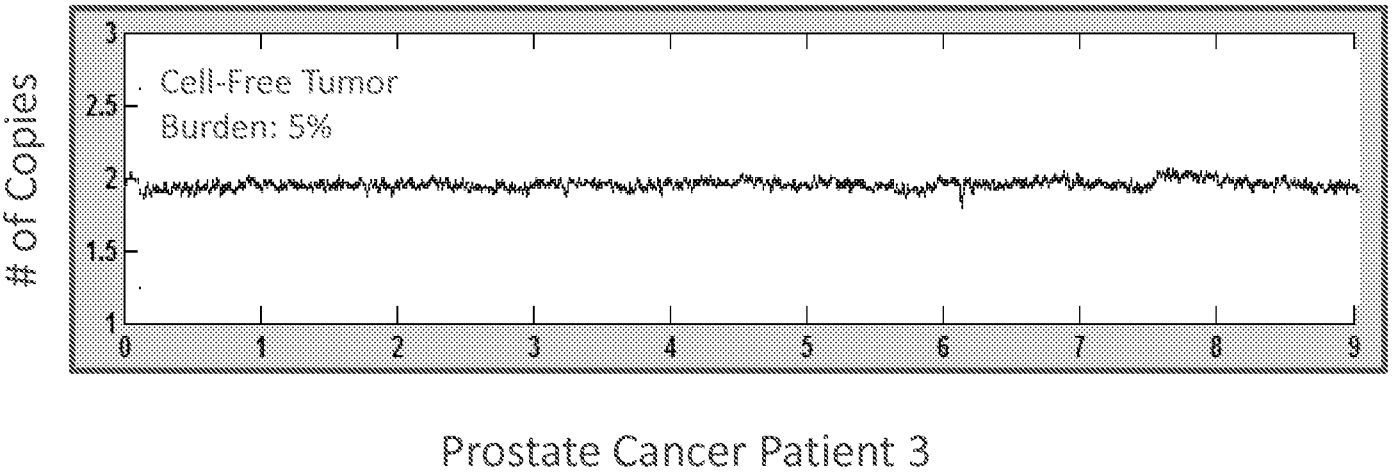

Chip having array of microwells for sequencing reactions
Sequencing apparatus
User
Modem to connect to internet
Software
Sample
Software
Handheld device to provide sequencing information to remote user
Computer system

C

Fig. 4

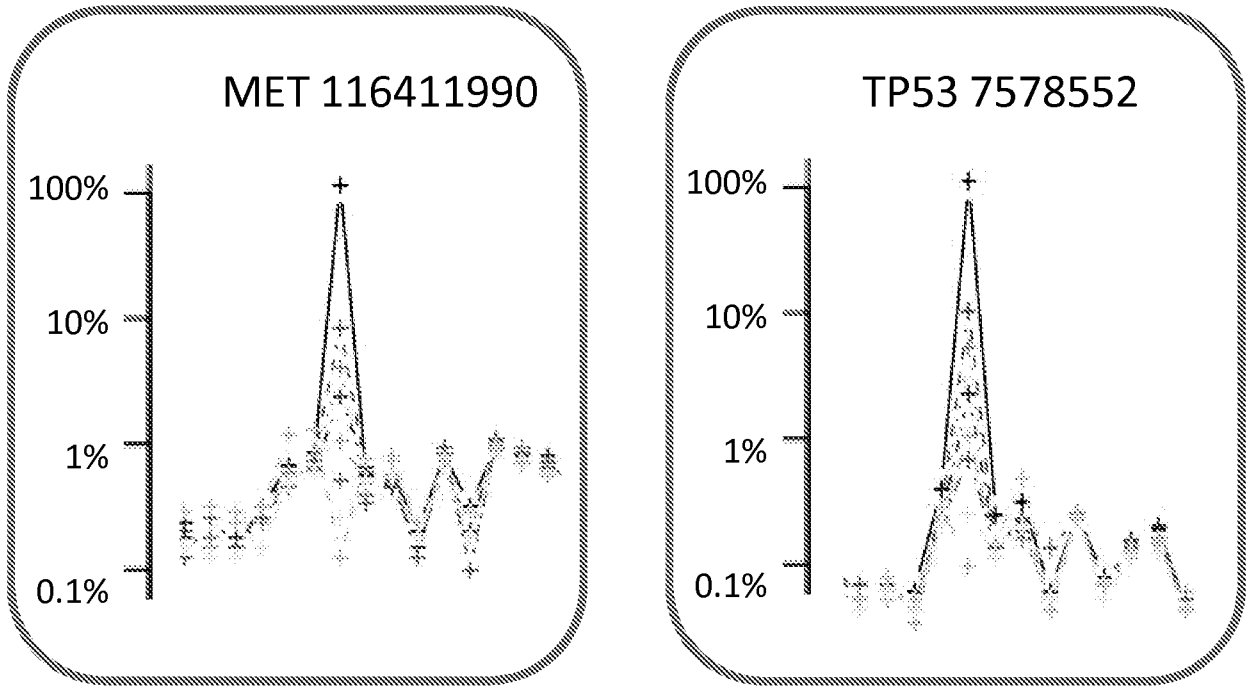

Output of CNV sub-engine
of Copies
3
2.5
2
1.5
1
0
1
2
3
4
5
6
7
8
9
Prostate Cancer Patient 2

A

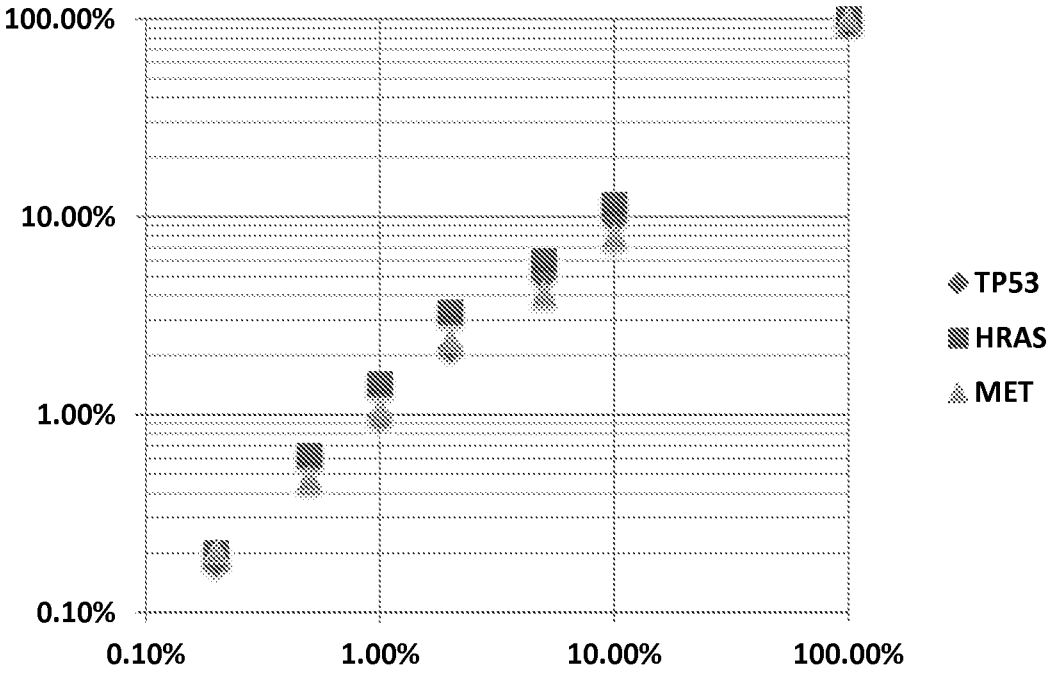

of Copies
Cell-Free Tumor
Burden: 5%
3
2.5
2
1.5
1
0
1
2
3
4
5
6
7
8
9
Prostate Cancer Patient 3

B

Fig. 5

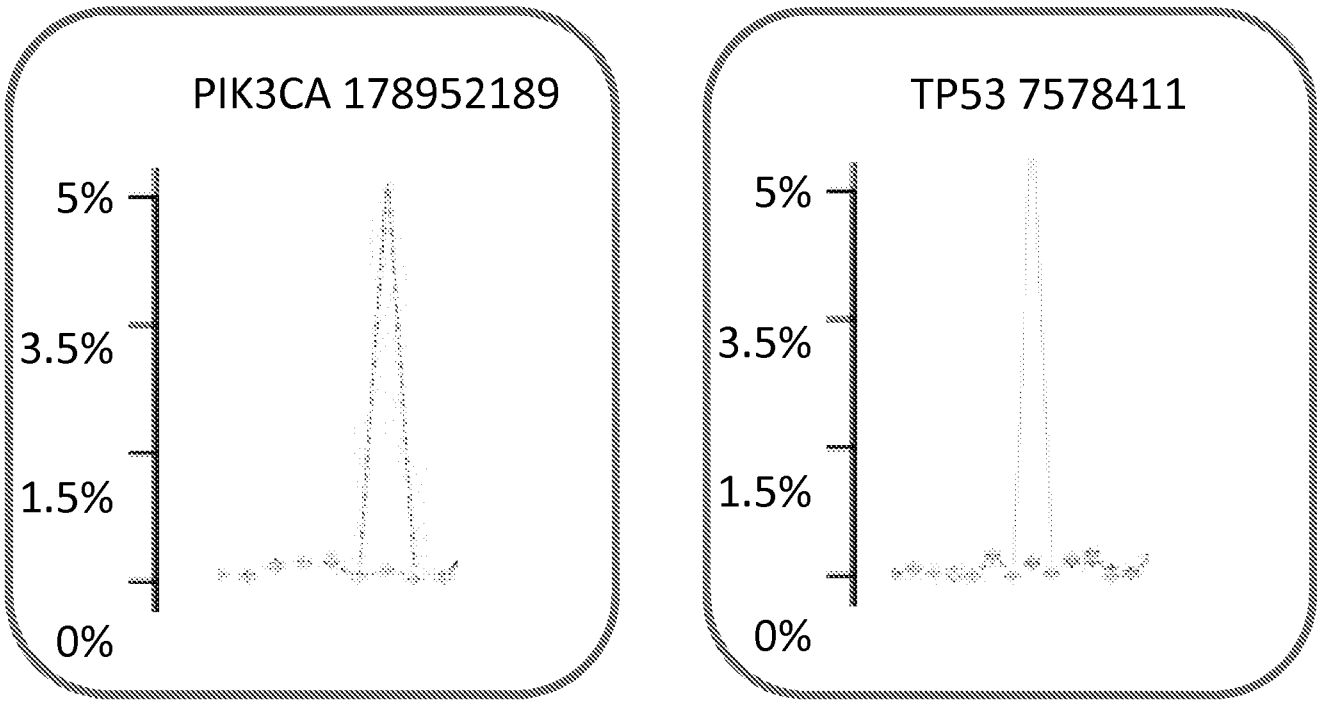

MET 116411990
100%
10%
1%
0.1%
TP53 7578552
100%
10%
1%
0.1%

A

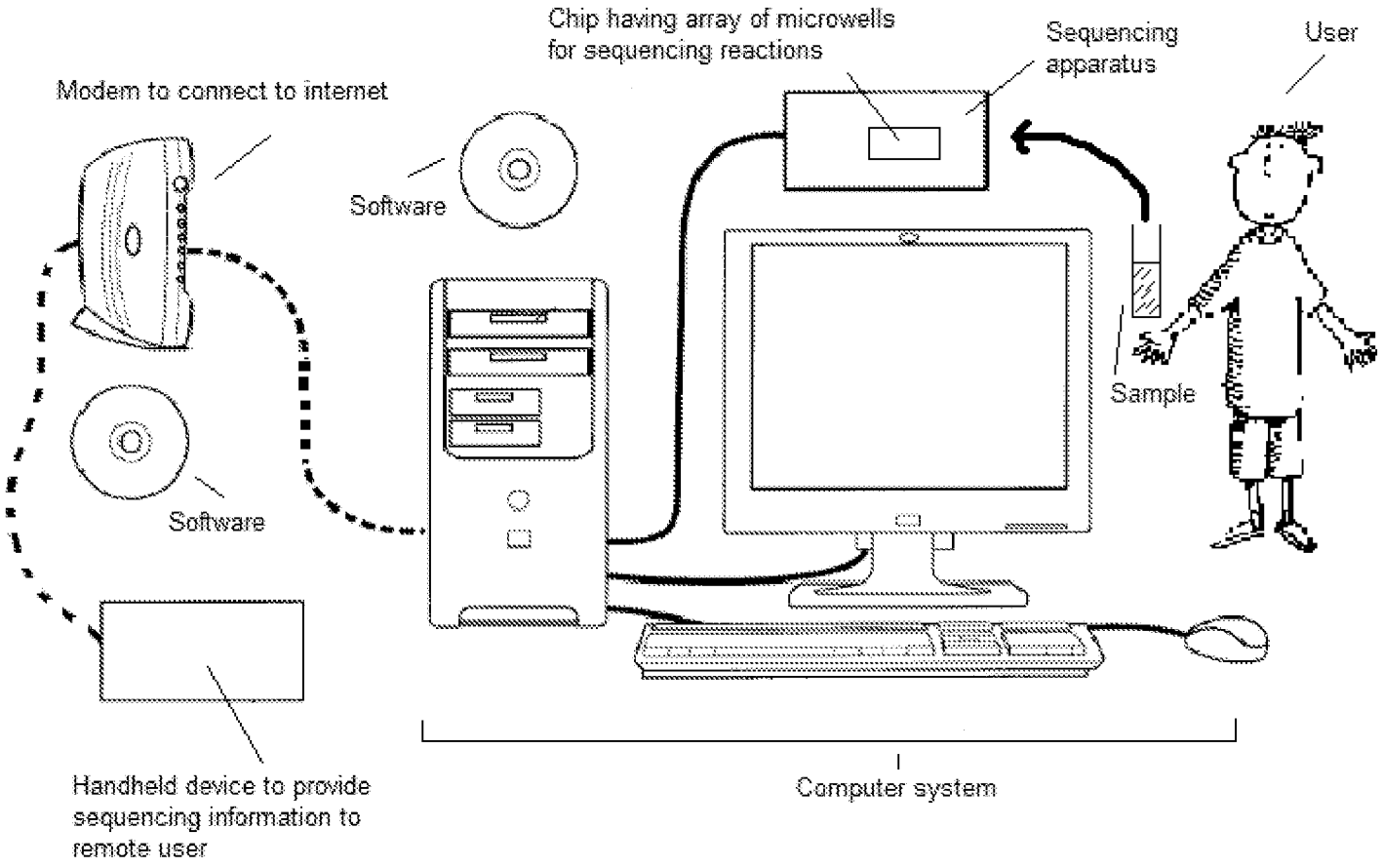

100.00%
10.00%
1.00%
0.10%
0.10%
1.00%
10.00%
100.00%
TP53
HRAS
MET

B

Fig. 6


PIK3CA 178952189
5%
3.5%
1.5%
0%
TP53 7578411
5%
3.5%
1.5%
0%

A


Chip having array of microwells for sequencing reactions
Sequencing apparatus
User
Modem to connect to internet
Software
Sample
Software
Handheld device to provide sequencing information to remote user
Computer system

B

Fig. 7

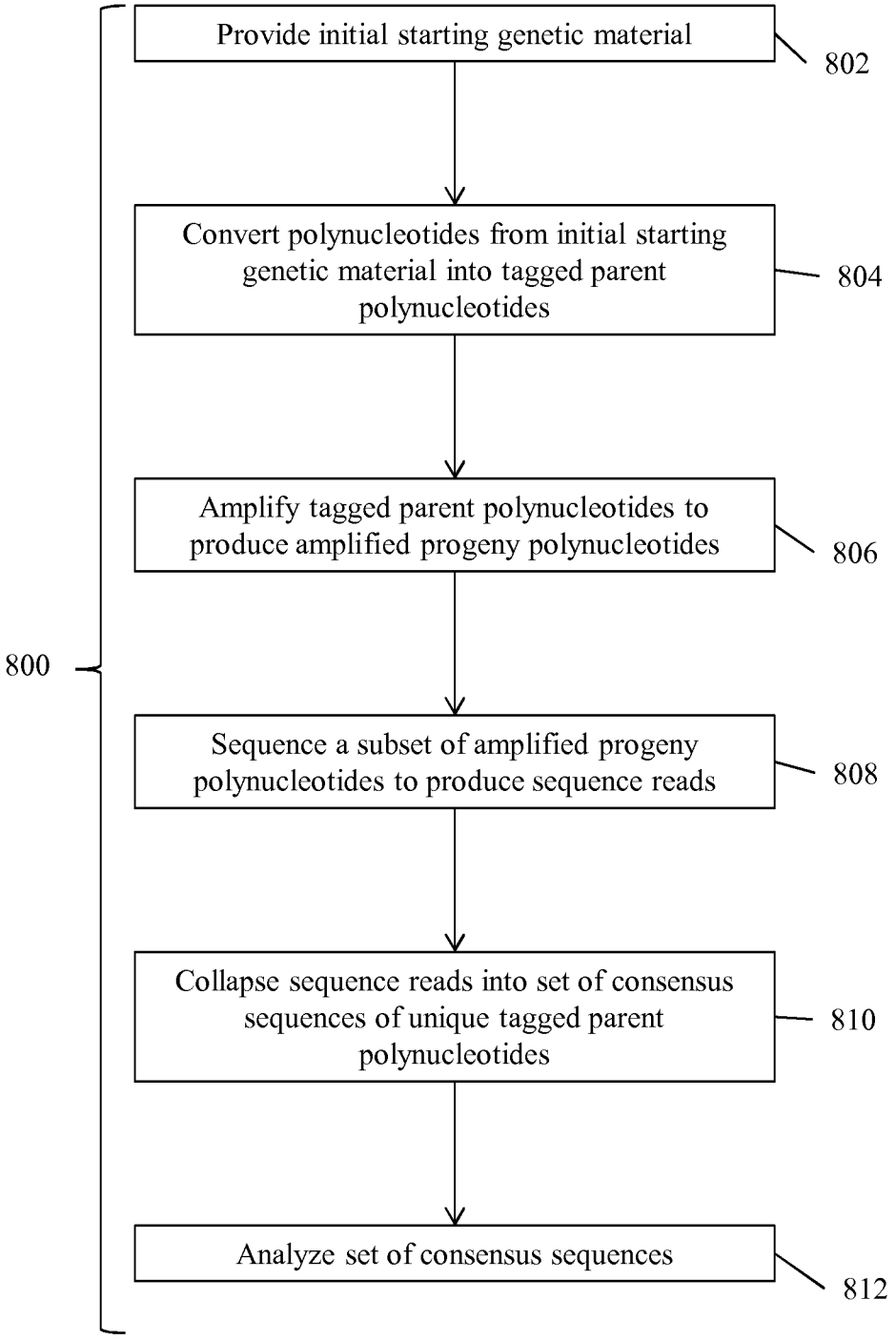

Provide initial starting genetic material
802
Convert polynucleotides from initial starting genetic material into tagged parent polynucleotides
804
Amplify tagged parent polynucleotides to produce amplified progeny polynucleotides
806
800
Sequence a subset of amplified progeny polynucleotides to produce sequence reads
808
Collapse sequence reads into set of consensus sequences of unique tagged parent polynucleotides
810
Analyze set of consensus sequences
812

Fig. 8

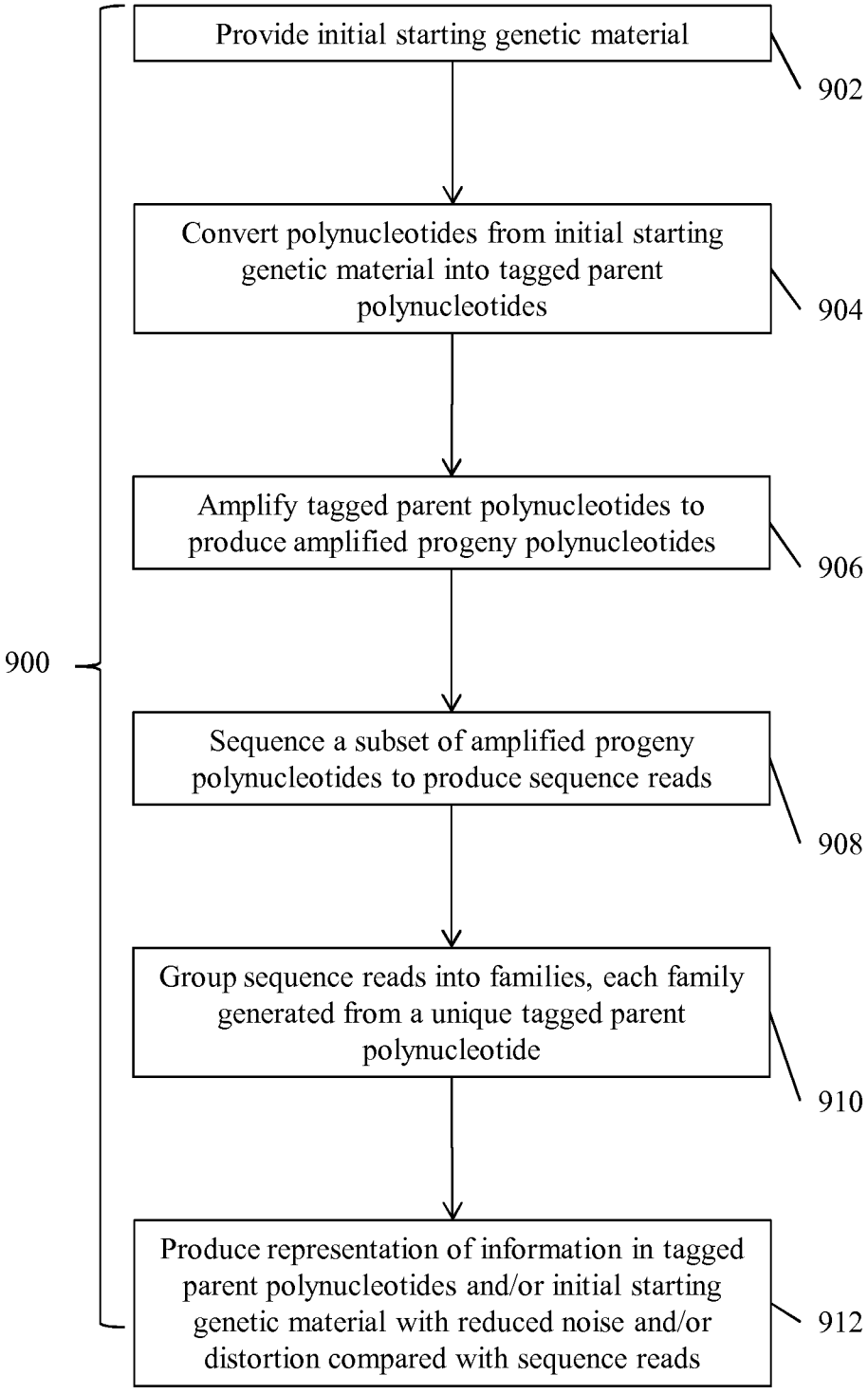

Provide initial starting genetic material
902
Convert polynucleotides from initial starting genetic material into tagged parent polynucleotides
904
Amplify tagged parent polynucleotides to produce amplified progeny polynucleotides
906
900
Sequence a subset of amplified progeny polynucleotides to produce sequence reads
908
Group sequence reads into families, each family generated from a unique tagged parent polynucleotide
910
Produce representation of information in tagged parent polynucleotides and/or initial starting genetic material with reduced noise and/or distortion compared with sequence reads
912

Fig. 9

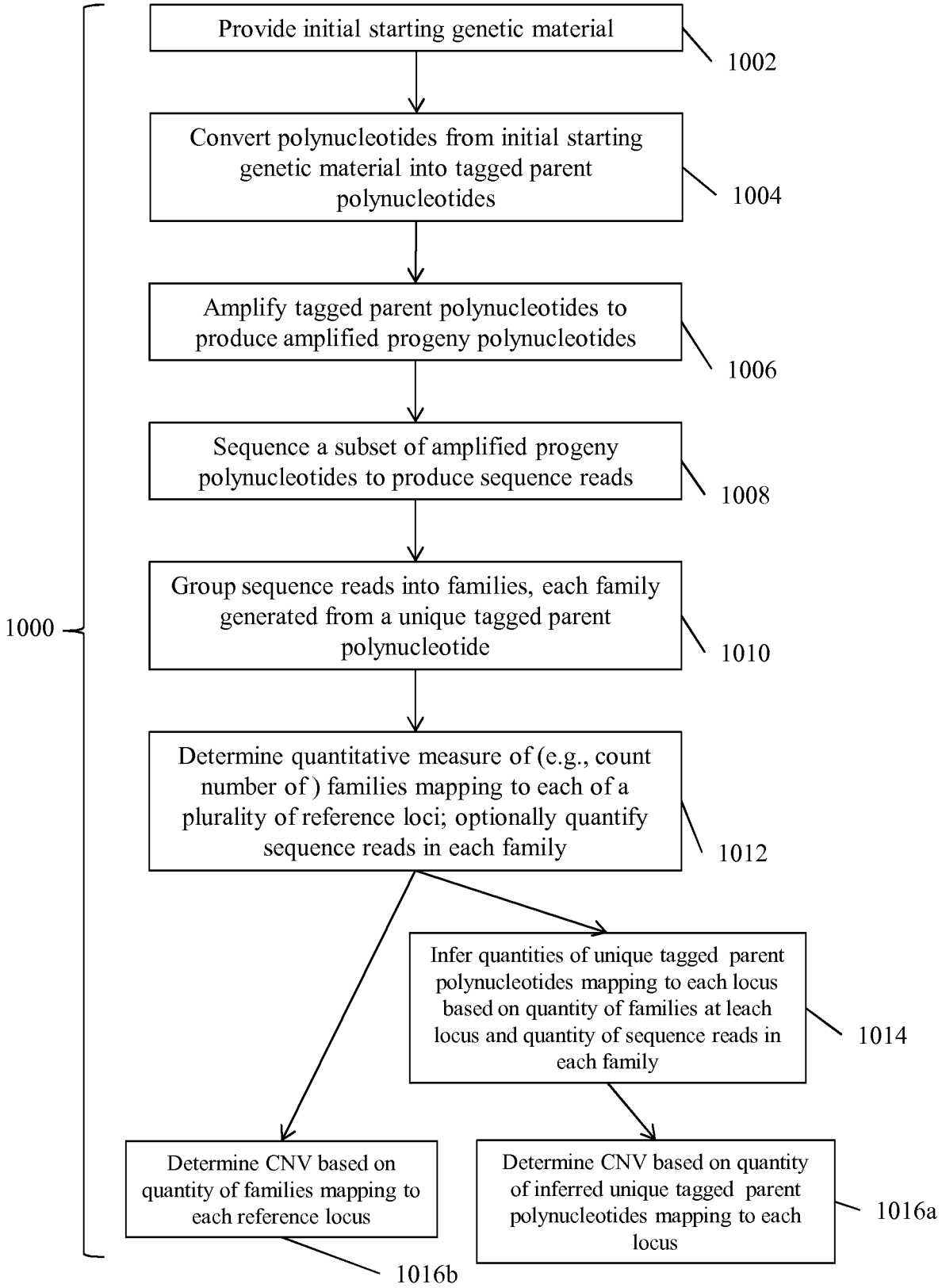

Provide initial starting genetic material
1002
Convert polynucleotides from initial starting genetic material into tagged parent polynucleotides
1004
Amplify tagged parent polynucleotides to produce amplified progeny polynucleotides
1006
Sequence a subset of amplified progeny polynucleotides to produce sequence reads
1008
Group sequence reads into families, each family generated from a unique tagged parent polynucleotide
1010
1000
Determine quantitative measure of (e.g., count number of) families mapping to each of a plurality of reference loci; optionally quantify sequence reads in each family
1012
Infer quantities of unique tagged parent polynucleotides mapping to each locus based on quantity of families at leach locus and quantity of sequence reads in each family
1014
Determine CNV based on quantity of families mapping to each reference locus
1016b
Determine CNV based on quantity of inferred unique tagged parent polynucleotides mapping to each locus
1016a

Fig. 10

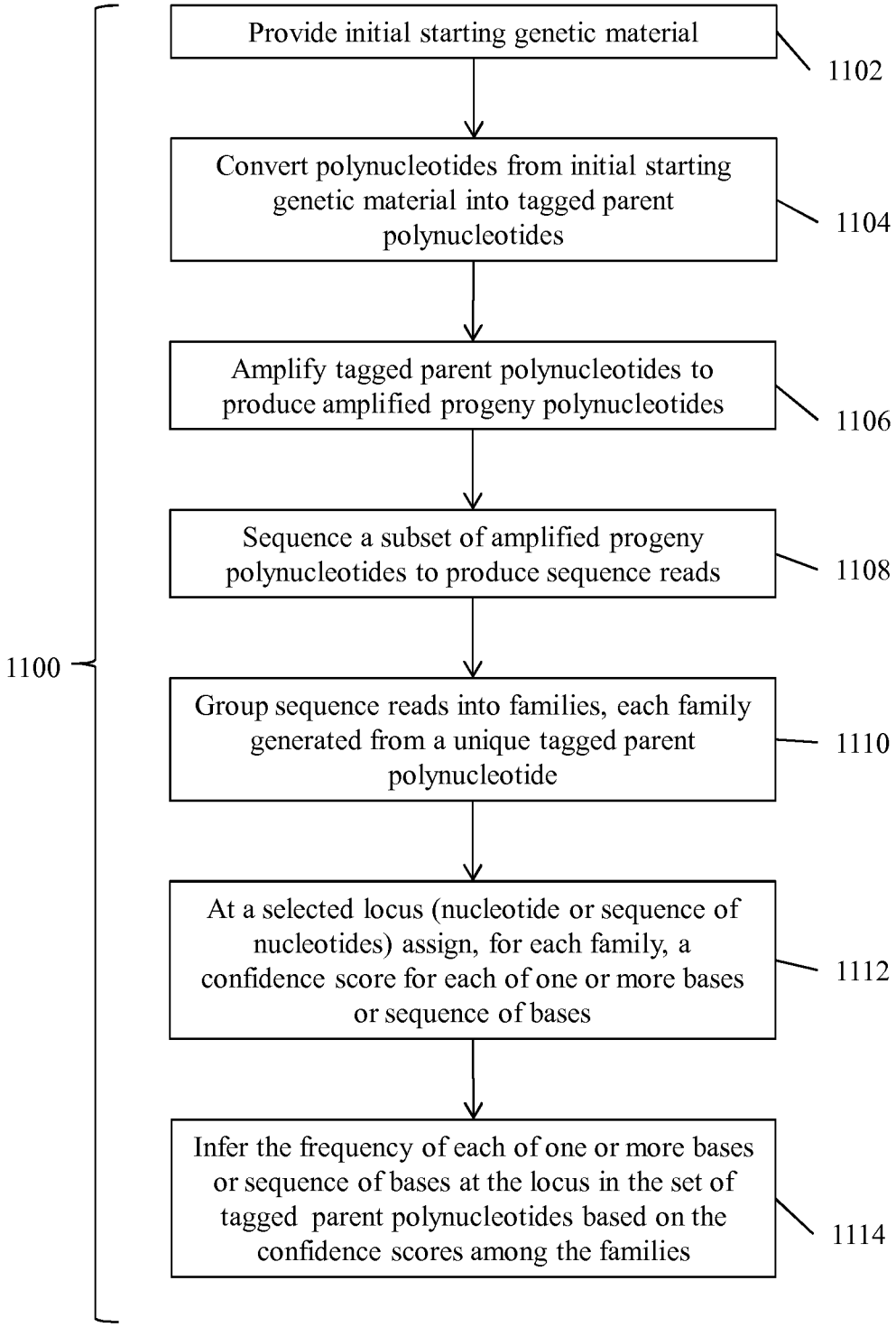

Provide initial starting genetic material
1102
Convert polynucleotides from initial starting genetic material into tagged parent polynucleotides
1104
Amplify tagged parent polynucleotides to produce amplified progeny polynucleotides
1106
Sequence a subset of amplified progeny polynucleotides to produce sequence reads
1108
1100
Group sequence reads into families, each family generated from a unique tagged parent polynucleotide
1110
At a selected locus (nucleotide or sequence of nucleotides) assign, for each family, a confidence score for each of one or more bases or sequence of bases
1112
Infer the frequency of each of one or more bases or sequence of bases at the locus in the set of tagged parent polynucleotides based on the confidence scores among the families
1114

Fig. 11

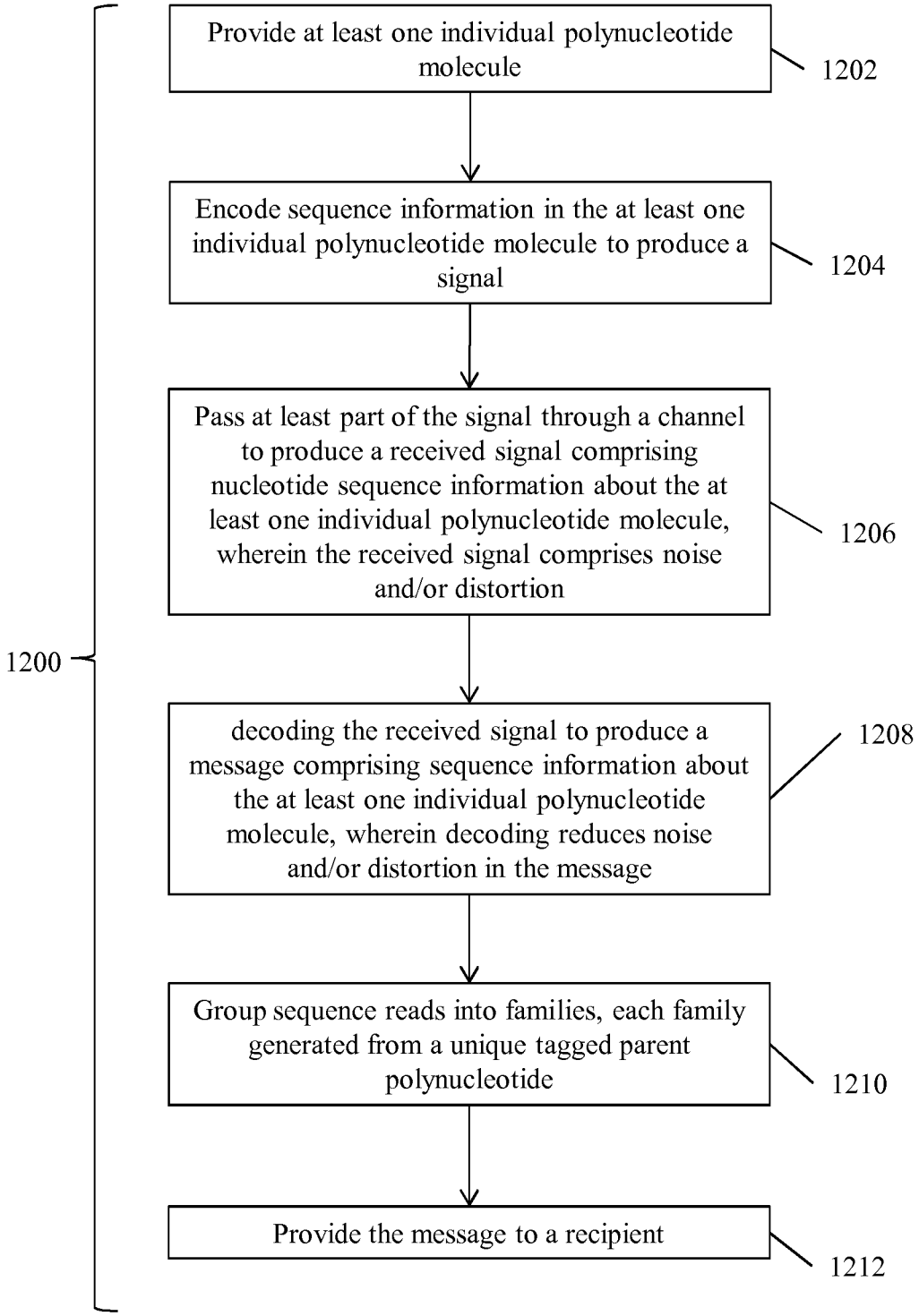

Provide at least one individual polynucleotide molecule
1202
Encode sequence information in the at least one individual polynucleotide molecule to produce a signal
1204
Pass at least part of the signal through a channel to produce a received signal comprising nucleotide sequence information about the at least one individual polynucleotide molecule, wherein the received signal comprises noise and/or distortion
1206
1200
decoding the received signal to produce a message comprising sequence information about the at least one individual polynucleotide molecule, wherein decoding reduces noise and/or distortion in the message
1208
Group sequence reads into families, each family generated from a unique tagged parent polynucleotide
1210
Provide the message to a recipient
1212

Fig. 12

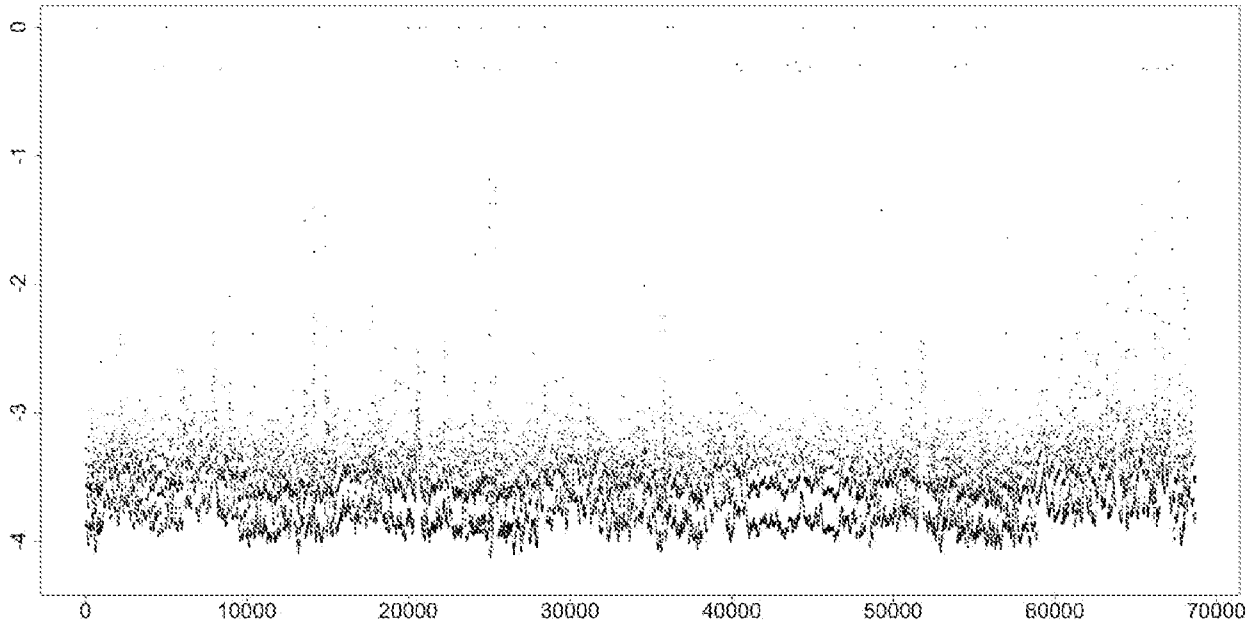


Fig. 13A

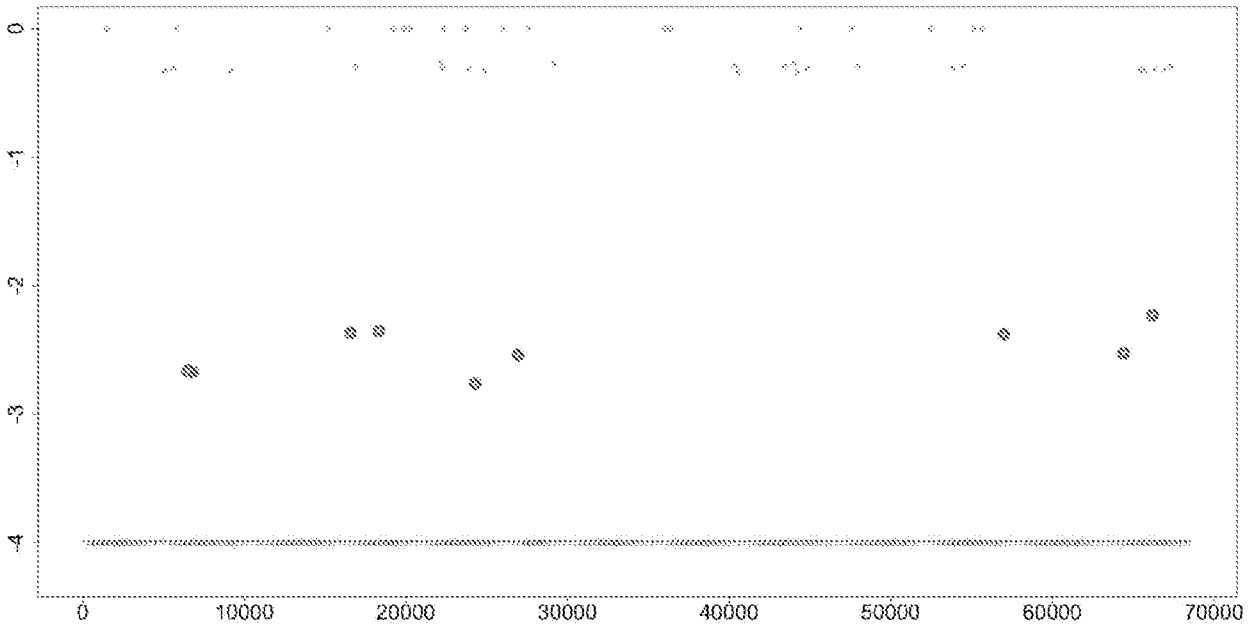


Fig. 13B

Fig. 13

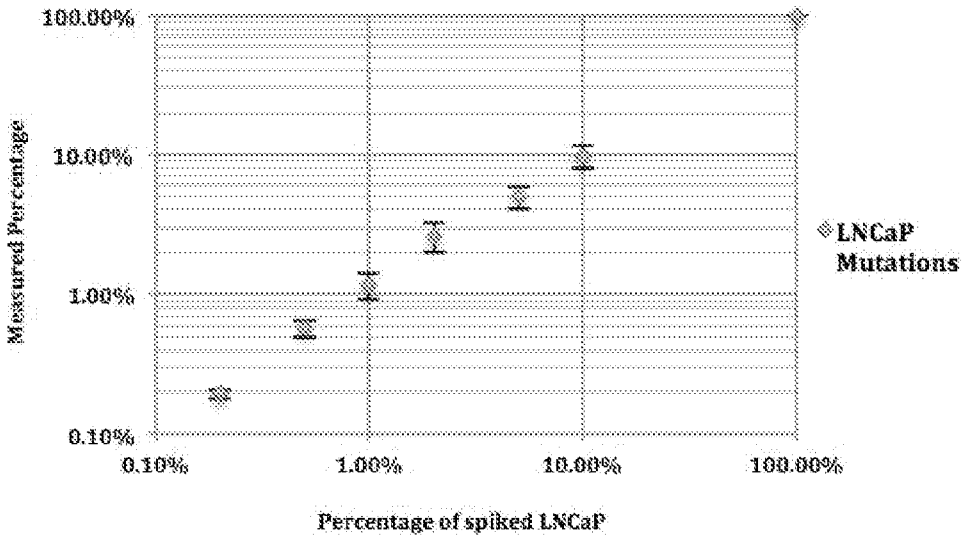

100.00%
10.00%
1.00%
0.10%
Measured Percentage
0.10%
1.00%
10.00%
100.00%
Percentage of spiked LNCaP
LNCaP Mutations

Fig. 14

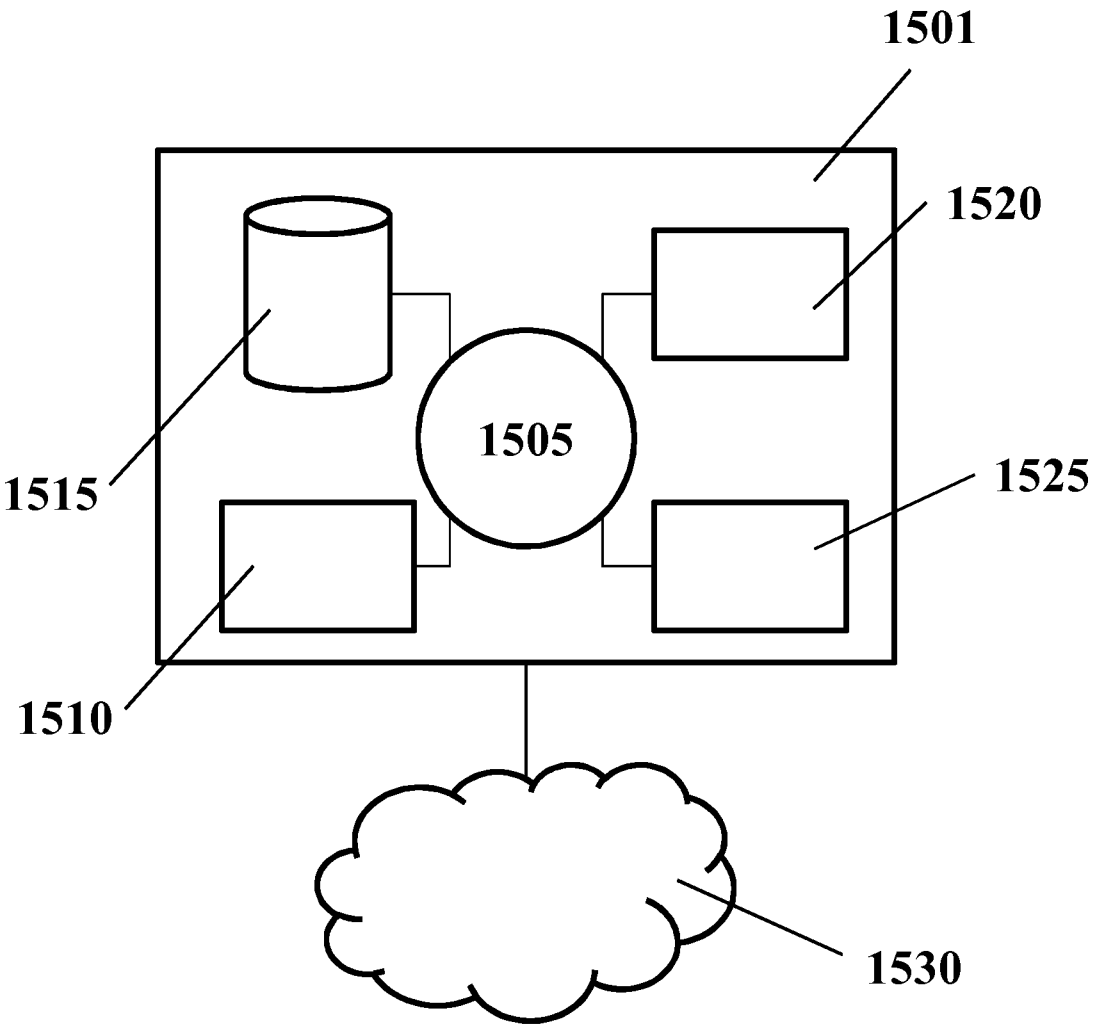


Fig. 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/058061

A. CLASSIFICATION OF SUBJECT MATTER

C12Q 1/68(2006.01)i, C12N 15/11(2006.01)i

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)
C12Q 1/68; C40B 20/00; G06F 19/00; C12N 15/11

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & keywords: copy number variation, rare mutation, computer readable medium, sequencing read, non transitory machine executable code

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2012-0100548 A1 (RAVA et al.) 26 April 2012 See abstract and pages 5, 8, 9, 10, 12-15, 28, 34-39.	1,6-8,10-43,45-54,57,59-63,68,69-74,76,116-124
Y		125-133
A		2,3,44,77-115,143-175,184-206
X	WO 2012-106559 A1 (TRANSLATIONAL GENOMICS RESEARCH INSTITUTE) 09 Augest 2012 See abstract and claims 1-3, 17-29.	3
Y	TSAI et al. `Discovery of rare mutations in populations: TILLING by sequencing` Plant Physiology, 29 April 2011, Vol.156, No.3, pp.1257-1268. See abstract and pages 1258, 1266, 1267.	125-133
A	US 2012-0046877 A1 (HYLAND et al.) 23 February 2012 See abstract and claims 1, 10, 11.	1-3,6-8,10-54,57,59-63,68-74,76-133,143-175,184-206

☒ 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 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	Date of mailing of the international search report
15 November 2013 (15.11.2013)	**18 November 2013 (18.11.2013)**

Name and mailing address of the ISA/KR
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon Metropolitan City, 302-701, Republic of Korea
Facsimile No. +82-42-472-7140

Authorized officer
KIM, Seung Beom
Telephone No. +82-42-481-3371

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/058061

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SCHMITT et al. `Detection of ultra-rare mutations by next-generation sequencing` PNAS, 01 August 2012 Vol.109, No.36, pp.14508-14513. See abstract, pages 1, 2 and figure 1.	1-3,6-8,10-54,57,59-63,68-74,76-133,143-175,184-206
A	US 2012-0214678 A1 (RAVA et al.) 23 August 2012 See abstract and claim 1	1-3,6-8,10-54,57,59-63,68-74,76-133,143-175,184-206

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/058061

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.: 55,56
because they relate to subject matter not required to be searched by this Authority, namely:
Claims 55 and 56 pertain to diagnostic methods for the human body, as well as treatment methods, and thus relate to a subject matter which this International Searching Authority is not required, under Article 17(2)(a)(i) of the PCT and Rule 39.1(iv) of the Regulations under the PCT, to search.

2. ☒ Claims Nos.: 4,5,9,58,64-67,141,181,183
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:
(See suplemental Box)

3. ☒ Claims Nos.: 75,134-140,142,176-180,182
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:

1. ☐ As all required addtional 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 an additional fee, this Authority did not invite payment of any additional fee.

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.:

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.

(Continuation of Box No.II)

2. Claims Nos.: 4,5,9,58,64-67,141,181,183

Claims 4 and 5 are worded in reference to the prevalence of claim 1, 2, or 3. However, said the prevalence has not been worded in claim 1, 2, or 3.

Claim 9 is worded in reference to the isolating of claim 1, 2, or 3. However, said the isolating has not been worded in claim 1, 2, or 3.

Claim 58 is worded in reference to the growth of a tumor of claim 1, 2, or 3. However, said the growth of a tumor has not been worded in claim 1, 2, or 3.

Claims 64-67 are worded in reference to the measurement data of claim 1, or 2. However, said the measurement data has not been worded in claim 1, or 2.

Claims 141, 181 and 183 are unclear since they are referring to the multiple dependent claims which do not comply with PCT Rule 6.4(a).

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2013/058061

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0100548 A1	26/04/2012	US 2012-149582 A1	14/06/2012
		US 2012-149583 A1	14/06/2012
WO 2012-106559 A1	09/08/2012	None	
US 2012-0046877 A1	23/02/2012	WO 2012-006291 A2	12/01/2012
		WO 2012-006291 A3	12/04/2012
US 2012-0214678 A1	23/08/2012	US 2012-0010085 A1	12/01/2012
		WO 2011-090556 A1	28/07/2011

Espacenet

Bibliographic data: WO2007019444 (A2) — 2007-02-15

SUBTRACTIVE SEPARATION AND AMPLIFICATION OF NON-RIBOSOMAL TRANSCRIBED RNA (NRRNA)

Inventor(s): FREIJE WADIHA [US]; BRIKUN IGOR [US] ± (FREIJE, WADIHA, ; BRIKUN, IGOR)

Applicant(s): EUCLID DIAGNOSTICS LLC [US]; FREIJE WADIHA [US]; BRIKUN IGOR [US] ± (EUCLID DIAGNOSTICS LLC, ; FREIJE, WADIHA, ; BRIKUN, IGOR)

Classification: **- international:** ***C12N15/09***
- cooperative: **C12N15/1096**; **C12Q1/6813**; **C12Q1/6853** more

Application number: WO2006US30721 20060804 Global Dossier

Priority number (s): US20050705964P 20050805

Also published as: WO2007019444 (A3) US2009137415 (A1)

Abstract of WO2007019444 (A2)

The invention provides a method of separating non-ribosomal transcribed RNA (nrRNA) fragments from ribosomal RNA (rRNA) and rRNA fragments. The method comprises (i) providing a sample comprising rRNA, rRNA fragments, and nrRNA fragments, and (ii) providing a plurality of probes. The probes hybridize to RNA targeting sequences of at least 50% of the contiguous regions of the rRNA and to rRNA fragments comprising the rRNA targeting sequences. The method further comprises (iii) adding the plurality of probes to the sample, (iv) hybridizing the probes to the rRNA and rRNA fragments to form rRNA-probe complexes and rRNA fragment-probe complexes, and (v) separating the rRNA-probe complexes and rRNA fragment-probe complexes.; The invention also provides a method of amplifying an nrRNA fragment, a method of analyzing nrRNA expression, a method of determining the level of nrRNA in a sample, and a kit and system useful in any of the foregoing methods.

IDT Exhibit 1020.4141



IDT Exhibit 1020.4142

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 February 2007 (15.02.2007)

PCT

(10) International Publication Number
WO 2007/019444 A2

(51) International Patent Classification:
C12N 15/09 (2006.01)

(21) International Application Number:
PCT/US2006/030721

(22) International Filing Date: 4 August 2006 (04.08.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/705,964 5 August 2005 (05.08.2005) US

(71) Applicant *(for all designated States except US)*: **EUCLID DIAGNOSTICS LLC** [US/US]; Chicago Technology Park, 2201 West Campbell Park Drive, Chicago, Illinois 60612 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **FREIJE, Wadiha** [LB/US]; 117 Elgin Avenue, Forest Park, il 60130 (US). **BRIKUN, Igor** [US/US]; 117 Elgin Avenue, Forest Park, Illinois 60130 (US).

(74) Agents: **KILYK, JR, John** et al.; Leydig, Voit & Mayer, Ltd, 180 North Stetson Avenue, Two Prudential Plaza, Suite 4900, Chicago, Illinois 60601-6780 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: SUBTRACTIVE SEPARATION AND AMPLIFICATION OF NON-RIBOSOMAL TRANSCRIBED RNA (NRRNA)

28S rRNA sequence:

cgcgacctcagatcagacgtggcgacccgctgaatttaagcatattagtcagcggaggaaaagaaactaaccagg
attccctcagtaacggcgagtgaacagggaagagcccagcgccgaatccccgccccgcgggcgcgggacatgt
ggcgtacggaagacccgctccccggcgccgctcgtgggggcccaagtccttctgatcgaggcccagcccgtgga
cggtgtgaggccggtagcggccggcgcgcgcccgggtcttcccggagtcgggttgcttgggaatgcagcccaaag
cgggtggtaaactccatctaaggctaaatacccggcacgagaccgatagtcaacaagtaccgtaagggaaagttga
aaagaacttgaagagagagttcaagagggcgtgaaaccgttaagaggtaaacgggtggggtccgcgcagtccg
cccggaggattcaacccggcggcgggtccggccgtgtcggcggcccggcggatctttcccgcccccctttcctcccg
accccctccacccgccctccctttccccgccgcccctcctcctcctcccggaggggggcgggctccggcgggtgcgg
gggtgggcgggcgggccgggggtgggtcggcgggggaccgtccccggaccggcgaccggccgccgccggg
gcgcatttccaggcggtgcgccgcgaccggctccgggacggctgggaaggcccggcggggaaggtggctcggg
gggccccgtccgtccgtccgtcctcctcctccccgtctccgcccccggcccgcgtcctccctcgggagggcgcgc
gggtcggggcggcggcggcggcggcggtggcggcggcggcgggggcggcgggaccgaaaccccccccgagt
gttacagccccccggcagcagcactcgccgaatcccgggggccgagggagcgagacccgtcgccgcgctctccc
ccctcccggcgcccaccccgcgcgggaatccccgcgagggggtctccccccggcgcggcgccggcgtctcctcgtg
gggggccgggccaccctcccacggcgcgaccgctctcccaccctcctccccgcgcccccgccccggcgcacg
gggggggtgccgcgcgcgggtcgggggcggggcggactgtcccagtgcgcccccgggcgggtcgcgccgtcg
ggcccgggggaggttctctcgggggccacgcgcgctcccccgaagaggggagcggcggagcgagccgcacggg
gtcggcggcgacgtcggctacccacccgacccgtcttgaaacacggaccaaggagtctaacacgtgcgcgagtcg
ggggctcgcacgaaagccgccgtggcgcaatgaaggtgaaggccggcgcgctcgccggccgaggtgggatccc
gaggcctctccagtccgccgaggggcaccaccggcccgtctcgcccgccgcgccggggaggtggagcacgagc
gcacgtgttaggacccgaaagatggtgaactatgcctgggcagggcgaagccagaggaaactctggtggaggtcc
gtagcggtcctgacgtgcaaatcggtcgtccgacctgggtataggggcgaaagactaatcgaaccatctagtagctg
gttccctccgaagtttccctcaggatagctggcgctctcgcagacccgacgcacccccgccacgcagttttatccggta
aagcgaatgattagaggtcttggggccgaaacgatctcaacctattctcaaactttaaatgggtaagaagcccggctc
gctggcgtggagccggggtggaatgcgagtgcctagtgggccacttttggtaagcagaactggcgctgcgggatga
accgaacgccgggttaaggcgcccgatgccgacgctcatcagaccccagaaaaggtgttggttgatatagacagc
aggacggtggccatggaagtcggaatccgctaaggagtgtgtaacaactcacctgccgaatcaactagccctgaaa
atggatggcgctggagcgtcgggcccatacccggccgtcgccggcagtcgagagtggacgggagcggcggggg
cggcggcgcgcgcgcgcgtgtggtgtgcgtcggagggcggcggcggcggcggcggcggggtgtgggtccttc
ccccgccccccccccacgcctcctccctcctcccgcccacgcccgctcccgcccccggagccccgcggagc
tacgccgcgacgagtaggagggcgctgcggtgagccttgaagcctagggcgcgggcccgggtggaggccgcc
gcaggtgcagatcttggtggtagtagcaaatattcaaacgagaactttgaaggccgaagtggagaagggttccatgt
gaacagcagttgaacatgggtcagtcggtcctgagagatgggcgagcgccgttccgaagggacgggcgatggcct
ccgttgccctcggccgatcgaaagggagtcgggttcagatcccgaatccggagtggcggagatgggcgccgcga
ggcgtccagtgcggtaacgcgaccgatcccggagaagccggcgggagccccggggagagttctcttttctttgtgaa
gggcagggcgccctggaatgggttcgccccgagagaggggcccgtgccttggaaagcgtcgcggttccggcggc

(57) Abstract: The invention provides a method of separating non-ribosomal transcribed RNA (nrRNA) fragments from ribosomal RNA (rRNA) and rRNA fragments. The method comprises (i) providing a sample comprising rRNA, rRNA fragments, and nrRNA fragments, and (ii) providing a plurality of probes. The probes hybridize to RNA targeting sequences of at least 50% of the contiguous regions of the rRNA and to rRNA fragments comprising the rRNA targeting sequences. The method further comprises (iii) adding the plurality of probes to the sample, (iv) hybridizing the probes to the rRNA and rRNA fragments to form rRNA-probe complexes and rRNA fragment-probe complexes, and (v) separating the rRNA-probe complexes and rRNA fragment-probe complexes. The invention also provides a method of amplifying an nrRNA fragment, a method of analyzing nrRNA expression, a method of determining the level of nrRNA in a sample, and a kit and system useful in any of the foregoing methods.

WO 2007/019444 A2

IDT Exhibit 1020.4143

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SUBTRACTIVE SEPARATION AND AMPLIFICATION OF NON-RIBOSOMAL TRANSCRIBED RNA (nrRNA)

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This patent application claims the benefit of U.S. Provisional Patent Application No. 60/705,964, filed August 5, 2005.

TECHNICAL FIELD OF THE INVENTION

[0002] The invention relates to methods comprising separating messenger RNA (mRNA) and/or non-coding RNA (ncRNA), referred to collectively herein as non-ribosomal transcribed RNA (nrRNA), from a sample, such as a sample of total RNA. The invention also relates to kits and systems for use in such methods.

BACKGROUND OF THE INVENTION

[0003] Most mRNA isolation kits rely on the oligo dT-mediated purification of mRNA. This method is disadvantageous when mRNA is degraded and, hence, fragmented. The oligo dT only binds to the polyA tail of mRNA and, therefore, only purifies intact mRNA and molecules of mRNA that contain polyA tails. Molecules of mRNA that do not contain polyA tails are not purified. For this reason, when performing expression analysis using RNA isolated from fixed tissues, scientists have relied on the reverse transcription of total RNA using random hexamers even though over 95% of the RNA present in the sample is ribosomal (Godfrey et al., J. Molec. Diagnostics 2(2): 84-91 (2000)). This approach is disadvantageous because it is not sensitive and does not enable optimal detection of mRNA present in low copy numbers.

[0004] To overcome this limitation, scientists have used oligonucleotide primers designed to anneal to a plurality of genes for the reverse transcription reaction (see, for example, U.S. Patent Application Publication Nos. 2004/0259105 and 2005/0095634). Using this approach, only the expression level of a limited number of genes can be determined. Furthermore, only small portions of those genes, which are upstream of the primer site, can be analyzed, since the majority of the RNA fragments are less than 500 base pairs in length. In addition, if the fragments are very short, such as less than 50 base pairs, it may be extremely difficult to design an assay to determine expression levels, whether by polymerase chain reaction, ligation-mediated amplification, or microarray analysis, for example.

[0005] Murphy and Whitley (U.S. Patent Application Publication No. 2003/0175709 and International Patent Application Publication No. WO 03/054162) have proposed the use of a bridging nucleic acid, which includes at least one targeting region and at least one bridging region, and a capture nucleic acid, which includes a capture region and a non-reacting structure. The targeting region is complementary to a targeted region of a targeted nucleic acid to be removed. The bridging region is complementary to the capture region. The nonreacting structure is a compound that does not react with a nucleic acid. The bridging nucleic acid can include up to 10 or more targeting regions that are complementary to different, non-overlapping targeted regions from the same or different targeted nucleic acids. This method is disadvantageous in that the use of multiple targeting regions on a single bridging nucleic acid can decrease the efficiency of hybridization between the targeting regions and the complementary targeted regions. In addition, the use of multiple targeting regions on a single bridging nucleic acid can interfere with hybridization between the bridging region of the bridging nucleic acid and the capture region of the capture nucleic acid. These effects also can decrease efficiency of recovery of non-ribosomal RNA. The methods disclosed by Murphy and Whitley also result in increased background hybridization due to the length of the oligonucleotides used.

[0006] The invention seeks to overcome the disadvantages of the currently available isolation kits by providing a method that enables the purification of nrRNA fragments that do not contain polyA tails as well as intact mRNA and/or ncRNA. This and other objects and advantages of the invention, as well as additional inventive features, will become apparent from the detailed description provided herein.

BRIEF SUMMARY OF THE INVENTION

[0007] The invention provides a method of separating non-ribosomal transcribed RNA (nrRNA) fragments from ribosomal RNA (rRNA) and rRNA fragments. The method comprises (i) providing a sample comprising rRNA, rRNA fragments, and nrRNA fragments, wherein the rRNA comprises multiple contiguous regions of about 100 base pairs. Each contiguous region comprises an rRNA targeting sequence, and the rRNA fragments comprise one or more of the rRNA targeting sequences. The method further comprises (ii) providing a plurality of probes. The probes hybridize to (a) different RNA targeting sequences of at least 50% of the contiguous regions of the rRNA and (b) rRNA

fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA. The method further comprises (iii) adding the plurality of probes to the sample, (iv) hybridizing the probes to the rRNA and rRNA fragments to form rRNA-probe complexes and rRNA fragment-probe complexes, and (v) separating the rRNA-probe complexes and rRNA fragment-probe complexes from the sample, thereby separating nrRNA fragments from rRNA and rRNA fragments.

[0008] The invention further provides a method of amplifying a cDNA complementary to an nrRNA fragment. The method comprises (i) separating an nrRNA fragment from rRNA and rRNA fragments according to the inventive method, wherein the nrRNA fragment comprises a 5' end. The method also comprises (ii) ligating a first oligonucleotide to the 5' end of the nrRNA fragment, (iii) hybridizing to the nrRNA fragment a second oligonucleotide, (iv) extending a DNA strand from the second oligonucleotide to generate a first cDNA, and (v) amplifying the first cDNA, thereby amplifying a cDNA complementary to an nrRNA fragment.

[0009] In addition, the invention provides a method of amplifying nrRNA fragments. The method comprises (i) separating nrRNA fragments from rRNA and rRNA fragments according to the inventive method. The method further comprises (ii) amplifying the nrRNA fragments. Alternatively, the method further comprises (ii) ligating a first oligonucleotide to the 5' end of the nrRNA fragment; (iii) hybridizing to the nrRNA fragment a second oligonucleotide; and (iv) extending a DNA strand from the second oligonucleotide to generate a first cDNA having a 5' end and a 3' end. The method then comprises (v) degrading the nrRNA fragment and, optionally, removing any free oligonucleotides and/or free RNA fragments; (vi) synthesizing a complementary strand to the first cDNA; (vii) optionally purifying the double-stranded DNA copy; and (viii) transcribing RNA from the double-stranded DNA copy. Synthesizing a complementary strand to the first cDNA in step (vi) comprises (a) contacting the first cDNA with a third oligonucleotide such that the third oligonucleotide hybridizes to the 3' end of the first cDNA and (b) extending the complementary strand from the third oligonucleotide, whereupon a double-stranded DNA copy of the nrRNA fragment is generated. The third oligonucleotide hybridizes to at least a portion of the first oligonucleotide and optionally contains an RNA polymerase promoter sequence.

[0010] Still further provided by the invention is a method of analyzing nrRNA expression. The method comprises (i) separating nrRNA fragments from rRNA and rRNA fragments according to the inventive method and (ii) analyzing nrRNA expression. Additionally, the invention provides a method of determining the level of nrRNA in a sample. The method comprises (i) separating nrRNA fragments from rRNA and rRNA fragments according to the inventive method and (ii) labeling the separated nrRNA fragments with a detectable label. The method of determining the level of nrRNA in a sample further comprises (iii) providing an array comprising a collection of fixed DNAs that hybridize to a plurality of target sequences located within an expression sequence of a gene of interest. The fixed DNAs comprise from about 15 nucleotides to about 750 nucleotides, and at least one target sequence is interspersed approximately every 500 base pairs of the expression sequence. The method additionally comprises (iv) applying the labeled nrRNA fragments to the array under hybridization conditions; (v) optionally removing unbound and nonspecifically bound nrRNA fragments from the array; (vi) detecting the labeled nrRNA fragments that have specifically annealed to the fixed DNAs in the array; and (vii) quantitating the level of labeled nrRNA fragments detected in step (vi) to determine the level of nrRNA in the sample.

[0011] The invention also provides a kit for separating nrRNA fragments from a sample comprising rRNA and rRNA fragments. The kit comprises a plurality of probes that hybridize to (a) different rRNA target sequences of at least 50% of contiguous regions of about 100 base pairs of the rRNA, which contiguous regions comprise the rRNA targeting sequences and (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA. Preferably, the kit also comprises instructions.

[0012] Likewise, the invention provides a system for separating nrRNA fragments from a sample comprising rRNA and rRNA fragments. The system comprises a device adapted for the separation of nucleic acids by hybridization. The system also comprises a plurality of probes that hybridize to (a) different rRNA target sequences of at least 50% of contiguous regions of about 100 base pairs of the rRNA, which contiguous regions comprise the rRNA targeting sequences, and (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA. The

plurality of probes is positioned within the device to separate the rRNA and rRNA fragments from nrRNA.

BRIEF DESCRIPTION OF THE FIGURES

[0013] Figures 1A - 1D show the nucleotide sequences of the 28S (SEQ ID NO: 1), 18S (SEQ ID NO: 2), 5.8S (SEQ ID NO: 3), 5S (SEQ ID NO: 4), 16S (SEQ ID NO: 5), and 12S (SEQ ID NO: 6) rRNAs. Sequences are presented from left to right, top to bottom, in accordance with convention.

[0014] Figures 2A - 2D show a set of oligonucleotides (SEQ ID NOs: 7-110) that can hybridize to rRNA and are suitable for use in the method. The "Start" positions are given relative to the sequences of Figures 1A - 1D.

[0015] Figures 3A - 3C show a set of forward and reverse primers (SEQ ID NOs: 111-165 and 191-245) that can amplify fragments corresponding to six rRNA species. The "Left" and "Right" designations indicate forward and reverse primers, respectively.

[0016] Figures 4A - 4M compare 28S rRNA sequences and provide the consensus sequence (SEQ ID NOs: 166-184).

[0017] Figure 5 is a diagram illustrating the preparation of double-stranded cDNA for amplification.

DETAILED DESCRIPTION OF THE INVENTION

[0018] The invention is predicated on the discovery that the use of an exhaustive set of probes enables the efficient removal of ribosomal RNA (rRNA) (i.e., intact rRNA and/or rRNA fragments) from a sample of total RNA. In doing so, the invention enables the separation and/or purification and/or enrichment of non-ribosomal transcribed RNA (nrRNA) fragments that do not contain polyA tails, such as are commonly found in degraded RNA. The invention does not suffer from the disadvantages attendant the use of currently available methods.

[0019] The invention provides a method of separating nrRNA, such as messenger RNA (mRNA) and non-coding RNA (ncRNA), from a sample, such as a sample comprising total RNA. The inventive method can be performed on any sample suitable for separation of RNA species. Indeed, any suitable sample of, for example, total RNA can be used. Preferably, degradation of RNA in the sample has been minimized. However, the invention is advantageous in that it can be used on degraded RNA, such as RNA recovered from

IDT Exhibit 1020.4149

preserved tissue samples, e.g., paraffin-embedded tissue samples, biopsies, surgical specimens, and any other source containing degraded RNA. Such RNA is usually less than 500 base pairs in size and may not represent the entire population of mRNA species present in original tissue samples. The invention also permits the recovery of ncRNA species that are present in a total RNA sample, such as small interfering RNA (siRNA) and small nuclear RNA (snRNA). Researchers have great difficulty in efficiently isolating such small RNA species using currently known methods.

[0020] The method comprises (i) removing rRNA from the sample of total RNA by hybridizing the rRNA with a plurality of probes comprising at least one probe that can hybridize to a complementary sequence within about 100 contiguous base pairs of the rRNA to be removed such that at least one probe hybridizes to a complementary sequence within at least about every 100 contiguous base pairs of the rRNA to be removed, and (ii) purifying the nrRNA.

[0021] Alternatively, the invention provides a method of separating nrRNA fragments from rRNA and rRNA fragments. The method comprises: (i) providing a sample comprising rRNA, rRNA fragments, and nrRNA fragments; (ii) providing a plurality of probes; (iii) adding the plurality of probes to the sample; (iv) hybridizing the probes to the rRNA and rRNA fragments to form rRNA-probe complexes and rRNA fragment-probe complexes, and (v) separating the rRNA-probe complexes and rRNA fragment-probe complexes from the sample (i.e., separating the hybridized RNA from the sample of RNA). Optionally, simultaneously with or subsequently to removing rRNA from the sample (e.g., the sample of total RNA), the method can further comprise removing any other RNA that is not of interest, such as unwanted mRNA species that are abundantly expressed or tRNA(s). It will be understood that "separating" the rRNA-probe complexes and/or rRNA fragment-probe complexes (or any other RNA) from the sample does not require complete elimination of rRNA or rRNA fragments from the sample. If desired, additional rounds of hybridization in accordance with the inventive method can be performed to further remove unwanted RNA. For example, the method can be performed on a sample of total RNA to achieve a sample comprising no more than 20% rRNA. Preferably, the method yields a sample or composition comprising no more than 10% rRNA, and more preferably the resulting composition comprises no more than 5% rRNA. Most preferably, the composition resulting from the inventive method comprises no more than 1% rRNA. The inventive

method preferably further comprises (vi) recovering and purifying the nrRNA fragments. In view of the above, the method is particularly suited for enriching a composition for nrRNA and fragments thereof by removing unwanted RNAs, such as rRNA and rRNA fragments.

[0022] The invention comprises employing a plurality of probes to saturate unwanted rRNA in a sample, thereby providing multiple "handles" for removing the unwanted material using any suitable separation technique. In this regard, rRNA and rRNA fragments can be conceptually divided into multiple contiguous regions of about 100 base pairs. For example, an rRNA or rRNA fragment that comprises 1000 base pairs comprises ten contiguous regions of about 100 base pairs. Each contiguous region comprises one or more RNA targeting sequences. Preferably, two or more of the multiple contiguous regions comprise different rRNA targeting sequences. An "rRNA targeting sequence" refers to a nucleotide sequence within the rRNA or rRNA fragment that is complementary to, and thereby hybridizes with, one or more probes of the plurality. The rRNA fragments of the sample desirably comprise one or more of the rRNA targeting sequences.

[0023] A "plurality of probes" is a collection of molecules, such as oligonucleotides, that bind rRNA target sequences within the rRNA or rRNA fragments of the sample. Preferably, the probes hybridize to (a) different RNA targeting sequences of at least 50% of the contiguous regions of the rRNA and (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA. In other words, if an rRNA comprises ten contiguous regions of about 100 base pairs, the probes hybridize to RNA targeting sequences in at least five of the ten contiguous regions. By "different" RNA targeting sequences is meant separate, individual RNA targeting sequences, which may or may not be unique. Desirably, multiple probes hybridize over the length of the unwanted RNA to ensure efficient removal. For example, the probes of the inventive method can hybridize to rRNA targeting sequences of at least 60% of the contiguous regions (e.g., the probes hybridize to rRNA targeting sequences of at least 70% of the contiguous regions). Even more preferably, the probes hybridize to rRNA targeting sequences of at least 80% of the contiguous regions (e.g., the probes hybridize to rRNA targeting sequences of at least 90% of the contiguous regions). Still more preferably, the probes hybridize to rRNA targeting sequences of at least 95% of the contiguous regions. Even more preferably, the probes hybridize to rRNA targeting sequences of all of the

contiguous regions (i.e., rRNA targeting sequences located in each and every contiguous region).

[0024] In this regard, rRNA and fragments thereof and, optionally, any other RNA that is not of interest, can be removed from a sample, e.g., a sample of total RNA, by hybridizing the RNA to a plurality of probes comprising, for example, at least one tagged oligonucleotide (i.e., DNA or RNA, including modified base analogs), tagged single-stranded nucleic acid molecule (i.e., DNA or RNA), or tagged DNA mimic, such as a locked nucleic acid (LNA; Koshkin et al., Tetrahedron 54: 3607-3630 (1998); Koshkin et al., JACS 120: 13252-13253 (1998); and Wahlestedt et al., PNAS 97: 5633-5638 (2000)) and/or a DNA analog with a non-phosphodiester backbone (such as a peptide nucleic acid (PNA; Nielson, Molec. Biotech. 26: 233-248 (2004); and Egholm et al., Nature 365(6446): 566-568 (1993)) or morpholino derivatives (Summerton, BioChim. Biophys. Acta 1489:141-158 (1999)), which is complementary to the RNA that is to be removed.

[0025] The rRNA and the rRNA fragments in the sample can comprise eukaryotic rRNA and eukaryotic rRNA fragments, although a sample also can comprise bacterial RNA species. In addition, the probes, including any tagged oligonucleotides, tagged single-stranded nucleic acid molecules, and/or DNA mimics, can be complementary to 18S, 28S, 5S, 5.8S, 12S, or 16S rRNA (see Figures 1A - 1D and 2A - 2D). The sequences of different rRNA species are highly conserved; however, the probes (e.g., oligonucleotides or molecules complementary to the rRNA species) can contain sequences complementary to polymorphic regions within the rRNA sequences to maximize removal of rRNA species from the sample (see, e.g., Figures 4A - 4M for examples of such polymorphic regions). Such probes (e.g., oligonucleotides and/or molecules complementary to the rRNA species) can be used in various combinations. The rRNA and rRNA fragments can comprise at least two rRNAs selected from the group consisting of 28S, 18S, 5.8S, 5S, 12S, and 16S rRNAs. Thus, the plurality of probes can comprise one that is complementary to 18S rRNA and another that is complementary to 28S rRNA. Under certain circumstances, such as when an nrRNA of interest is present in low copy number, it can be desirable to remove unwanted RNA fragments that are abundant in the sample. It may be desirable to use an additional oligonucleotide and/or single-stranded nucleic acid molecule probe in addition to probes that are complementary to 18S and 28S, such as one or more or even all of 12S, 16S, 5.8S, and 5S rRNAs, alone or in further combination with probes (such as an oligonucleotide, a

IDT Exhibit 1020.4152

single-stranded nucleic molecule and/or a DNA mimic) that are complementary to tRNA and/or that are complementary to any mRNA that may be abundantly expressed in the RNA sample under investigation and that is not of interest.

[0026] Probes (e.g., oligonucleotide probes) can be synthesized using any *in vitro* chemical synthesis known in the art. Oligonucleotide probes can be as short as around 12 nucleotides or as long as around 150 nucleotides, such as around 15 or 17 nucleotides up to around 100 nucleotides, whereas a single-stranded nucleic acid molecule probe can be as short as 75 nucleotides up to the full-length of the RNA molecule that it complements. Oligonucleotides can be selected using a Primer selection program (e.g., Vector NTI (Invitrogen, Carlsbad CA) or Primer 3 (Rozen et al., "Primer 3 on the WWW for general users and for biologist programmers." In: Krawetz and Misener, eds. Bioinformatics Methods and Protocols, Methods in Molecular Biology. Humana Press, Totowa, NJ (2000), pp. 365-386)) using suitable parameters for primer length, base composition, and melting temperatures. It is preferred that the oligonucleotide probe have a GC content of about 35-80%, and a melting temperature of about 50-75°C. The optimal oligonucleotide length and melting temperature for the probes can be determined by the composition of the buffers used for the hybridization. Primer design programs select oligonucleotides based on the user's input of relevant parameters, including minimizing secondary structures, palindromic sequences, and hairpin loops. The oligonucleotides can be further analyzed by sequence homology search programs (e.g., BLAST) to minimize cross-hybridization to other genomic sequences, especially homology to expressed transcripts. It is preferred that the set of oligonucleotides used as probes for the hybridization contain at least one oligonucleotide complementary to each fragment (i.e., contiguous region) of the rRNA, wherein each fragment (i.e., contiguous region) is from about 100 base pairs to about 250 base pairs.

[0027] Single-stranded nucleic acid molecules can be synthesized *in vitro* (i.e., PCR or transcription-coupled amplification) or *in vivo* (i.e., replicated in living cells using recombinant DNA vectors). Such molecules can be synthesized as fragments or full-length molecules, and can be employed as probes in the context of the invention.

[0028] The probes of the invention can be tagged. By "tagged" is meant adding a molecule (a "tag") to the probe (e.g., oligonucleotide, single-stranded nucleic acid molecule, and/or DNA mimic) that will permit downstream manipulations, such as removal. For example, the tag can be biotin, which, upon binding to streptavidin, such as streptavidin

attached to a support (e.g., a magnetic bead or plastic surface), results in the removal of rRNA and any other RNA that is not of interest from the sample. Alternatively, the tag can be digoxygenin, which, upon binding to an immobilized anti-digoxygenin antibody, such as an antibody attached to a substrate (e.g., a magnetic bead or a microtiter plate), results in the removal of rRNA and any other RNA that is not of interest from the sample. Alternatively, the tag can be an amine linker that can be used to immobilize the oligonucleotide on a solid support.

[0029] The method of the invention comprises hybridizing the probes to the rRNA and rRNA fragments to form rRNA-probe complexes and rRNA fragment-probe complexes, which are then separated from the sample. Hybridization refers to the annealing of single stranded oligonucleotides to form double stranded oligonucleotides in an environment below the melting temperature of the double stranded oligonucleotide. Hybridization conditions are well known in the art and can be adapted to accommodate specific probes, rRNA sequences, and desired level of specificity of unwanted RNA removal. In this regard, "stringency" of hybridization reactions is determined by probe length, homology between the sequences, washing temperature, and salt concentration. In general, longer probes require higher temperatures for proper annealing, while shorter probes need lower temperatures. In addition, the higher the degree of homology between the probe and rRNA or rRNA fragment, the higher the relative temperature that can be used. Hybridization reactions are further described in, for example, Ausubel et al., Current Protocols in Molecular Biology, Wiley Interscience Publishers, (1995), and U.S. Patent No. 7,081,340.

[0030] The hybridization conditions can include, but are not limited to, "stringent conditions" or "high stringency conditions" which typically: (1) employ low ionic strength and high temperature for washing, for example 0.015 M sodium chloride/0.0015 M sodium citrate/0.1% sodium dodecyl sulfate at 50°C; (2) employ, during hybridization, a denaturing agent, such as formamide (e.g., 50% (v/v) formamide), 50 mM sodium phosphate buffer at pH 6.5 with 750 mM sodium chloride, 75 mM sodium citrate at 42°C; or (3) employ 50% formamide, 5X SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium phosphate (pH 6.8), 0.1% sodium pyrophosphate, optionally followed by a high-stringency wash consisting of 0.1X SSC containing EDTA at 55°C.

[0031] "Moderately stringent conditions" are described in, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, New York: Cold Spring Harbor Press, 1989.

Moderately stringent conditions can be achieved using a washing solution and hybridization conditions (e.g., temperature and ionic strength) less stringent that those described above. For example, moderately stringent conditions can comprise an overnight incubation at 37°C in a solution comprising: 20% formamide, 5X SSC (150 mM NaCl, 15 mM trisodium citrate), 50 mM sodium phosphate (pH 7.6). The skilled artisan will recognize how to adjust the temperature, ionic strength, etc. as necessary to accommodate factors such as probe length and the like. DNA recovered in the remaining wash buffer can be added to the original hybridization mix, and hybridization can be repeated to further reduce the amount of rRNA in the sample. If desired, the rRNA of the sample is discarded, and the nrRNA is recovered in the hybridization buffer. In this regard, the hybridization step of the inventive method can be repeated to reduce the concentration of unwanted RNA, such as unwanted rRNA or rRNA fragments, in the sample. The hybridization step can be repeated by administering fresh probes to the reaction to form additional rRNA-probe complexes and rRNA fragment-probe complexes with the residual rRNA and rRNA fragments that may remain in the sample. Further reducing the concentration of rRNA and rRNA fragments further enriches the composition with desired nrRNA.

[0032] In view of the above, the invention further provides a composition comprising mRNA, which has been separated from a sample of total RNA and which includes N-terminal and/or internal molecules of mRNA. "N-terminal molecule of mRNA" means a molecule of mRNA that includes the transcription start site. "Internal molecule of mRNA" means a molecule of mRNA that does not include the transcription start site or the polyA tail. The mRNA can be amplified.

[0033] Also provided by the invention is an improved method of amplifying nrRNA fragments (such as mRNA or fragments thereof). The method comprises (i) separating nrRNA fragments from rRNA and rRNA fragments according to the inventive method described herein, and (ii) amplifying the nrRNA fragments. The improvement provided by the invention comprises using either of (i) nrRNA that has been separated from a sample of total RNA in accordance with the above-described method or (ii) the above-described composition. Although the inventive method enriches a sample for desired nrRNA, the composition can be further enriched with desired nrRNA by amplifying the nrRNA fragments in the resulting composition. The amplification of RNA, and oligonucleotides in general, has been the subject of numerous patents (see, e.g., U.S. Patent Nos. 5,545,522;

5,716,785; 6,291,170; 6,642,034; 6,593,086; and 6,794,141, and U.S. Patent Application Publication No. 2005/0009101). Examples of methods that can be used to amplify mRNA include, but are not limited to, concatenation of RNA fragments, reverse transcription to generate the first cDNA strand, and strand displacement amplification; and attachment of an RNA polymerase promoter to the first or second strand cDNA followed by a transcription reaction to generate a multitude of copies of the starting cDNAs. A promoter is a sequence that can be recognized and bound by an RNA polymerase with subsequent transcription of the DNA fragment that is located downstream of the 3' end of the promoter. Examples of such promoters include, but are not limited to, the T3, T4, T7, SP6, and Q beta replicase promoters.

[0034] The invention further provides an improved method of performing expression analysis of nrRNA. The method of analyzing nrRNA expression comprises (i) separating nrRNA fragments from rRNA and rRNA fragments according to the inventive method described herein and (ii) analyzing nrRNA expression. Examples of methods that can be used to perform expression analysis include, but are not limited to, PCR, quantitative PCR, and microarrays (PCR protocols, Bartlett et al., eds. Humana Press, Totowa, NJ (2003); RT-PCR protocols, O'Connell, ed. Humana Press (2002); Gene cloning and analysis by RT-PCR, Siebert et al., eds. Eaton Publishing, Natick, MA (1998); A-Z quantitative PCR, Bustin, ed. IUL press, LaJolla, CA (2004); DNA microarrays: Gene expression applications, Jordan, ed. Springer-Verlag, New York, NY (2001); Microarrays Methods and Applications: Nuts and Bolts, Hardiman, ed. DNA Press, Eagleville, PA (2003); and Guide to analysis of microarray data, 2nd ed., Knudsen, ed. Wiley & Sons, Indianapolis, IN (2004)). The improvement provided by the invention comprises using either of (i) mRNA that has been separated from a sample of total RNA in accordance with the above-described method or (ii) the above-described composition.

[0035] A kit is also provided by the invention. The kit comprises instructions and at least one reagent for separating nrRNA from a sample of total RNA by (i) removing rRNA from the sample of total RNA, and, optionally, simultaneously or subsequently removing any other RNA that is not of interest, and (ii) purifying the nrRNA. The at least one reagent can comprise at least one oligonucleotide, single-stranded nucleic acid molecule, or DNA mimic that is complementary to (i) an rRNA selected from the group consisting of 18S, 28S, 12S, 16S, 5.8S and 5S and/or (ii) a tRNA.

[0036] In addition, the invention provides a kit for separating nrRNA fragments from a sample comprising rRNA and rRNA fragments (e.g., eukaryotic rRNA and eukaryotic rRNA fragments). Preferably, the rRNA and rRNA fragments comprise 28S, 18S, 5.8S, 5S, 16S, and 12S rRNAs or fragments thereof. The inventive kit comprises a plurality of probes that hybridize to different rRNA target sequences of at least 50% of contiguous regions of about 100 base pairs of the rRNA, which contiguous regions comprise the rRNA targeting sequences. The plurality of probes also hybridize to rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA. The kit also, optionally, further comprises instructions.

[0037] To increase specificity and efficiency of removal of unwanted rRNA from the sample, the probes of the kit preferably hybridize to rRNA targeting sequences of at least 60% of the contiguous regions (more preferably at least 70% of the contiguous regions). More preferably, the probes hybridize to rRNA targeting sequences of at least 80% of the contiguous regions (e.g., at least 90% of the contiguous regions). Even more preferably, the probes hybridize to rRNA targeting sequences of at least 95% of the contiguous regions (e.g., the probes hybridize to rRNA targeting sequences of all of the contiguous regions).

[0038] Also provided is a method of amplifying nrRNA fragments, such as those present in degraded RNA. The method comprises (i) separating an nrRNA fragment from rRNA and rRNA fragments according to the inventive method. The method further comprises (ii) ligating a first oligonucleotide to the 5' end of nrRNA fragments, which have been separated from a sample (e.g., a sample of total RNA) in accordance with the above method. The first oligonucleotide (a) can comprise a first unique sequence, (b) preferably is of sufficient length to hybridize to a complementary oligonucleotide under annealing (i.e., hybridization) conditions as are known in the art, and (c) optionally contains an RNA polymerase promoter sequence. The first oligonucleotides serves as an anchor sequence at the 5' end of nrRNA fragments, thereby facilitating the subsequent analysis of the fragments, in particular short fragments, such as those of less than about 50 base pairs. Desirably, the first oligonucleotide does not contain a sequence that is present in the genome of the species from which the nrRNA sample is derived. Preferably, the first oligonucleotide is at least about 10 nucleotides in length, more preferably, at least about 25 nucleotides in length, and most preferably, at least about 35 nucleotides in length. Examples of RNA polymerase promoter sequences include, but are not limited to, T3, T7,

SP6, T4, and Qβ replicase. If desired, non-ligated nucleotides can be removed from the reaction.

[0039] The method further comprises (iii) annealing (i.e., hybridizing) to the nrRNA fragment a second oligonucleotide. The second oligonucleotide preferably comprises a second unique sequence and a random sequence, wherein the random sequence is of sufficient length to hybridize to the nrRNA under annealing (i.e. hybridization) conditions as are understood in the art. Desirably, the unique sequence of the second oligonucleotide does not contain a sequence that is present in the genome of the species from which the nrRNA sample is derived. Optionally, the unique sequence of the second oligonucleotide comprises an RNA polymerase promoter. As such, it is possible to obtain a nucleotide chain comprising RNA polymerase promoters capable of directing transcription in opposite directions. The random sequence can be from about 6 to about 35 nucleotides in length, preferably from about 6 to about 12 nucleotides in length. The random sequence must anneal to the nrRNA to be amplified. One of ordinary skill in the art will appreciate that step (iii) involves heating the sample and then cooling the sample to allow the second oligonucleotide to anneal to the nrRNA (e.g., mRNA) and incubating the sample at the annealing temperature in accordance with methods known in the art.

[0040] The method still further comprises (iv) extending a DNA strand from the second oligonucleotide to generate a first cDNA having a 5' and a 3' end. The DNA strand can be extended using any suitable method such as, for example, reverse transcription. Several methods of reverse-transcription for obtaining cDNA are known in the art. If the sample of nrRNA is small, it may be desirable to amplify the cDNA. Therefore, if desired, the cDNA may be amplified using any suitable method, such as PCR. Then, the method comprises (v) degrading the nrRNA (i.e., the nrRNA fragment). The nrRNA can be degraded using one or more RNA-degrading enzymes, such as RNases. The enzymes will need to be removed prior to transcription, such as by phenol-chloroform extraction in accordance with methods known in the art. The method optionally further comprises removing any free oligonucleotides and/or RNA fragments in accordance with methods known in the art, such as size-exclusion column purification, preferential precipitation, and the like.

[0041] Afterwards, the method comprises (vi) synthesizing the complementary strand to the first cDNA by annealing a third oligonucleotide to the 3' of the first cDNA and extending a strand from the third oligonucleotide. In other words, a complementary strand

to the first cDNA is synthesized by (a) contacting the first cDNA with a third oligonucleotide such that the third oligonucleotide hybridizes to the 3' end of the first cDNA, and (b) extending the complementary strand from the third oligonucleotide, whereupon a double-stranded DNA copy of the nrRNA fragment is generated. The third oligonucleotide is of sufficient length to hybridize to at least a portion of the first oligonucleotide under annealing conditions and optionally contains an RNA polymerase promoter sequence as previously described. The third oligonucleotide can comprise a unique sequence that is not present in the genome of the species from which the nrRNA sample was extracted. If the free oligonucleotides and/or RNA fragments are not removed in step (v), then step (vi) should be conducted at an elevated annealing temperature to prevent the second oligonucleotide and any RNA fragments from annealing to the first cDNA, in which case a thermostable DNA polymerase, preferably with proofreading capabilities, is used for synthesis of the complementary strand to the first cDNA.

[0042] The method optionally comprises (vii) purifying the double-stranded DNA copy of the nrRNA (i.e., the nrRNA fragment) in accordance with methods known in the art (e.g., binding to glass milk or ion-exchange resin, proteinase K digestion and phenol-chloroform extraction, etc.). In addition, the double-stranded DNA copy can be amplified using, for instance, PCR. The method further comprises (viii) transcribing RNA from the double-stranded DNA in accordance with methods known in the art, such as with commercially available kits (e.g., such as those available from Ambion (Austin, TX) or Epicentre (Madison, WI)). The purified double-stranded copy of the nrRNA carries two RNA polymerase promoters that can be used to generate RNA copies of the coding or the noncoding strands.

[0043] In addition, the invention provides a method of amplifying a cDNA complementary to an nrRNA fragment. The method comprises (i) separating an nrRNA fragment from rRNA and rRNA fragments according to the invention. The method further comprises (ii) ligating a first oligonucleotide, such as the first oligonucleotide described above, to the 5' end of the nrRNA fragment. The method then comprises (iii) hybridizing to the nrRNA fragment a second oligonucleotide and (iv) extending a DNA strand from the second oligonucleotide to generate a first cDNA. The first cDNA is then (v) amplified, thereby amplifying a cDNA complementary to an nrRNA fragment.

[0044] In view of the above, the invention provides a kit comprising instructions and at least one reagent for generating a double-stranded DNA copy of mRNA in accordance with the above-described kit. The kit can further comprise the kit described above, in which case the at least one reagent comprises at least one tagged oligonucleotide, single-stranded nucleic acid molecule, or DNA mimic that is complementary to (i) an rRNA selected from the group consisting of 18S, 28S, 12S, 16S, 5.8S and 5S and/or (ii) a tRNA.

[0045] The invention further provides a method of determining the level of nrRNA in a sample of nrRNA, which has been separated from total RNA as described above. The method comprises (i) separating nrRNA fragments from rRNA and rRNA fragments according to the inventive method, and (ii) labeling the separated nrRNA with a detectable label. The nrRNA can be labeled by any suitable method as is known in the art. Examples of suitable labels include biotin, fluorescent tags, radioactive labels, and the like. The method further comprises (iii) providing an array comprising a collection of fixed DNAs that hybridize to a plurality of target sequences located within an expression sequence of a gene of interest. The fixed DNAs (e.g., DNA fragments) comprise from about 15 nucleotides to about 750 nucleotides, such as from about 30 nucleotides to about 750 nucleotides. At least one target sequence is interspersed approximately every 500 base pairs of the expression sequence. In other words, preferably, a fixed DNA hybridizes to a target sequence that is within 500 base pairs of another target sequence. More preferably, a target sequence recognized by a fixed DNA is located less than 500 base pairs from other target sequences recognized by fixed DNAs. Also preferably, the collection of fixed DNAs comprises a sufficient number of fixed DNAs to hybridize to target sequences of about 50% or more of the expression sequence. In other words, the collection preferably comprises DNA fragments that correspond to at least about 50% of each gene of interest and, preferably, the collection comprises at least one unique or discrete DNA fragment for every 500 contiguous base pairs of each gene of interest. The method further comprises (iv) applying the labeled nrRNA fragments to the array under hybridization conditions, i.e., under conditions that allow annealing of the labeled nrRNA with complementary fragments in the array of DNA fragments, and (v) optionally removing unbound and nonspecifically bound nrRNA fragments from the array. The method then comprises (vi) detecting the labeled nrRNA fragments that have specifically annealed to the fixed DNAs in the array (i.e., the nrRNA that has annealed to complementary fragments in the array of DNA

fragments); and (vii) quantitating the level of labeled nrRNA fragments detected in (iv) to thereby determine the level of nrRNA in the sample. The method can comprise (viii) correlating the level of nrRNA in the sample to a level of expression of a gene of interest. For example, the amount of nrRNA in the sample indicates whether the gene of interest is highly or poorly expressed.

[0046] An array for use in the inventive method can include any one-dimensional, two-dimensional or substantially two-dimensional (as well as a three-dimensional) arrangement of addressable regions bearing particular probes associated with that region. The arrays for use in the invention generally are arrays of probes as described above, and can include nucleic acids, including oligonucleotides, polynucleotides, cDNAs, RNAs, synthetic mimetics thereof, and the like. Oligonucleotide probes can be attached to the array substrate at any point along the nucleic acid chain, but are generally attached at one of their termini (e.g., the 3' or 5' terminus).

[0047] Any suitable array substrate can be used in the context of the invention including, for example, siliceous materials, e.g., glass, fused silica, ceramics and the like; metals such as stainless steel, aluminum, and alloys thereof; polymers, e.g., plastics and other polymeric materials such as polysulfone, poly(vinylidene fluoride), poly(ethyleneterephthalate), polyurethane, e.g., nonporous polyurethane, fluoropolymers such as polytetrafluoroethylene (e.g., Teflon^{TM}) or the like, polypropylene, polystyrene, polycarbonate, PVC, nylon, and blends thereof. With arrays that read using fluorescence, the substrate may be of a material that emits low fluorescence upon illumination with the excitation light. Additionally, the substrate may be relatively transparent to reduce the absorption of the incident illuminating laser light and subsequent heating if the focused laser beam travels too slowly over a region. Any given substrate may carry one, two, four or more arrays disposed on a front surface of the substrate. Depending upon the use, multiple arrays can be used which are the same or different from one another. Each array may cover an area of less than 100 cm^2, or even less than 50 cm^2, 10 cm^2 or 1 cm^2. The substrate carrying one or more arrays can be any shape, such as a rectangular solid (although other shapes are possible). Arrays are further described in, for example, U.S. Patent No. 7,022,157.

[0048] Arrays can be fabricated in a number of ways, including, for example, drop deposition, direct synthesis, light directed fabrication, or by printing. Methods of

generating arrays are further described in, for example, U.S. Patent Nos. 6,171,797; 6,180,351; 6,232,072; 6,242,266; and 6,323,043, as well as U.S. Patent Application No. 09/302,898, filed Apr. 30, 1999 by Caren et al. (a counterpart of which has been published as U.S. Patent Application No. 2004/0203138 A1), and the references cited therein, which are hereby incorporated by reference. For example, the labeled nrRNA can be incubated with an array as described herein in a commercially available buffer (e.g., such as those available from Sigma Chemical Co., St. Louis, MO), for example.

[0049] Unbound and nonspecifically bound nrRNA can be removed by washing in a salt-containing buffer at various temperatures/stringencies. The method employed in detecting the labeled nrRNA will be determined largely by the label employed in accordance with well-known methods. The level of detected labeled nrRNA can be quantitated after subtracting the background. If desired, the nrRNA can be initially amplified in accordance with the above-described methods.

[0050] Accordingly, in view of the above, the invention also provides a microarray for analysis of nrRNA fragments for genes of interest. The microarray comprises a collection of DNA fragments, wherein each fragment ranges in size from about 15 nucleotides to about 750 nucleotides (e.g., from about 30 base pairs to about 750 base pairs), such as from about 50 base pairs to about 600 base pairs or from about 60 base pairs to about 500 base pairs. The collection preferably comprises at least one unique DNA fragment (or discrete DNA fragment) for every 500 contiguous base pairs of each gene of interest. In other words, the collection preferably comprises a DNA fragment (i.e., probe) that hybridizes to a target DNA (e.g., gene) of interest within 500 base pairs of another DNA fragment of the collection. Preferably, the collection comprises DNA fragments that correspond to at least about 50% of each gene of interest, as described further herein.

[0051] Thus, further provided is a kit comprising instructions and at least one reagent for separating nrRNA from a sample of total RNA by (i) removing rRNA from the sample of total RNA, and, optionally, simultaneously or subsequently removing any other RNA that is not of interest, and (ii) purifying the nrRNA, and the above-described microarray. The at least one reagent comprises at least one tagged oligonucleotide, single-stranded nucleic acid molecule, or DNA mimic that is complementary to (i) an rRNA selected from the group consisting of 18S, 28S, 12S, 16S, 5.8S and 5S and/or (ii) a tRNA.

[0052] In addition, the invention provides a system for separating nrRNA fragments from a sample comprising rRNA and rRNA fragments, such as eukaryotic rRNA (e.g., comprise 28S, 18S, 5.8S, 5S, 16S, and 12S rRNAs or fragments thereof). The system comprises (i) a device adapted for the separation of nucleic acids by hybridization; and (ii) a plurality of probes positioned within or upon the device to separate the rRNA and rRNA fragments from nrRNA. The probes hybridize to (a) different rRNA target sequences of at least 50% of contiguous regions of about 100 base pairs of the rRNA, which contiguous regions comprise the rRNA targeting sequences. Preferably, the probes of the plurality hybridize to rRNA targeting sequences of at least 60% of the contiguous regions (more preferably at least 70% of the contiguous regions) to further increase the efficiency of separation. More preferably, the probes hybridize to rRNA targeting sequences of at least 80% of the contiguous regions (e.g., at least 90% of the contiguous regions). Even more preferably, the probes hybridize to rRNA targeting sequences of at least 95% of the contiguous regions (e.g., the probes hybridize to rRNA targeting sequences of all of the contiguous regions). The probes also hybridize to (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA.

[0053] Any device that employs hybridization to separate of nucleic acids can be used in the context of the inventive method. For example, the device can be an array, such as the array described above, where probes are attached to a substrate to facilitate separating target oligonucleotides from a sample. The device can be a filtration apparatus wherein probes are attached to filters to allow oligonucleotide separation following centrifugation. Likewise, the device can be a microfluidic device. Microfluidic devices can comprise biochannels or microchannels comprising arrays of probes to capture target RNA in samples. Alternatively, a microfluidic reaction apparatus can be configured to have one or more individual reaction chambers in direct communication with a microarray of oligonucleotide probes to perform multiple, parallel, thermally controlled hybridization reactions. Microfluidic-style devices have been developed for oligonucleotide amplification, detection, and/or hybridization assays, as described in International Patent Applications WO 96/15450; WO 96/15576; WO 97/27324; WO 96/39252; WO 96/39260; WO 97/16561; WO 97/16835; WO 97/37755; WO 97/43629; and WO 98/13683; U.S. Patent Nos. 5,061,336; 5,071531; 5,110,745; 5,126,022; 5,135,627; 5,147,607; 5,296,375; 5,304,487;

5,486,335; 5,498,392; 5,569,364; 5,585,069; 5,587,128; 5,593,838; 5,603,351; 5,631,337; 5,632,876; 5,637,469; 5,643,738; 5,681,484; 5,726,026; 5,747,169; 5,750,015; 5,755,942; 5,770,029; 5,843,767; and U.S. Patent Application Publication No. 2005/0009101 A1.

[0054] The following examples serve to illustrate the invention but are not intended to limit its scope in any way.

EXAMPLE 1

[0055] This example describes a method of recovering mRNA from paraffin-embedded tissues by subtraction with oligonucleotides complementary to rRNA.

[0056] Five 10 micron sections of paraffin-embedded, benign prostatic hyperplasia tissues are extracted 3X with xylene to remove the paraffin and 3X with ethanol. After air drying, the tissue is incubated in 100 µl of proteinase K buffer at 50°C for four days, with additional aliquots of proteinase K buffer added every 12 hours. The proteinase K buffer comprises 10 millimolar Tris (pH 8), 5 millimolar EDTA, 100 millimolar NaCl, 0.1% SDS, and 2 µg/mL proteinase K. The RNA is then isolated using total RNA isolation methods, such as lysis in guanidium thiocyanate followed by phenol-chloroform extraction and ethanol precipitation. A number of commercial kits are suitable for this purpose, such as Totally RNA (Ambion), Perfect RNA Eukaryotic Kit (Eppendorf, Westbury, NY), and RNeasy kit (Qiagen, Valencia, CA). The RNA is extracted according to manufacturer's directions. Following extraction and precipitation, the RNA is subjected to DNase treatment, phenol chloroform extraction, and precipitation. The RNA resuspended in H_2O, and stored at –80°C.

[0057] To remove the rRNA, one microgram of the total RNA is incubated with biotinylated oligonucleotides in a hybridization buffer to promote the annealing of the oligonucleotide to the rRNA fragments. A sample of suitable oligonucleotides is shown in Figures 2a - 2d. Any suitable buffer can be used for this step, including buffers containing tetramethylammonium chloride (TMACl; see, e.g., U.S. Patent No. 5,633,134 in regard to the use of quaternary ammonium salt). One such buffer is composed of 50 mM Tris HCl, pH 8.0, 150 mM NaCl, 10 mM EDTA, and 0.1% SDS. The oligonucleotide mix includes an equimolar amount of all oligonucleotides that are complementary to 28S, 18S, 5.8S, 5S, 12S, and 16S rRNA. One microgram of RNA and 2 micrograms of oligonucleotides are

resuspended in 10 µl of H_2O and incubated at 75°C for 10 minutes, followed by immersion in an ice bath for 10 minutes. Ten microliters of 10X buffer and 80 µl of H_2O are added, and the solution is incubated at 65°C for between 10 minutes and several hours. Following the incubation step, a 50 µl aliquot of streptavidin-coated magnetic beads (Dynal, Oslo, Norway) is added to the hybridization solution, which is then mixed gently to allow for binding of the biotinylated oligonucleotides to the strepavidin-coated beads. The beads are then precipitated using a magnet, and the supernatant is transferred to a fresh tube. If desired, the hybridization step can be repeated one or more times by adding a fresh aliquot of the oligonucleotides. The supernatant is extracted with phenol-chloroform, and the RNA is precipitated with the addition of 2 volumes of ethanol. The RNA is resuspended in H_2O, quantitated using the picogreen DNA quantitation kit (Invitrogen, Carlsbad, CA), and stored at –80°C. The RNA is then ready for reverse transcription.

[0058] This example illustrates a method of separating nrRNA from rRNA and rRNA fragments using the inventive method.

EXAMPLE 2

[0059] This example demonstrates a method of preparing probes for use in the invention.

[0060] To generate single-stranded DNA fragments complementary to rRNA, PCR primers are designed to amplify segments of the rRNA species that range in size from around about 100 base pairs up to the full-length rRNA template. The primers listed in Figures 3A - 3C are designed to amplify fragments corresponding to all six rRNA species. The amplification reaction is performed using MasterTaq kit from Eppendorf according to the manufacturer's direction. The buffer is supplemented with 10% DMSO or 5% formamide when needed to improve the amplification of GC rich sequences. The amplification is performed for a total of 35 cycles of denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds, and extension for 1 minute at 72 °C in the presence of 25 pmoles of the forward and reverse primers. The amplified rRNA fragments are used as templates to generate the single-stranded DNA complementary to the rRNA using asymmetric PCR reactions. This is achieved by reducing the concentration of the forward primer to 2.5 pmoles during the PCR amplification, which leads to an excess of the desired strand. Following asymmetric PCR, the DNA is purified using phenol-chloroform

extraction and ethanol precipitation. The DNA fragments are quantitated and mixed in equimolar amounts.

[0061] To remove rRNA from a sample comprising total RNA, 1 μg of the total RNA is incubated with 2 μg of the single-stranded DNA fragments complementary to rRNA. The hybridization and rRNA removal are performed as described above.

[0062] This example illustrates a method of preparing probes for use in the inventive method, kit, and system.

EXAMPLE 3

[0063] This example describes a method of analyzing nrRNA expression. In particular, this example describes reverse transcription and detection of p53 mRNA.

[0064] Reverse transcription is performed using the cMaster RT $_{plus}$ PCR kit (Eppendorf) using random primers or a combination of random primers and oligo dT primers according to the supplier's directions. Twenty nanograms of selected RNA (e.g., nrRNA isolated according to the method of Example 1) are used for each reaction. The reaction is stopped by heating at 70°C for 15 minutes, followed by the addition of 1 unit of Rnase H and incubation at 37°C for 20 minutes to degrade the RNA. The cDNA is now ready for RT-PCR. The transcription reaction is stored at –20°C.

[0065] To detect the p53 transcript, 1 μl of the cDNA is amplified using primers designed to amplify a portion of the p53 transcript. Examples of suitable primer pairs are:

F1: cttgccgtcccaagcaatggatg (SEQ ID NO: 185); R1: ggagcttcatctggacctgggtc (SEQ ID NO: 246) (89 base pairs),

F2: caataggtgtgcgtcagaagcacc (SEQ ID NO: 186); R2: caaaacaccagtgcaggccaacttg (SEQ ID NO: 247) (86 base pairs), and

F3: ggtctcacagtgttgcccaggctg (SEQ ID NO: 187); R3: ttgtaatcccagcactctgggag (SEQ ID NO: 248) (86 base pairs).

One microliter of the cDNA is amplified using the primer pairs shown above for 40 cycles using the MasterTaq PCR amplification kit (Eppendorf) in a 15 μl reaction containing 25 pmoles of each primer. The annealing temperature is 60°C. The reactions are separated on a 7.5% acrylamide gel and stained with ethidium bromide to detect the amplification products.

[0066] This example illustrates a method of analyzing nrRNA expression from a sample of nrRNA separated from rRNA in accordance with the invention.

EXAMPLE 4

[0067] This example describes amplification of mRNA fragments using transcription-coupled amplification.

[0068] One hundred nanograms of selected RNA fragments (e.g., nrRNA fragments isolated according to the method of Example 1) are ligated to 1 picomole of RNA oligonucleotide (GGAUGAACGUAGGAAGCUUG (SEQ ID NO: 188)) using 2 units of T4 RNA ligase in a 20 μl-reaction (NEB, Beverly, MA) according to supplier's recommendations. Following the ligation reaction, the RNA is purified by adding 80 μl of TE8 to the ligation reaction and then gently pipetting the entire volume to the top of a Sephadex G25 column (Pharmacia, Peapack, NJ). The reaction is spun at 4,000 rpm for 1 minute, and the recovered RNA is precipitated by adding 0.1 volumes of 5.2 M sodium acetate and 2 volumes of ethanol. The RNA is resuspended in 20 μl of TE8. The reverse transcription reaction is performed as described in the Superscript Choice System for cDNA synthesis (Invitrogen), except that the primers used to initiate the priming of the first strand are composed of a T7 promoter followed by around 6 to 12 random nucleotides (GGAATTAATACGACTCACTATAGGGN9 (SEQ ID NO: 189)). The reaction is treated with RNase H at 30°C for 30 minutes. The second strand synthesis is initiated using the following primer, which includes a T3 promoter (GCGCGAAATTAACCCTCACTAAA GGGATGAACGTAGGAAGCTTG (SEQ ID NO: 190)). Following the second strand synthesis, the cDNA is extracted with phenol-chloroform, ethanol precipitated, and resuspended in TE8. The cDNA (Figure 5) is ready for the amplification step using the MEGAshortscript high yield transcription kit (Ambion). The transcription reaction is performed according to the manufacturer's recommendations. Depending on the choice of promoter used for the transcription reaction, the sense or the antisense strand can be generated. The amplified RNA can serve as a template for reverse transcription and RT-PCR as described in Example 3 or as a probe for use in microarray hybridization.

[0069] This example illustrates a method of amplifying nrRNA separated from rRNA according to the method of the invention.

[0070] All references, including publications, patent applications, and patents, cited herein are hereby incorporated by reference to the same extent as if each reference were individually and specifically indicated to be incorporated by reference and were set forth in its entirety herein.

[0071] The use of the terms "a" and "an" and "the" and similar referents in the context of describing the invention (especially in the context of the following claims) are to be construed to cover both the singular and the plural, unless otherwise indicated herein or clearly contradicted by context. The terms "comprising," "having," "including," and "containing" are to be construed as open-ended terms (i.e., meaning "including, but not limited to,") unless otherwise noted. Recitation of ranges of values herein are merely intended to serve as a shorthand method of referring individually to each separate value falling within the range, unless otherwise indicated herein, and each separate value is incorporated into the specification as if it were individually recited herein. All methods described herein can be performed in any suitable order unless otherwise indicated herein or otherwise clearly contradicted by context. The use of any and all examples, or exemplary language (e.g., "such as") provided herein, is intended merely to better illuminate the invention and does not pose a limitation on the scope of the invention unless otherwise claimed. No language in the specification should be construed as indicating any non-claimed element as essential to the practice of the invention.

[0072] Preferred embodiments of this invention are described herein, including the best mode known to the inventors for carrying out the invention. Variations of those preferred embodiments may become apparent to those of ordinary skill in the art upon reading the foregoing description. The inventors expect skilled artisans to employ such variations as appropriate, and the inventors intend for the invention to be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications and equivalents of the subject matter recited in the claims appended hereto as permitted by applicable law. Moreover, any combination of the above-described elements in all possible variations thereof is encompassed by the invention unless otherwise indicated herein or otherwise clearly contradicted by context.

WHAT IS CLAIMED IS:

1. A method of separating non-ribosomal transcribed RNA (nrRNA) fragments from ribosomal RNA (rRNA) and rRNA fragments, wherein the method comprises:

(i) providing a sample comprising rRNA, rRNA fragments, and nrRNA fragments, wherein the rRNA comprises multiple contiguous regions of about 100 base pairs, each contiguous region comprising an rRNA targeting sequence, and wherein the rRNA fragments comprise one or more of the rRNA targeting sequences;

(ii) providing a plurality of probes, wherein the probes hybridize to (a) different RNA targeting sequences of at least 50% of the contiguous regions of the rRNA and (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA;

(iii) adding the plurality of probes to the sample,

(iv) hybridizing the probes to the rRNA and rRNA fragments to form rRNA-probe complexes and rRNA fragment-probe complexes, and

(v) separating the rRNA-probe complexes and rRNA fragment-probe complexes from the sample,

thereby separating nrRNA fragments from rRNA and rRNA fragments.

2. The method of claim 1, wherein the method further comprises (vi) recovering and purifying the nrRNA fragments.

3. The method of claim 1 or claim 2, wherein the sample comprises total RNA.

4. The method of any of claims 1-3, wherein the probes hybridize to rRNA targeting sequences of at least 60% of the contiguous regions.

5. The method of claim 4, wherein the probes hybridize to rRNA targeting sequences of at least 70% of the contiguous regions.

6. The method of claim 5, wherein the probes hybridize to rRNA targeting sequences of at least 80% of the contiguous regions.

7. The method of claim 6, wherein the probes hybridize to rRNA targeting sequences of at least 90% of the contiguous regions.

8. The method of claim 7, wherein the probes hybridize to rRNA targeting sequences of at least 95% of the contiguous regions.

9. The method of claim 8, wherein the probes hybridize to rRNA targeting sequences of all of the contiguous regions.

10. The method of any of claims 1-9, wherein the rRNA and the rRNA fragments comprise eukaryotic rRNA and eukaryotic rRNA fragments.

11. The method of any of claims 1-10, wherein the rRNA and rRNA fragments comprise at least two rRNAs selected from the group consisting of 28S, 18S, 5.8S, 5S, 12S and 16S rRNAs.

12. A method of amplifying nrRNA fragments, wherein the method comprises:

(i) separating nrRNA fragments from rRNA and rRNA fragments according to the method of any of claims 1-11, and

(ii) amplifying the nrRNA fragments.

13. A method of analyzing nrRNA expression, wherein the method comprises:

(i) separating nrRNA fragments from rRNA and rRNA fragments according to the method of any of claims 1-11, and

(ii) analyzing nrRNA expression.

14. A method of amplifying a cDNA complementary to an nrRNA fragment, wherein the method comprises:

(i) separating an nrRNA fragment from rRNA and rRNA fragments according to the method of any of claims 1-11, wherein the nrRNA fragment comprises a 5' end;

(ii) ligating a first oligonucleotide to the 5' end of the nrRNA fragment;

(iii) hybridizing to the nrRNA fragment a second oligonucleotide;

(iv) extending a DNA strand from the second oligonucleotide to generate a first cDNA;

(v) amplifying the first cDNA,

thereby amplifying a cDNA complementary to an nrRNA fragment.

15. A method of amplifying an nrRNA fragment, wherein the method comprises:

(i) separating an nrRNA fragment from rRNA and rRNA fragments according to the method of any of claims 1-11, wherein the nrRNA fragment comprises a 5' end and a 3' end;

(ii) ligating a first oligonucleotide to the 5' end of the nrRNA fragment;

(iii) hybridizing to the nrRNA fragment a second oligonucleotide;

(iv) extending a DNA strand from the second oligonucleotide to generate a first cDNA having a 5' end and a 3' end;

(v) degrading the nrRNA fragment and, optionally, removing any free oligonucleotides and/or free RNA fragments;

(vi) synthesizing a complementary strand to the first cDNA by

(a) contacting the first cDNA with a third oligonucleotide such that the third oligonucleotide hybridizes to the 3' end of the first cDNA, wherein the third oligonucleotide hybridizes to at least a portion of the first oligonucleotide and wherein the third oligonucleotide optionally contains an RNA polymerase promoter sequence, and

(b) extending the complementary strand from the third oligonucleotide, whereupon a double-stranded DNA copy of the nrRNA fragment is generated;

(vii) optionally purifying the double-stranded DNA copy; and

(viii) transcribing RNA from the double-stranded DNA copy,

thereby amplifying the nrRNA fragment.

16. The method of claim 15, wherein the second oligonucleotide comprises a unique sequence, a random sequence, and, optionally, an RNA polymerase promoter sequences, wherein the random sequence is of sufficient length to hybridize to the nrRNA.

17. A method of determining the level of nrRNA in a sample, wherein the method comprises:

(i) separating nrRNA fragments from rRNA and rRNA fragments according to the method of any of claims 1-11;

(ii) labeling the separated nrRNA fragments with a detectable label;

(iii) providing an array comprising a collection of fixed DNAs that hybridize to a plurality of target sequences located within an expression sequence of a gene of interest, wherein the fixed DNAs comprise from about 15 nucleotides to about 750 nucleotides, wherein at least one target sequence is interspersed approximately every 500 base pairs of the expression sequence;

(iv) applying the labeled nrRNA fragments to the array under hybridization conditions;

(v) optionally removing unbound and nonspecifically bound nrRNA fragments from the array;

(vi) detecting the labeled nrRNA fragments that have specifically annealed to the fixed DNAs in the array; and

(vii) quantitating the level of labeled nrRNA fragments detected in step (vi) to thereby determine the level of nrRNA in the sample.

18. The method of claim 17, wherein the method further comprises (viii) correlating the level of nrRNA to a level of expression of the gene of interest.

19. A kit for separating non-ribosomal transcribed RNA (nrRNA) fragments from a sample comprising ribosomal RNA (rRNA) and rRNA fragments, wherein the kit comprises a plurality of probes that hybridize to (a) different rRNA target sequences of at least 50% of contiguous regions of about 100 base pairs of the rRNA, which contiguous regions comprise the rRNA targeting sequences and (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the rRNA.

20. The kit of claim 19, wherein the probes hybridize to rRNA targeting sequences of at least 60% of the contiguous regions.

21. The kit of claim 20, wherein the probes hybridize to rRNA targeting sequences of at least 70% of the contiguous regions.

22. The kit of claim 21, wherein the probes hybridize to rRNA targeting sequences of at least 80% of the contiguous regions.

23. The kit of claim 22, wherein the probes hybridize to rRNA targeting sequences of at least 90% of the contiguous regions.

24. The kit of claim 23, wherein the probes hybridize to rRNA targeting sequences of at least 95% of the contiguous regions.

25. The kit of claim 24, wherein the probes hybridize to rRNA targeting sequences of all of the contiguous regions.

26. The kit of any of claims 19-25, wherein the rRNA and rRNA fragments comprise eukaryotic rRNA and eukaryotic rRNA fragments.

27. The kit of any of claims 19-26, wherein the rRNA and rRNA fragments comprise 28S, 18S, 5.8S, 5S, 16S, and 12S rRNAs or fragments thereof.

28. The kit of any of claims 19-27 further comprising instructions.

29. A system for separating non-ribosomal transcribed RNA (nrRNA) fragments from a sample comprising ribosomal RNA (rRNA) and rRNA fragments, the system comprising:

(i) a device adapted for the separation of nucleic acids by hybridization; and

(ii) a plurality of probes that hybridize to (a) different rRNA target sequences of at least 50% of contiguous regions of about 100 base pairs of the rRNA, which contiguous regions comprise the rRNA targeting sequences, and (b) rRNA fragments comprising the rRNA targeting sequences hybridized to by the probes in the contiguous regions of the

rRNA, wherein the plurality of probes is positioned within the device to separate the rRNA and rRNA fragments from nrRNA.

30. The system of claim 29, wherein the probes hybridize to rRNA targeting sequences of at least 60% of the contiguous regions.

31. The system of claim 30, wherein the probes hybridize to rRNA targeting sequences of at least 70% of the contiguous regions.

32. The system of claim 31, wherein the probes hybridize to rRNA targeting sequences of at least 80% of the contiguous regions.

33. The system of claim 32, wherein the probes hybridize to rRNA targeting sequences of at least 90% of the contiguous regions.

34. The system of claim 33, wherein the probes hybridize to rRNA targeting sequences of at least 95% of the contiguous regions.

35. The system of claim 34, wherein the probes hybridize to rRNA targeting sequences of all of the contiguous regions.

36. The system of any of claims 29-35, wherein the rRNA and rRNA fragments comprise 28S, 18S, 5.8S, 5S, 16S, and 12S rRNAs or fragments thereof.

28S rRNA sequence:

cgcgacctcagatcagacgtggcgacccgctgaatttaagcatattagtcagcggaggaaaagaaactaaccagg
attccctcagtaacggcgagtgaacagggaagagcccagcgccgaatccccgccccgcggggcgcgggacatgt
ggcgtacggaagacccgctccccggcgccgctcgtgggggggcccaagtccttctgatcgaggcccagcccgtgga
cggtgtgaggccggtagcggccggcgcgcgcccgggtcttcccggagtcgggttgcttgggaatgcagcccaaag
cgggtggtaaactccatctaaggctaaataccggcacgagaccgatagtcaacaagtaccgtaagggaaagttga
aaagaactttgaagagagagttcaagagggcgtgaaaccgttaagaggtaaacgggtggggtccgcgcagtccg
cccggaggattcaacccggcggcgggtccggccgtgtcggcggcccggcggatctttcccgccccccgttcctcccg
acccctccacccgccctcccttcccccgccgcccctcctcctcctcccggaggggggcgggctccggcgggtgcgg
gggtgggcgggcggggccgggggtggggtcggcgggggaccgtcccccggaccggcgaccggccgccgccgg
gcgcatttccaggcggtgcgccgcgaccggctccgggacggctgggaaggcccggcggggaaggtggctcggg
gggccccgtccgtccgtccgtcctcctcctcccccgtctccgcccccccggccccgcgtcctccctcgggagggcgcgc
gggtcggggcggcggcggcggcggcggtggcggcggcggcgggggcggcgggaccgaaaccccccccgagt
gttacagccccccccggcagcagcactcgccgaatcccggggccgagggagcgagacccgtcgccgcgctctccc
ccctcccggcgcccaccccccgcgggaatccccgcgagggggtctccccccggcgcggcgccggcgtctcctcgtg
gggggggccgggccacccctcccacggcgcgaccgctctcccacccctcctccccgcgcccccgcccccggcgacg
gggggggtgccgcgcgcgggtcgggggggcggggcggactgtccccagtgcgccccgggcgggtcgcgccgtcg
ggcccgggggaggttctctcggggccacgcgcgcgtcccccgaagaggggggacggcggagcgagcgcacggg
gtcggcggcgacgtcggctacccacccgacccgtcttgaaacacggaccaaggagtctaacacgtgcgcgagtcg
ggggctcgcacgaaagccgccgtggcgcaatgaaggtgaaggccggcgcgctcgccggccgaggtgggatccc
gaggcctctccagtccgccgaggggcaccaccggcccgtctcgcccgccgcgccggggaggtggagcacgagc
gcacgtgttaggacccgaaagatggtgaactatgcctgggcagggcgaagccagaggaaactctggtggaggtcc
gtagcggtcctgacgtgcaaatcggtcgtccgacctgggtataggggcgaaagactaatcgaaccatctagtagctg
gttccctccgaagtttccctcaggatagctggcgctctcgcagacccgacgcacccccgccacgcagttttatccggta
aagcgaatgattagaggtcttggggccgaaacgatctcaacctattctcaaactttaaatgggtaagaagcccggctc
gctggcgtggagccggggtggaatgcgagtgcctagtgggccacttttggtaagcagaactggcgctgcgggatga
accgaacgccgggttaaggcgcccgatgccgacgctcatcagaccccagaaaaggtgttggttgatatagacagc
aggacggtggccatggaagtcggaatccgctaaggagtgtgtaacaactcacctgccgaatcaactagccctgaaa
atggatggcgctggagcgtcgggcccatacccggccgtcgccggcagtcgagagtggacgggagcggcggggg
cggcggcgcgcgcgcgcgtgtggtgtgcgtcggagggcggcggcggcggcggcgggggtgtggggtccttc
ccccgccccccccccccacgcctcctcccctcctcccgcccacgccccgctccccgccccccggagccccgcggagc
tacgccgcgacgagtaggagggccgctgcggtgagccttgaagcctagggcgcgggcccgggtggaggccgcc
gcaggtgcagatcttggtggtagtagcaaatattcaaacgagaactttgaaggccgaagtggagaagggttccatgt
gaacagcagttgaacatgggtcagtcggtcctgagagatgggcgagcgccgttccgaagggacgggcgatggcct
ccgttgccctcggccgatcgaaagggagtcgggttcagatccccgaatccggagtggcggagatgggcgccgcga
ggcgtccagtgcggtaacgcgaccgatcccggagaagccggcgggagccccggggagagttctcttttctttgtgaa
gggcagggcgccctggaatgggttcgccccgagagaggggcccgtgccttggaaagcgtcgcggttccggcggc

FIG. 1A

IDT Exhibit 1020.4175

gtccggtgagctctcgctggcccttgaaaatccgggggagagggtgtaaatctcgcgccgggccgtacccatatccg
cagcaggtctccaaggtgaacagcctctggcatgttggaacaatgtaggtaagggaagtcggcaagccggatccgt
aacttcgggataaggattggctctaagggctgggtcggtcgggctggggcgcgaagcggggctgggcgcgcgccg
cggctggacgaggcgcgcgccccccccacgcccggggcacccccctcgcggccctcccccgccccacccgcgc
gcgccgctcgctccctccccacccccgcgccctctctctctctctcccccgctccccgtcctcccccctccccgggggga
gcgccgcgtgggggcgcggcgggggagaagggtcggggcggcaggggccgcgcggcggccgccgggggcg
gccggcgggggcaggtccccgcgagggggccccggggacccgggggccggcggcggcgcggactctgga
cgcgagccgggcccttcccgtggatcgccccagctgcggcgggcgtcgcggccgcccccgggggagcccggcgg
cggcgcggcgcgcccccccacccccacccccacgtctcggtcgcgcgcgcgtccgctgggggcgggagcggtcggg
cggcggcggtcggcgggcggcggggcggggcggttcgtcccccccgccctacccccccggccccgtccgcccccc
gttcccccctcctcctcggcgcgcggcggcggcggcaggcggcggagggggccgcgggccggtcccccccgc
cgggtccgcccccgggggccgcggttccgcgcgcgcctcgcctcggccggcgcctagcagccgacttagaactggt
gcggaccaggggaatccgactgtttaattaaaacaaagcatcgcgaaggcccgcggcgggtgttgacgcgatgtg
atttctgcccagtgctctgaatgtcaaagtgaagaaattcaatgaagcgcgggtaaacggcgggagtaactatgactc
tcttaaggtagccaaatgcctcgtcatctaattagtgacgcgcatgaatggatgaacgagattcccactgtccctaccta
ctatccagcgaaaccacagccaagggaacgggcttggcggaatcagcggggaaagaagaccctgttgagcttga
ctctagtctggcacggtgaagagacatgagaggtgtagaataagtgggaggcccccggcgcccccccggtgtccc
cgcgaggggcccgggggcggggtccgcggccctgcgggccgccggtgaaataccactactctgatcgttttttcactg
acccggtgaggcgggggggcgagcccgaggggctctcgcttctggcgccaagcgcccgcccggccgggcgcga
cccgctccggggacagtgccaggtggggagtttgactggggcggtacacctgtcaaacggtaacgcaggtgtccta
aggcgagctcagggaggacagaaacctcccgtggagcagaagggcaaaagctcgcttgatcttgattttcagtacg
aatacagaccgtgaaagcggggcctcacgatccttctgacctttgggttttaagcaggaggtgtcagaaaagttacc
acagggataactggcttgtggcggccaagcgttcatagcgacgtcgcttttgatccttcgatgtcggctcttcctatcatt
gtgaagcagaattcgccaagcgttggattgttcacccactaatagggaacgtgagctgggtttagaccgtcgtgagac
aggttagttttaccctactgatgatgtgttgttgccatggtaatcctgctcagtacgagaggaaccgcaggttcagacattt
ggtgtatgtgcttggctgaggagccaatggggcgaagctaccatctgtgggattatgactgaacgcctctaagtcaga
atcccgcccaggcgaacgatacggcagcgccgcggagcctcggttggcctcggatagccggtcccccgcctgtcc
ccgccggcgggccgcccccccctccacgcgccccgccgcgggagggcgcgtgccccgccgcgcgccgggacc
ggggtccggtgcggagtgcccttcgtcctgggaaacggggcgcggccggaaaggcggccgccccctcgcccgtc
acgcaccgcacgttcgtggggaacctggcgctaaaccattcgtagacgacctgcttctgggtcggggtttcgtacgta
gcagagcagctccctcgctgcgatctattgaaagtcagccctcgacacaagggtttgtc [SEQ ID NO: 1]

18S rRNA sequence:

cgctgctcctcccgtcgccgtccgggcccgtccgtccgtccgtcgtcctcctcgctnnnncggggcgccgggcc
cgtcctcacnggcccccgnnnnngtccnggcccgtcggggcctcgccgcgctctaccttacctacctggttgatcctg
ccagtagcatatgcttgtctcaaagattaagccatgcatgtctaagtacgcacggccggtacagtgaaactgcgaatg
gctcattaaatcagttatggttcctttggtcgctcgctcctctcctacttggataactgtggtaattctagagctaatacatgc

FIG. 1B

IDT Exhibit 1020.4176

cgacgggcgctgacccccttcgcgggggggatgcgtgcatttatcagatcaaaaccaacccggtcagcccctctcc
ggccccggccgggggcgggcgccggcggctttggtgactctagataacctcgggccgatcgcacgcccccgtg
gcggcgacgacccattcgaacgtctgccctatcaactttcgatggtagtcgccgtgcctaccatggtgaccacgggtg
acggggaatcagggttcgattccggagagggagcctgagaaacggctaccacatccaaggaaggcagcaggcg
cgcaaattacccactcccgacccggggaggtagtgacgaaaaataacaatacaggactctttcgaggccctgtaatt
ggaatgagtccactttaaatcctttaacgaggatccattggagggcaagtctggtgccagcagccgcggtaattccag
ctccaatagcgtatattaaagttgctgcagttaaaaagctcgtagttggatcttgggagcgggcgggcggtccgccgc
gaggcgagccaccgcccgtccccgcccttgcctctcggcgccccctcgatgctcttagctgagtgtcccgcggggc
ccgaagcgtttactttgaaaaaattagagtgttcaaagcaggcccgagccgcctggataccgcagctaggaataatg
gaataggaccgcggttctattttgttggtttcggaactgaggccatgattaagagggacggccgggggcattcgtattg
cgccgctagaggtgaaattccttggaccggcgcaagacggaccagagcgaaagcatttgccaagaatgttttcatta
atcaagaacgaaagtcggaggttcgaagacgatcagataccgtcgtagttccgaccataaacgatgccgaccggc
gatgcggcggcgttattcccatgacccgccgggcagcttccgggaaaccaaagtctttgggttccgggggggagtatg
gttgcaaagctgaaacttaaaggaattgacggaagggcaccaccaggagtggagcctgcggcttaatttgactcaa
cacgggaaacctcacccggcccggacacggacaggattgacagattgatagctctttctcgattccgtgggtggtggt
gcatggccgttcttagttggtggagcgatttgtctggttaattccgataacgaacgagactctggcatgctaactagttac
gcgacccccgagcggtcggcgtcccccaacttcttagagggacaagtggcgttcagccacccgagattgagcaata
acaggtctgtgatgcccttagatgtccggggctgcacgcgcgctacactgactggctcagcgtgtgcctaccctacgc
cggcaggcgcgggtaacccgttgaaccccattcgtgatggggatcggggattgcaattattccccatgaacgaggaa
ttcccagtaagtgcgggtcataagcttgcgttgattaagtccctgccctttgtacacaccgcccgtcgctactaccgattg
gatggtttagtgaggccctcggatcggccccgccggggtcggcccacggccctggcggagcgctgagaagacggt
cgaacttgactatctagaggaagtaaaagtcgtaacaaggtttccgtaggtgaacctgcggaaggatcattaacgga
gcccggacggcggcccgcggcggcgccgcgccgcgcttccctccgcacacccacccccccaccgcgacggcgc
gtgcgggcggggccgtgcccgttcgttcgctcgctcgttcgttcgccgcccggcccggccgcgagagccgagaactc
gggagggagacgggggagagagagagagagagagagagagagagagagagagagaaagaagggcgt
gt [SEQ ID NO: 2]

5.8S rRNA sequence:

cgactcttagcggtggatcactcggctcgtgcgtcgatgaagaacgcagctagctgcgagaattaatgtgaattgcag
gacacattgatcatcgacacttcgaacgcacttgcggccccgggttcctcccggggctacgcctgtctgagcgtcgct
[SEQ ID NO: 3]

5S rRNA sequence:

gtctacggccataccaccctgaacgcgcccgatctcgtctgatctcggaagctaagcagggtcgggcctggttagtac
ttggatgggagaccgcctgggaataccgggtgctgtaggcttt [SEQ ID NO: 4]

FIG. 1C

IDT Exhibit 1020.4177

16S rRNA sequence:

sequence:gctaaacctagccccaaacccactccaccttactaccagacaaccttagccaaaccatttacccaaat
aaagtataggcgatagaaattgaaacctggcgcaatagatatagtaccgcaagggaaagatgaaaaattataacc
aagcataatatagcaaggactaacccctataccttctgcataatgaattaactagaaataactttgcaaggagagcca
aagctaagacccccgaaaccagacgagctacctaagaacagctaaaagagcacacccgtctatgtagcaaaata
gtgggaagatttataggtagaggcgacaaacctaccgagcctggtgatagctggttgtccaagatagaatcttagttc
aactttaaatttgcccacagaaccctctaaatccccttgtaaatttaactgttagtccaaagaggaacagctctttggaca
ctaggaaaaaaaccttgtagagagagtaaaaaatttaacacccatagtaggcctaaaagcagccaccaattaagaa
agcgttcaagctcaacacccactacctaaaaaatcccaaacatataactgaactcctcacacccaattggaccaatc
tatcaccctatagaagaactaatgttagtataagtaacatgaaaacattctcctccgcataagcctgcgtcagattaaa
acactgaactgacaattaacagcccaatatctacaatcaaccaacaagtcattattaccctcactgtcaacccaacac
aggcatgctcataaggaaaggttaaaaaaagtaaaaggaactcggcaaatcttaccccgcctgtttaccaaaaaca
tcacctctagcatcaccagtattagaggcaccgcctgcccagtgacacatgtttaacggccgcggtaccctaaccgtg
caaaggtagcataatcacttgttccttaaatagggacctgtatgaatggctccacgagggttcagctgtctcttacttttaa
ccagtgaaattgacctgcccgtgaagaggcgggcataacacagcaagacgagaagaccctatggagctttaattta
ttaatgcaaacagtacctaacaaacccacaggtcctaaactaccaaacctgcattaaaaatttcggttggggcgacct
cggagcagaacccaacctccgagcagtacatgctaagacttcaccagtcaaagcgaactactatactcaattgatcc
aataacttgaccaacggaacaagttaccctagggataacagcgcaatcctattctagagtccatatcaacaataggg
tttacgacctcgatgttggatcaggacatcccgatggtgcagccgctattaaaggttcgtttgttcaacgattaaagtcct
acgtgatctgagttcagaccggagtaatccaggtcggtttctatctaccttcaaattcctccctgtacgaaaggacaaga
gaaataaggcctacttcacaaagcgccttcccccgtaaatgatatcatctcaacttagtattatacccacacccaccca
agaacagggtt [SEQ ID NO: 5]

12S rRNA sequence:

aataggtttggtcctagcctttctattagctcttagtaagattacacatgcaagcatccccgttccagtgagttcaccctcta
aatcaccacgatcaaaaggaacaagcatcaagcacgcagcaatgcagctcaaaacgcttagcctagccacaccc
ccacgggaaacagcagtgattaacctttagcaataaacgaaagtttaactaagctatactaaccccagggttggtca
atttcgtgccagccaccgcggtcacacgattaacccaagtcaatagaagccggcgtaaagagtgttttagatcaccc
cctccccaataaagctaaaactcacctgagttgtaaaaaactccagttgacacaaaatagactacgaaagtggcttta
acatatctgaacacacaatagctaagacccaaactgggattagataccccactatgcttagccctaaacctcaacag
ttaaatcaacaaaactgctcgccagaacactacgagccacagcttaaaactcaaaggacctggcggtgcttcatatc
cctctagaggagcctgttctgtaatcgataaaccccgatcaacctcaccacctcttgctcagcctatataccgccatcttc
agcaaaccctgatgaaggctacaaagtaagcgcaagtacccacgtaaagacgttaggtcaaggtgtagcccatga
ggtggcaagaaatgggctacattttctaccccagaaaactacgatagcccttatgaaacttaagggtcgaaggtggat
ttagcagtaaactaagagtagagtgcttagttgaacagggccctgaagcgcgtacacaccgcccgtcaccctcctca
agtatacttcaaaggacatttaactaaaaccccctacgcatttatatagaggagacaagtcgtaacatggtaagtgtact
ggaaagtgcacttggacgaa [SEQ ID NO: 6]

FIG. 1D

rRNA species	Sequence	Start	Length	SEQ ID NO:
18SRNA	ACGACGGACGGACGGACGGAC	50	21	SEQ ID NO. 7
18SrRNA	TAGGTAAGGTAGAGCGCGGCGA	142	22	SEQ ID NO. 8
18SrRNA	TTCACTGTACCGGCCGTGCGTA	224	22	SEQ ID NO. 9
18SrRNA	GCCCGTCGGCATGTATTAGCTC	326	22	SEQ ID NO. 10
18SrRNA	TCTAGAGTCACCAAAGCCGCCG	439	22	SEQ ID NO. 11
18SrRNA	TGGTCACCATGGTAGGCACGG	541	21	SEQ ID NO. 12
18SrRNA	GGTCGGGAGTGGGTAATTTGCG	642	22	SEQ ID NO. 13
18SrRNA	GCCCTCCAATGGATCCTCGTTA	743	22	SEQ ID NO. 14
18SrRNA	CCGCTCCCAAGATCCAACTACG	837	22	SEQ ID NO. 15
18SrRNA	GCGGGACACTCAGCTAAGAGCA	925	22	SEQ ID NO. 16
18SrRNA	TTCCTAGCTGCGGTATCCAGGC	1002	22	SEQ ID NO. 17
18SrRNA	CCGGTCCAAGGAATTTCACCTC	1117	22	SEQ ID NO. 18
18SrRNA	TCTGATCGTCTTCGAACCTCCG	1201	22	SEQ ID NO. 19
18SrRNA	GACTTTGGTTTCCCGGAAGCTG	1294	22	SEQ ID NO. 20
18SrRNA	TTAAGCCGCAGGCTCCACTCCT	1382	22	SEQ ID NO. 21
18SrRNA	CTAAGAACGGCCATGCACCACC	1486	22	SEQ ID NO. 22
18SrRNA	TCTAAGAAGTTGGGGGACGCCG	1589	22	SEQ ID NO. 23
18SrRNA	TAGGCACACGCTGAGCCAGTCA	1697	22	SEQ ID NO. 24
18SrRNA	CTGGGAATTCCTCGTTCATGGG	1789	22	SEQ ID NO. 25
18SrRNA	CCGATCCGAGGGCCTCACTAAA	1890	22	SEQ ID NO. 26
18SrRNA	GCTCCGTTAATGATCCTTCCGC	2017	22	SEQ ID NO. 27
18SrRNA	CGAACGAACGAGCGAGCGAAC	2133	21	SEQ ID NO. 28
28SrRNA	GTCGCCACGTCTGATCTGAG	26	20	SEQ ID NO. 29
28SrRNA	CTCTTCCCTGTTCACTCGCCGT	109	22	SEQ ID NO. 30
28SrRNA	CCTCGATCAGAAGGACTTGGGC	211	22	SEQ ID NO. 31
28SrRNA	TTTACCACCCGCTTTGGGCTG	310	21	SEQ ID NO. 32
28SrRNA	ACGGTTTCACGCCCTCTTGAAC	415	22	SEQ ID NO. 33
28SrRNA	CGCCGGGTTGAATCCTCC	469	18	SEQ ID NO. 34
28SrRNA	CCTCCGGGGAGGAGGAGGAGG	579	21	SEQ ID NO. 35
28SrRNA	GCACCGCCTGGAAATGCGCC	692	20	SEQ ID NO. 36

FIG. 2A

IDT Exhibit 1020.4179

rRNA species	Sequence	Start	Length	SEQ ID NO:
28SrRNA	GATTCGGCGAGTGCTGCT	930	18	SEQ ID NO. 37
28SrRNA	AGAGCGGTCGCGCCGTGGGAG	1082	21	SEQ ID NO. 38
28SrRNA	CGTGGCCCCGAGAGAACCTC	1220	20	SEQ ID NO. 39
28SrRNA	GTCCGTGTTTCAAGACGGGTCG	1311	22	SEQ ID NO. 40
28SrRNA	CGCCGGCCTTCACCTTCATT	1386	20	SEQ ID NO. 41
28SrRNA	TCGGGTCCTAACACGTGCGCT	1502	21	SEQ ID NO. 42
28SrRNA	TTTCGCCCCTATACCCAGGTCG	1611	22	SEQ ID NO. 43
28SrRNA	TCGGCCCCAAGACCTCTAATCA	1744	22	SEQ ID NO. 44
28SrRNA	ACTAGGCACTCGCATTCCACCC	1830	22	SEQ ID NO. 45
28SrRNA	GGGTCTGATGAGCGTCGGCAT	1915	21	SEQ ID NO. 46
28SrRNA	GGGCTAGTTGATTCGGCAGGTG	2015	22	SEQ ID NO. 47
28SrRNA	CCCTCCGACGCACACCACAC	2130	20	SEQ ID NO. 48
28SrRNA	AGGCTTCAAGGCTCACCGCAG	2288	21	SEQ ID NO. 49
28SrRNA	CCCTTCTCCACTTCGGCCTTCA	2384	22	SEQ ID NO. 50
28SrRNA	TCGCCCATCTCTCAGGACCGA	2437	21	SEQ ID NO. 51
28SrRNA	TTCGGGGATCTGAACCCGACT	2514	21	SEQ ID NO. 52
28SrRNA	CGCCCTGCCCTTCACAAAGAAA	2629	22	SEQ ID NO. 53
28SrRNA	TTTTCAAGGGCCAGCGAGAGC	2723	21	SEQ ID NO. 54
28SrRNA	TGTTCCAACATGCCAGAGGCTG	2812	22	SEQ ID NO. 55
28SrRNA	CGACCCAGCCCTTAGAGCCAA	2883	21	SEQ ID NO. 56
28SrRNA	GTGGGGAGGGAGCGAGCGG	3016	19	SEQ ID NO. 57
28SrRNA	CTGCCGCCCCGACCCTTCTC	3119	20	SEQ ID NO. 58
28SrRNA	CGGCTCGCGTCCAGAGTCC	3224	19	SEQ ID NO. 59
28SrRNA	CCGAGGAGGAGGGGGGAAC	3455	19	SEQ ID NO. 60
28SrRNA	CGCACCAGTTCTAAGTCGGCTG	3588	22	SEQ ID NO. 61
28SrRNA	TTGACATTCAGAGCACTGGGCA	3687	22	SEQ ID NO. 62
28SrRNA	CGTTCATCCATTCATGCGCGTC	3796	22	SEQ ID NO. 63
28SrRNA	TGATTCCGCCAAGCCCGTTC	3863	20	SEQ ID NO. 64
28SrRNA	CGGGGGCCTCCCACTTATTCTA	3949	22	SEQ ID NO. 65
28SrRNA	CGCCTCACCGGGTCAGTGAAA	4057	21	SEQ ID NO. 66

FIG. 2B

IDT Exhibit 1020.4180

rRNA species	Sequence	Start	Length	SEQ ID NO:
28SrRNA	TCAAACTCCCCACCTGGCACTG	4152	22	SEQ ID NO. 67
28SrRNA	CCTTCTGCTCCACGGGAGGTTT	4236	22	SEQ ID NO. 68
28SrRNA	ATCGTGAGGCCCCGCTTTCAC	4300	21	SEQ ID NO. 69
28SrRNA	CGACGTCGCTATGAACGCTTGG	4393	22	SEQ ID NO. 70
28SrRNA	GTGAACAATCCAACGCTTGGCG	4464	22	SEQ ID NO. 71
28SrRNA	TCTGAACCTGCGGTTCCTCTCG	4584	22	SEQ ID NO. 72
28SrRNA	GCCTGGGCGGGATTCTGACTTA	4680	22	SEQ ID NO. 73
28SrRNA	GTTTCCCAGGACGAAGGGCA	4850	20	SEQ ID NO. 74
28SrRNA	CGACCCAGAAGCAGGTCGTCTA	4952	22	SEQ ID NO. 75
28SrRNA	CAAACCCTTGTGTCGAGGGCTG	5023	22	SEQ ID NO. 76
5.8S rRNA	TCATCGACGCACGAGCCGAG	40	20	SEQ ID NO. 77
5.8S rRNA	GCAAGTGCGTTCGAAGTGTCG	113	21	SEQ ID NO. 78
5.8S rRNA	AGCGACGCTCAGACAGGCGTA	156	21	SEQ ID NO. 79
5SrRNA	CTCTTAGCGGTGGATCACTCGG	25	22	SEQ ID NO. 80
5SrRNA	GTCGATGATCAATGTGTCCTGCAA	96	24	SEQ ID NO. 81
5SrRNA	AGCGACGCTCAGACAGGCGTAG	156	22	SEQ ID NO. 82
12SRNA	ACTCACTGGAACGGGGATGCTTG	73	23	SEQ ID NO. 83
12SrRNA	TGCGTGCTTGATGCTTGTTCCT	121	22	SEQ ID NO. 84
12SrRNA	GGCACGAAATTGACCAACCCTG	244	22	SEQ ID NO. 85
12SrRNA	ACGCCGGCTTCTATTGACTTGG	290	22	SEQ ID NO. 86
12SrRNA	GGTCTTAGCTATTGTGTGTTCAG	418	23	SEQ ID NO. 87
12SrRNA	TGTGGCTCGTAGTGTTCTGGCG	507	22	SEQ ID NO. 88
12SrRNA	AGAGGGATATGAAGCACCGCCA	548	22	SEQ ID NO. 89
12SrRNA	AGAGGTGGTGAGGTTGATCGGG	598	22	SEQ ID NO. 90
12SrRNA	TCAGGGTTTGCTGAAGATGGCG	636	22	SEQ ID NO. 91
12SrRNA	TTGCCACCTCATGGGCTACACC	707	22	SEQ ID NO. 92
12SrRNA	GGCGGTGTGTACGCGCTTCAGG	840	22	SEQ ID NO. 93
12SrRNA	CCATGTTACGACTTGTCTCCTCT	923	23	SEQ ID NO. 94
16SrRNA	GTCTGGTAGTAAGGTGGAGTGG	41	22	SEQ ID NO. 95
16SrRNA	TTCATCTTTCCCTTGCGGTACT	132	22	SEQ ID NO. 96

FIG. 2C

IDT Exhibit 1020.4181

rRNA species	Sequence	Start	Length	SEQ ID NO:
16SrRNA	CTGTTCTTAGGTAGCTCGTCTGG	262	23	SEQ ID NO. 97
16SrRNA	AACCAGCTATCACCAGGCTCGG	351	22	SEQ ID NO. 98
16SrRNA	CCTAGTGTCCAAAGAGCTGTTCC	459	23	SEQ ID NO. 99
16SrRNA	TGGGTGTTGAGCTTGAACGCTT	550	22	SEQ ID NO. 100
16SrRNA	TCTGACGCAGGCTTATGCGGAG	680	22	SEQ ID NO. 101
16SrRNA	TGAGCATGCCTGTGTTGGGTTG	774	22	SEQ ID NO. 102
16SrRNA	CCCTCGTGGAGCCATTCATACA	977	22	SEQ ID NO. 103
16SrRNA	TCGTCTTGCTGTGTTATGCCCG	1050	22	SEQ ID NO. 104
16SrRNA	ACTGCTCGGAGGTTGGGTTCTG	1180	22	SEQ ID NO. 105
16SrRNA	AGGGTAACTTGTTCCGTTGGTC	1260	22	SEQ ID NO. 106
16SrRNA	TTAATAGCGGCTGCACCATCGG	1359	22	SEQ ID NO. 107
16SrRNA	CTCCGGTCTGAACTCAGATCAC	1412	22	SEQ ID NO. 108
16SrRNA	GGGGAAGGCGCTTTGTGAAGTA	1501	22	SEQ ID NO. 109
16SrRNA	ACCCTGTTCTTGGGTGGGTGTG	1557	22	SEQ ID NO. 110

FIG. 2D

IDT Exhibit 1020.4182

rRNA	Left Primer	SEQ ID NO:	Right primer	SEQ ID NO:	Product size	Left	Right
28SrRNA	GACCTCAGATCAGACGTGGCGA	SEQ ID NO. 111	AGCAACCCGACTCCGGGAAGA	SEQ ID NO. 191	278	4,22	281,21
28SrRNA	GTCGGGTTGCTTGGGAATGC	SEQ ID NO. 112	GACCCGCCGCCGGGTTGAAT	SEQ ID NO. 192	206	271,20	476,20
28SrRNA	TCCGCCCGGAGGATTCAA	SEQ ID NO. 113	CGCACCGCCTGGAAATGC	SEQ ID NO. 193	249	445,18	693,18
28SrRNA	GCATTTCCAGGCGGTGCG	SEQ ID NO. 114	GATTCGGCGAGTGCTGCT	SEQ ID NO. 194	255	676,18	930,18
28SrRNA	AGCAGCACTCGCCGAATC	SEQ ID NO. 115	TGGCCCCGAGAGAACCTC	SEQ ID NO. 195	306	913,18	1218,18
28SrRNA	ACCCCTCCCACGGCGCGA	SEQ ID NO. 116	CGTGTTTCAAGACGGGTCGG	SEQ ID NO. 196	251	1058,18	1308,20
28SrRNA	CACGGACCAAGGAGTCTAACA	SEQ ID NO. 117	CCAGGCATAGTTCACCATCTTT	SEQ ID NO. 197	219	1305,21	1523,22
28SrRNA	CGAAAGATGGTGAACTATGCCT	SEQ ID NO. 118	GCCCCAAGACCTCTAATCATTC	SEQ ID NO. 198	242	1500,22	1741,22
28SrRNA	CCGAAACGATCTCAACCTATTC	SEQ ID NO. 119	TGTTACACACTCCTTAGCGGATT	SEQ ID NO. 199	250	1741,22	1990,23
28SrRNA	CACCTGCCGAATCAACTAGCCC	SEQ ID NO. 120	AGGCTTCAAGGCTCACCGCA	SEQ ID NO. 201	295	1994,22	2288,20
28SrRNA	CCACGCCTCCTCCCCTCCT	SEQ ID NO. 121	CAGGACCGACTGACCCATGTTC	SEQ ID NO. 201	244	2182,19	2425,22
28SrRNA	GAACAGCAGTTGAACATGGGT	SEQ ID NO. 122	CTGCCCTTCACAAAGAAAAGAG	SEQ ID NO. 202	233	2393,21	2625,22
28SrRNA	CGGGGAGAGTTCTCTTTTCTTT	SEQ ID NO. 123	GCCGACTTCCCTTACCTACATT	SEQ ID NO. 203	241	2593,22	2833,22
28SrRNA	CAATGTAGGTAAGGGAAGTCGG	SEQ ID NO. 124	GAGAGAGAGAGAGAGAGGGCG	SEQ ID NO. 204	231	2811,22	3041,21
28SrRNA	CTCTCTCTCTCCCCCGCTCC	SEQ ID NO. 125	AGCTGGGGCGATCCACGGGA	SEQ ID NO. 205	220	3031,20	3250,20
28SrRNA	TCCCGTGGATCGCCCCAGCT	SEQ ID NO. 126	CCGAGGAGGAGGGGGGAACG	SEQ ID NO. 206	225	3231,20	3455,20
28SrRNA	GGCGGGGCGGGGCGGTTCGT	SEQ ID NO. 127	CGCACCAGTTCTAAGTCGGCTG	SEQ ID NO. 207	208	3381,20	3588,22

FIG. 3A

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.4183

rRNA	Left Primer	SEQ ID NO:	Right primer	SEQ ID NO:	Product size	Left	Right
28SrRNA	CCTAGCAGCCGACTTAGAACTG	SEQ ID NO. 128	GACAGTGGGAATCTCGTTCATC	SEQ ID NO. 208	249	3562,22	3810,22
28SrRNA	TCTCTTAAGGTAGCCAAATGCC	SEQ ID NO. 129	GGGCCTCCCACTTATTCTACA	SEQ ID NO. 209	210	3737,22	3946,21
28SrRNA	CTACTATCCAGCGAAACCACAG	SEQ ID NO. 130	AGAGTAGTGGTATTTCACCGGC	SEQ ID NO. 210	213	3816,22	4028,22
28SrRNA	CGGGGAAAGAAGACCCTGTTGA	SEQ ID NO. 131	CTCGCCTTAGGACACCTGCGT	SEQ ID NO. 211	336	3865,22	4200,21
28SrRNA	CTCAGGGAGGACAGAAACCTC	SEQ ID NO. 132	AATGATAGGAAGAGCCGACATC	SEQ ID NO. 212	229	4201,21	4429,22
28SrRNA	GATGTCGGCTCTTCCTATCATT	SEQ ID NO. 133	CGTTCAGTCATAATCCCACAGA	SEQ ID NO. 213	247	4408,22	4654,22
28SrRNA	ACTAATAGGGAACGTGAGCTGG	SEQ ID NO. 134	GGATTCTGACTTAGAGGCGTTC	SEQ ID NO. 214	205	4467,22	4671,22
28SrRNA	CTAAGTCAGAATCCCGCCCAGG	SEQ ID NO. 135	GTTTAGCGCCAGGTTCCCCAC	SEQ ID NO. 215	267	4658,22	4924,21
18SrRNA	TCGCCGCGCTCTACCTTACCTA	SEQ ID NO. 136	GCCCGTCGGCATGTATTAGCTC	SEQ ID NO. 216	206	121,22	326,22
18SrRNA	GAGCTAATACATGCCGACGGGC	SEQ ID NO. 137	GAATCGAACCCTGATTCCCCGT	SEQ ID NO. 217	265	305,22	569,22
18SrRNA	AACGTCTGCCCTATCAACTTTC	SEQ ID NO. 138	TGGACTCATTCCAATTACAGGG	SEQ ID NO. 218	222	488,22	709,22
18SrRNA	GGCCCTGTAATTGGAATGAGT	SEQ ID NO. 139	GGGACACTCAGCTAAGAGCATC	SEQ ID NO. 219	238	686,21	923,22
18SrRNA	TGCTCTTAGCTGAGTGTCCCGC	SEQ ID NO. 140	GCCGGTCCAAGGAATTTCACC	SEQ ID NO. 220	215	904,22	1118,21
18SrRNA	TAGAGGTGAAATTCCTTGGACC	SEQ ID NO. 141	GTTTCAGCTTTGCAACCATACTC	SEQ ID NO. 221	239	1094,22	1332,23
18SrRNA	TGAAACTTAAAGGAATTGACGGA	SEQ ID NO. 142	AGTTAGCATGCCAGAGTCTCGT	SEQ ID NO. 222	219	1327,23	1545,22
18SrRNA	CACGGACAGGATTGACAGATT	SEQ ID NO. 143	CATCTAAGGGCATCACAGACCT	SEQ ID NO. 223	234	1418,21	1651,22
18SrRNA	CCCAACTTCTTAGAGGGACAAGT	SEQ ID NO. 144	AGCTTATGACCCGCACTTACTG	SEQ ID NO. 224	233	1576,23	1808,22

FIG. 3B

SUBSTITUTE SHEET (RULE 26)

rRNA	Left Primer	SEQ ID NO:	Right primer	SEQ ID NO:	Product size	Left	Right
18SrRNA	CTGTGATGCCCTTAGATGTCC	SEQ ID NO. 145	CTCACTAAACCATCCAATCGGT	SEQ ID NO. 225	244	1634,21	1877,22
18SrRNA	ATGAACGAGGAATTCCCAGTAAG	SEQ ID NO. 146	CCTACGGAAACCTTGTTACGAC	SEQ ID NO. 226	218	1771,23	1988,22
18SrRNA	ACCTGCGGAAGGATCATTAAC	SEQ ID NO. 147	ACGCCCTTCTTTCTCTCTCTCT	SEQ ID NO. 227	242	1992,21	2233,22
16SrRNA	CCCACTCCACCTTACTACCAGA	SEQ ID NO. 148	GGTCTTAGCTTTGGCTCTCCTT	SEQ ID NO. 228	214	19,22	232,22
16SrRNA	CAAGGAGAGCCAAAGCTAAGAC	SEQ ID NO. 149	AGTGTCCAAAGAGCTGTTCCTC	SEQ ID NO. 229	247	210,22	456,22
16SrRNA	AGAACCCTCTAAATCCCCTTGT	SEQ ID NO. 150	ATTGGGTGTGAGGAGTTCAGTT	SEQ ID NO. 230	205	392,22	596,22
16SrRNA	CAAAGAGGAACAGCTCTTTGGA	SEQ ID NO. 151	CGGAGGAGAATGTTTTCATGTT	SEQ ID NO. 231	233	431,22	663,22
16SrRNA	TAACTGAACTCCTCACACCCAA	SEQ ID NO. 152	CTTTCCTTATGAGCATGCCTGT	SEQ ID NO. 232	210	574,22	783,22
16SrRNA	ACAGGCATGCTCATAAGGAAAG	SEQ ID NO. 153	TAAGAGACAGCTGAACCCTCGT	SEQ ID NO. 233	231	762,22	992,22
16SrRNA	ACCGTGCAAAGGTAGCATAATC	SEQ ID NO. 154	GCAGGTTTGGTAGTTTAGGACC	SEQ ID NO. 234	216	912,22	1127,22
16SrRNA	AGCAAGACGAGAAGACCCTATG	SEQ ID NO. 155	CTAGGGTAACTTGTTCCGTTGG	SEQ ID NO. 235	222	1041,22	1262,22
16SrRNA	TTCACCAGTCAAAGCGAACTAC	SEQ ID NO. 156	GGTAGATAGAAACCGACCTGGA	SEQ ID NO. 236	245	1193,22	1437,22
16SrRNA	GATCCAATAACTTGACCAACGG	SEQ ID NO. 157	CTCTTGTCCTTTCGTACAGGGA	SEQ ID NO. 237	244	1226,22	1469,22
12SrRNA	AAGATTACACATGCAAGCATCC	SEQ ID NO. 158	TGACTTGGGTTAATCGTGTGAC	SEQ ID NO. 238	239	38,22	276,22
12SrRNA	GTCACACGATTAACCCAAGTCA	SEQ ID NO. 159	GTAGTGTTCTGGCGAGCAGTT	SEQ ID NO. 239	245	255,22	499,21
12SrRNA	GATACCCCACTATGCTTAGCCC	SEQ ID NO. 160	CTTGCGCTTACTTTGTAGCCTT	SEQ ID NO. 240	229	432,22	660,22
12SrRNA	TTCTGTAATCGATAAACCCCGA	SEQ ID NO. 161	ATCCACCTTCGACCCTTAAGTT	SEQ ID NO. 241	218	560,22	777,22

FIG. 3C

rRNA	Left Primer	SEQ ID NO:	Right primer	SEQ ID NO:	Product size	Left	Right
12SrRNA	GATCAACCTCACCACCTCTTG	SEQ ID NO. 162	GCCCTGTTCAACTAAGCACTCT	SEQ ID NO. 242	239	580,21	818,22
12SrRNA	GGCTACATTTTCTACCCCAGAA	SEQ ID NO. 163	TGCACTTTCCAGTACACTTACCA	SEQ ID NO. 243	230	714,22	943,23
5_8SrRNA	GACTCTTAGCGGTGGATCACTC	SEQ ID NO. 164	TTCGAAGTGTCGATGATCAATGT	SEQ ID NO. 244	103	2,22	104,23
5SrRNA	GTCTACGGCCATACCACCCT	SEQ ID NO. 165	CGGTCTCCCATCCAAGTACTAA	SEQ ID NO. 245	93	1,20	93,22

FIG. 3D

IDT Exhibit 1020.4186

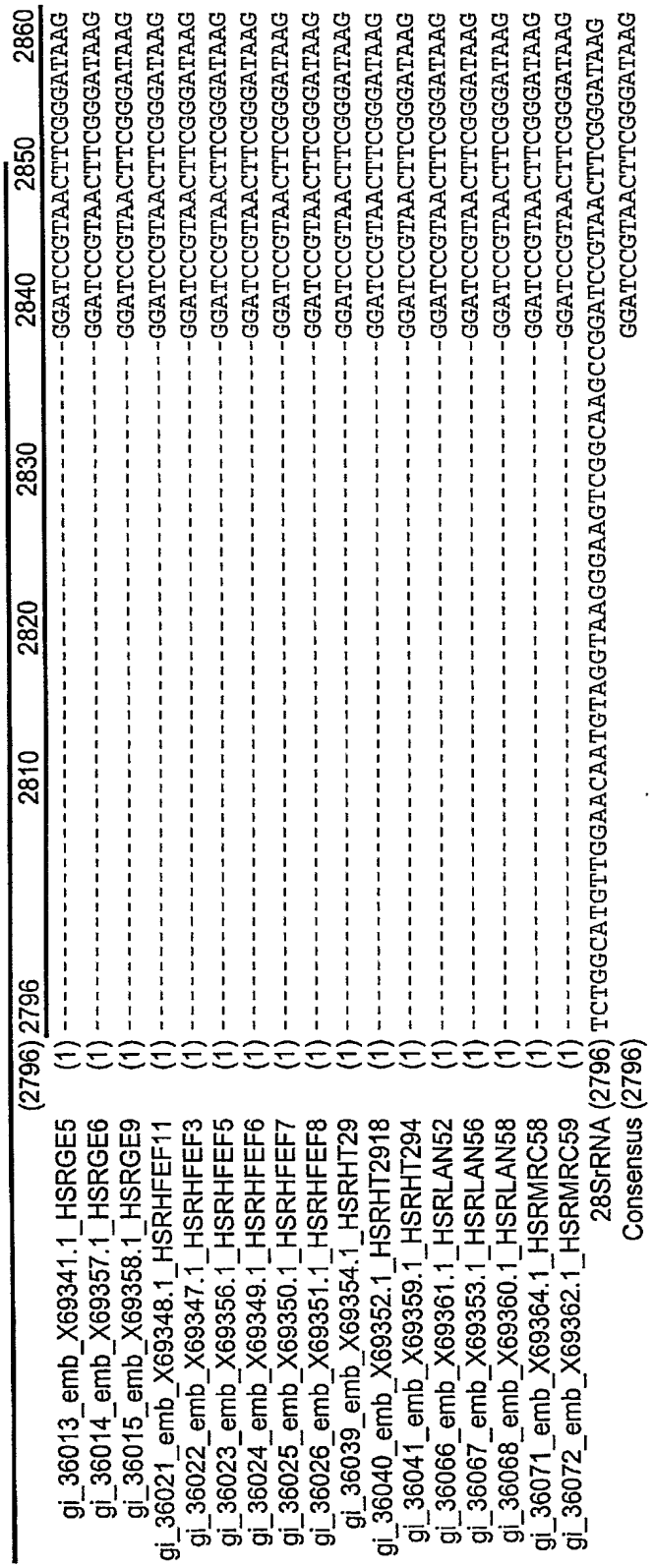

Section 44
(2796) 2796 2810 2820 2830 2840 2850 2860
gi_36013_emb_X69341.1_HSRGE5 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36014_emb_X69357.1_HSRGE6 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36015_emb_X69358.1_HSRGE9 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36021_emb_X69348.1_HSRHFEF11 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36022_emb_X69347.1_HSRHFEF3 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36023_emb_X69356.1_HSRHFEF5 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36024_emb_X69349.1_HSRHFEF6 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36025_emb_X69350.1_HSRHFEF7 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36026_emb_X69351.1_HSRHFEF8 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36039_emb_X69354.1_HSRHT29 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36040_emb_X69352.1_HSRHT2918 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36041_emb_X69359.1_HSRHT294 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36066_emb_X69361.1_HSRLAN52 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36067_emb_X69353.1_HSRLAN56 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36068_emb_X69360.1_HSRLAN58 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36071_emb_X69364.1_HSRMRC58 (1) ----------GGATCCGTAACTTCGGGATAAG
gi_36072_emb_X69362.1_HSRMRC59 (1) ----------GGATCCGTAACTTCGGGATAAG
28SrRNA (2796) TCTGGCATGTTGGAACAATGTAGGTAAGGGAAGTCGGCAAGCCGGATCCGTAACTTCGGGATAAG
Consensus (2796) GGATCCGTAACTTCGGGATAAG

FIG. 4A

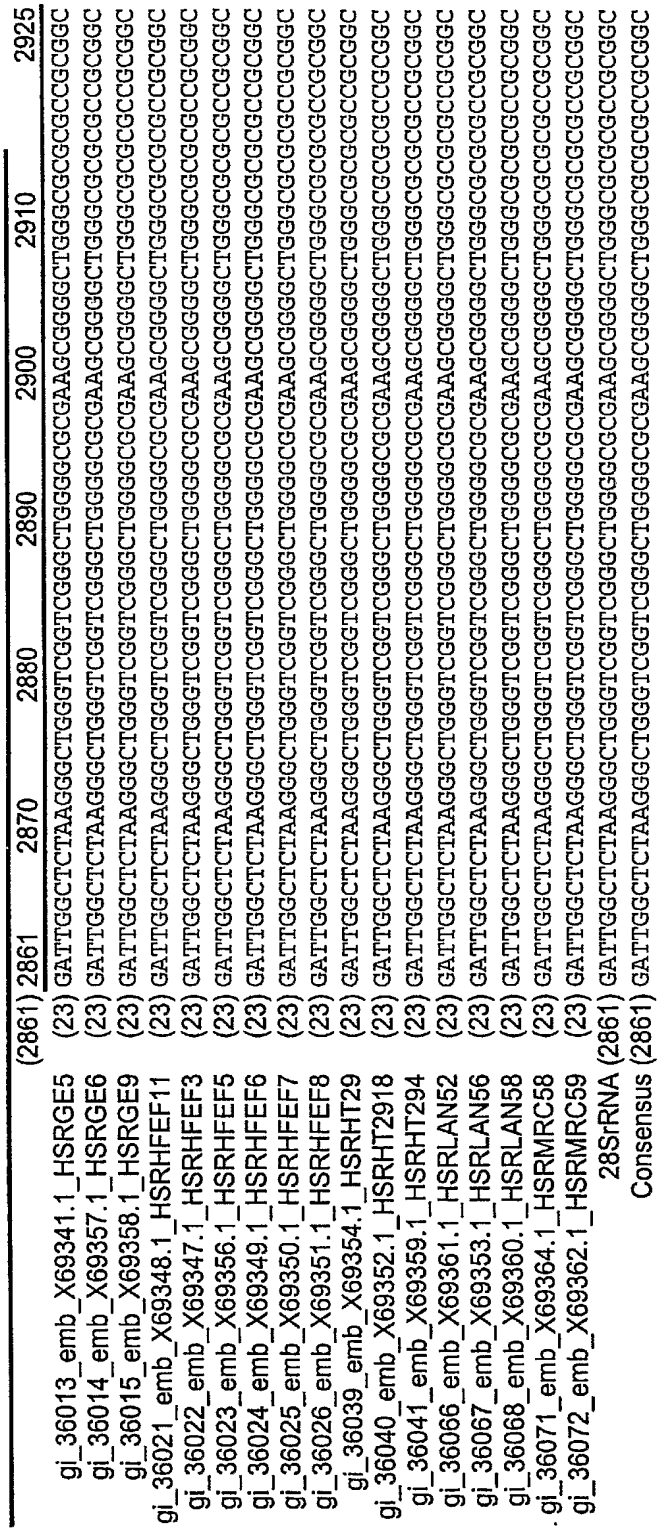

Section 45
2861
2870
2880
2890
2900
2910
2925
(2861)
gi_36013_emb_X69341.1_HSRGE5 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36014_emb_X69357.1_HSRGE6 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36015_emb_X69358.1_HSRGE9 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36021_emb_X69348.1_HSRHFEF11 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36022_emb_X69347.1_HSRHFEF3 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36023_emb_X69356.1_HSRHFEF5 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36024_emb_X69349.1_HSRHFEF6 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36025_emb_X69350.1_HSRHFEF7 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36026_emb_X69351.1_HSRHFEF8 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36039_emb_X69354.1_HSRHT29 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36040_emb_X69352.1_HSRHT2918 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36041_emb_X69359.1_HSRHT294 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36066_emb_X69361.1_HSRLAN52 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36067_emb_X69353.1_HSRLAN56 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36068_emb_X69360.1_HSRLAN58 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36071_emb_X69364.1_HSRMRC58 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
gi_36072_emb_X69362.1_HSRMRC59 (23) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
28SrRNA (2861) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC
Consensus (2861) GATTGGCTCTAAGGGCTGGGTCGGTCGGGCTGGGGCGCGAAGCGGGGCTGGGCGCGCCGGC

FIG. 4B

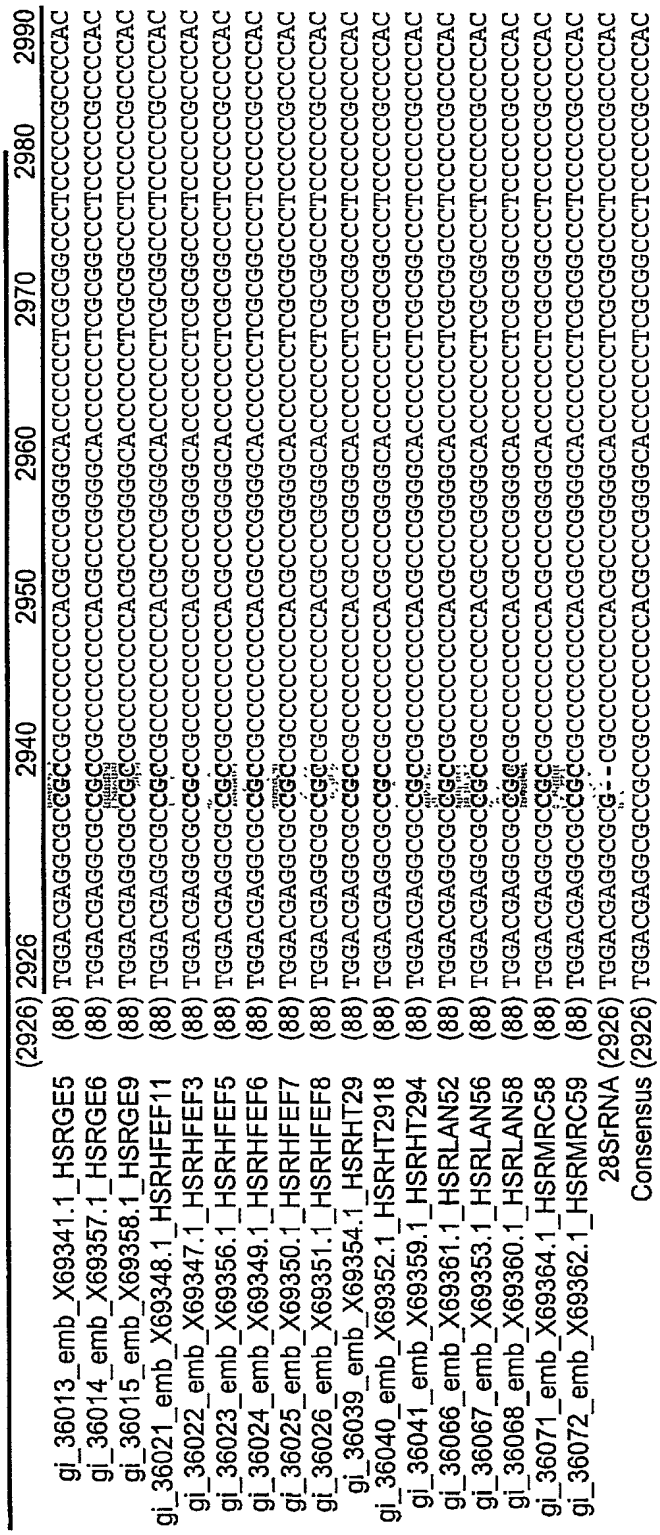

Section 46
2926
2940
2950
2960
2970
2980
2990
(2926)
gi_36013_emb_X69341.1_HSRGE5 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36014_emb_X69357.1_HSRGE6 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36015_emb_X69358.1_HSRGE9 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36021_emb_X69348.1_HSRHFEF11 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36022_emb_X69347.1_HSRHFEF3 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36023_emb_X69356.1_HSRHFEF5 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36024_emb_X69349.1_HSRHFEF6 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36025_emb_X69350.1_HSRHFEF7 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36026_emb_X69351.1_HSRHFEF8 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36039_emb_X69354.1_HSRHT29 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36040_emb_X69352.1_HSRHT2918 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36041_emb_X69359.1_HSRHT294 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36066_emb_X69361.1_HSRLAN52 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36067_emb_X69353.1_HSRLAN56 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36068_emb_X69360.1_HSRLAN58 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36071_emb_X69364.1_HSRMRC58 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
gi_36072_emb_X69362.1_HSRMRC59 (88) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
28SrRNA (2926) TGGACGAGGCGCG--CGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC
Consensus (2926) TGGACGAGGCGCCGCCCCCCCCACGCCCGGGGCACCCCCCTCGCGGCCCTCCCCCGCCCCAC

FIG. 4C

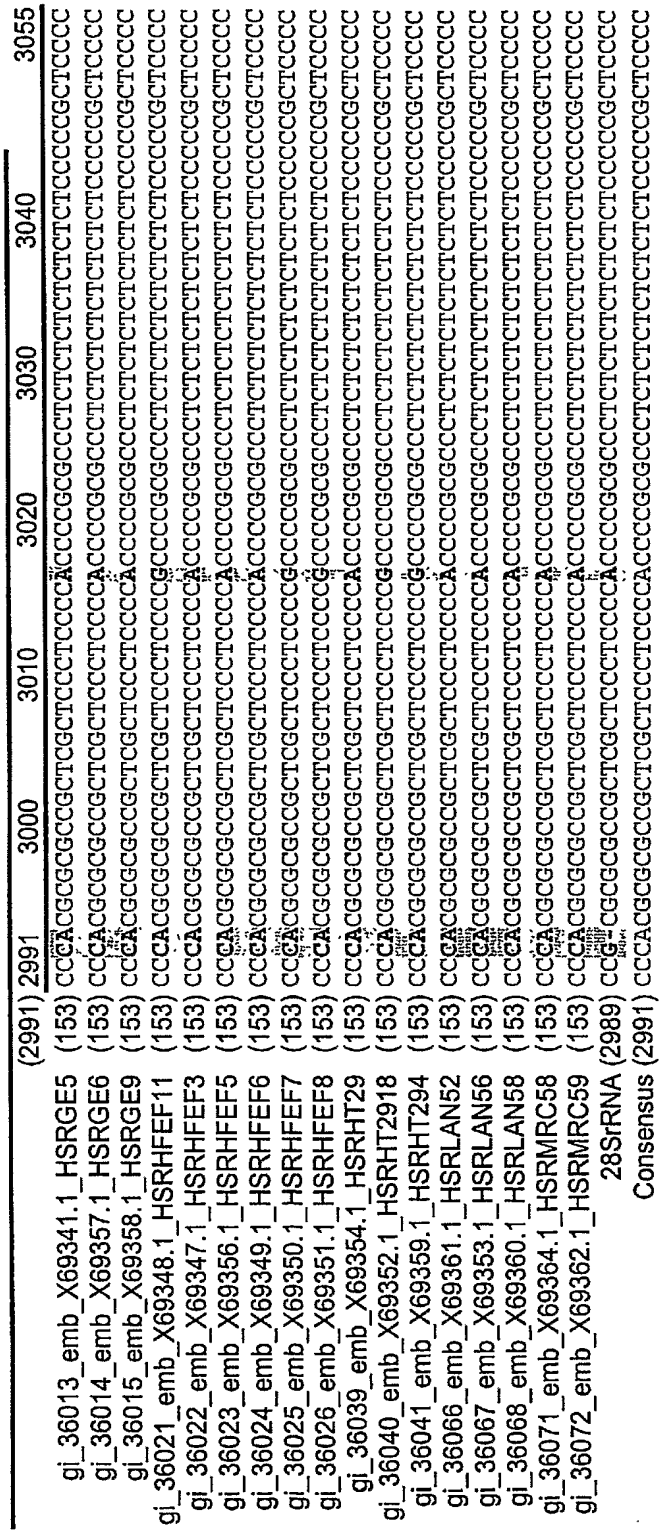

Section 47
(2991) 2991 3000 3010 3020 3030 3040 3055
gi_36013_emb_X69341.1_HSRGE5 (153)
gi_36014_emb_X69357.1_HSRGE6 (153)
gi_36015_emb_X69358.1_HSRGE9 (153)
gi_36021_emb_X69348.1_HSRHFEF11 (153)
gi_36022_emb_X69347.1_HSRHFEF3 (153)
gi_36023_emb_X69356.1_HSRHFEF5 (153)
gi_36024_emb_X69349.1_HSRHFEF6 (153)
gi_36025_emb_X69350.1_HSRHFEF7 (153)
gi_36026_emb_X69351.1_HSRHFEF8 (153)
gi_36039_emb_X69354.1_HSRHT29 (153)
gi_36040_emb_X69352.1_HSRHT2918 (153)
gi_36041_emb_X69359.1_HSRHT294 (153)
gi_36066_emb_X69361.1_HSRLAN52 (153)
gi_36067_emb_X69353.1_HSRLAN56 (153)
gi_36068_emb_X69360.1_HSRLAN58 (153)
gi_36071_emb_X69364.1_HSRMRC58 (153)
gi_36072_emb_X69362.1_HSRMRC59 (153)
28SrRNA (2989)
Consensus (2991)

FIG. 4D

Section 48
(3056) 3056 3070 3080 3090 3100 3110 3120
gi_36013_emb_X69341.1_HSRGE5 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGGAGAAGGGTCGGGGCGG
gi_36014_emb_X69357.1_HSRGE6 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36015_emb_X69358.1_HSRGE9 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36021_emb_X69348.1_HSRHFEF11 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36022_emb_X69347.1_HSRHFEF3 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36023_emb_X69356.1_HSRHFEF5 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36024_emb_X69349.1_HSRHFEF6 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36025_emb_X69350.1_HSRHFEF7 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36026_emb_X69351.1_HSRHFEF8 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36039_emb_X69354.1_HSRHT29 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36040_emb_X69352.1_HSRHT2918 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36041_emb_X69359.1_HSRHT294 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36066_emb_X69361.1_HSRLAN52 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36067_emb_X69353.1_HSRLAN56 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36068_emb_X69360.1_HSRLAN58 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36071_emb_X69364.1_HSRMRC58 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
gi_36072_emb_X69362.1_HSRMRC59 (218) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
28SrRNA (3053) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG
Consensus (3056) GTCCTCCCCCCTCCCGGGGAGCGCCGCGTGGGGGCGGCGGGGGGAGAAGGGTCGGGGCGG

FIG. 4E

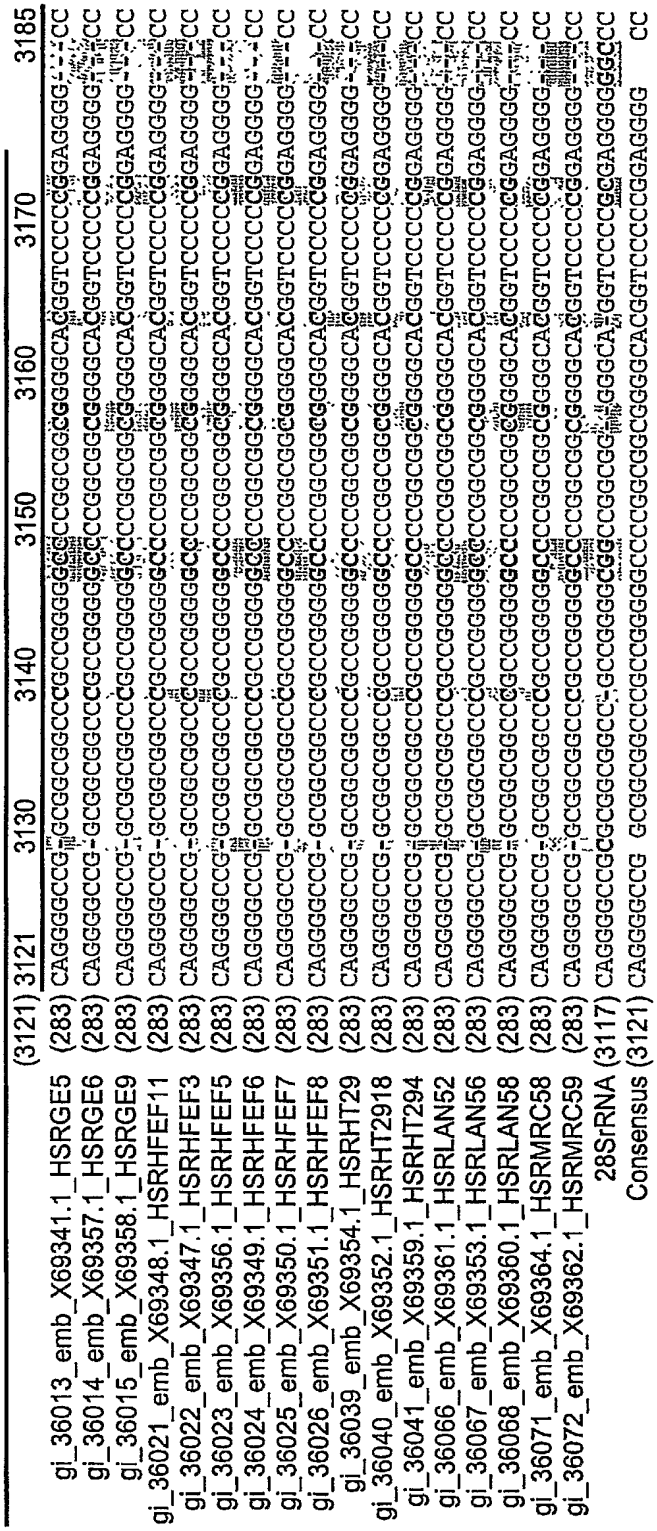

Section 49
(3121) 3121 3130 3140 3150 3160 3170 3185
gi_36013_emb_X69341.1_HSRGE5 (283)
gi_36014_emb_X69357.1_HSRGE6 (283)
gi_36015_emb_X69358.1_HSRGE9 (283)
gi_36021_emb_X69348.1_HSRHFEF11 (283)
gi_36022_emb_X69347.1_HSRHFEF3 (283)
gi_36023_emb_X69356.1_HSRHFEF5 (283)
gi_36024_emb_X69349.1_HSRHFEF6 (283)
gi_36025_emb_X69350.1_HSRHFEF7 (283)
gi_36026_emb_X69351.1_HSRHFEF8 (283)
gi_36039_emb_X69354.1_HSRHT29 (283)
gi_36040_emb_X69352.1_HSRHT2918 (283)
gi_36041_emb_X69359.1_HSRHT294 (283)
gi_36066_emb_X69361.1_HSRLAN52 (283)
gi_36067_emb_X69353.1_HSRLAN56 (283)
gi_36068_emb_X69360.1_HSRLAN58 (283)
gi_36071_emb_X69364.1_HSRMRC58 (283)
gi_36072_emb_X69362.1_HSRMRC59 (283)
28SrRNA (3117)
Consensus (3121)

FIG. 4F

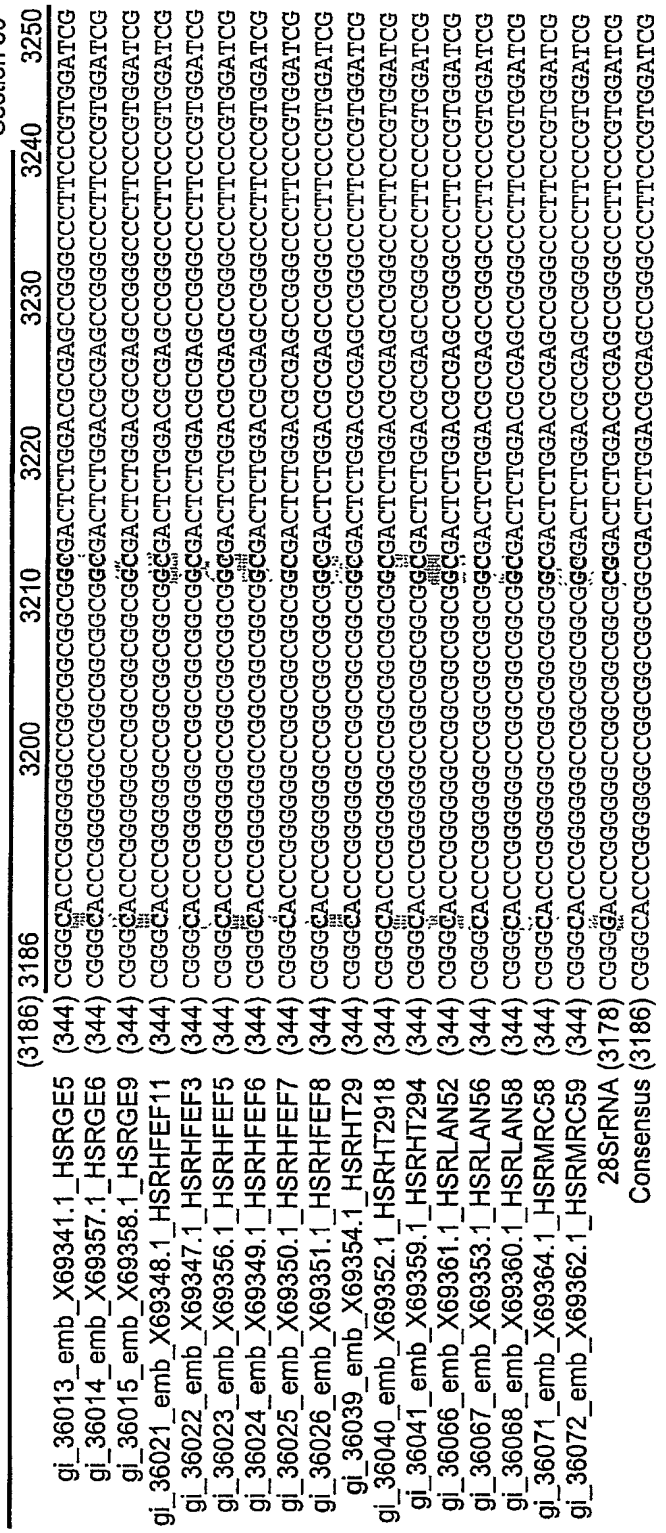

Section 50
(3186) 3186 3200 3210 3220 3230 3240 3250
gi_36013_emb_X69341.1_HSRGE5 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36014_emb_X69357.1_HSRGE6 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36015_emb_X69358.1_HSRGE9 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36021_emb_X69348.1_HSRHFEF11 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36022_emb_X69347.1_HSRHFEF3 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36023_emb_X69356.1_HSRHFEF5 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36024_emb_X69349.1_HSRHFEF6 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36025_emb_X69350.1_HSRHFEF7 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36026_emb_X69351.1_HSRHFEF8 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36039_emb_X69354.1_HSRHT29 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36040_emb_X69352.1_HSRHT2918 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36041_emb_X69359.1_HSRHT294 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36066_emb_X69361.1_HSRLAN52 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36067_emb_X69353.1_HSRLAN56 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36068_emb_X69360.1_HSRLAN58 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36071_emb_X69364.1_HSRMRC58 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
gi_36072_emb_X69362.1_HSRMRC59 (344) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
28SrRNA (3178) CGGGGACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG
Consensus (3186) CGGGCACCCGGGGGGCCGGCGGCGGCGACTCTGGACGCGAGCCGGGCCCTTCCCGTGGATCG

FIG. 4G

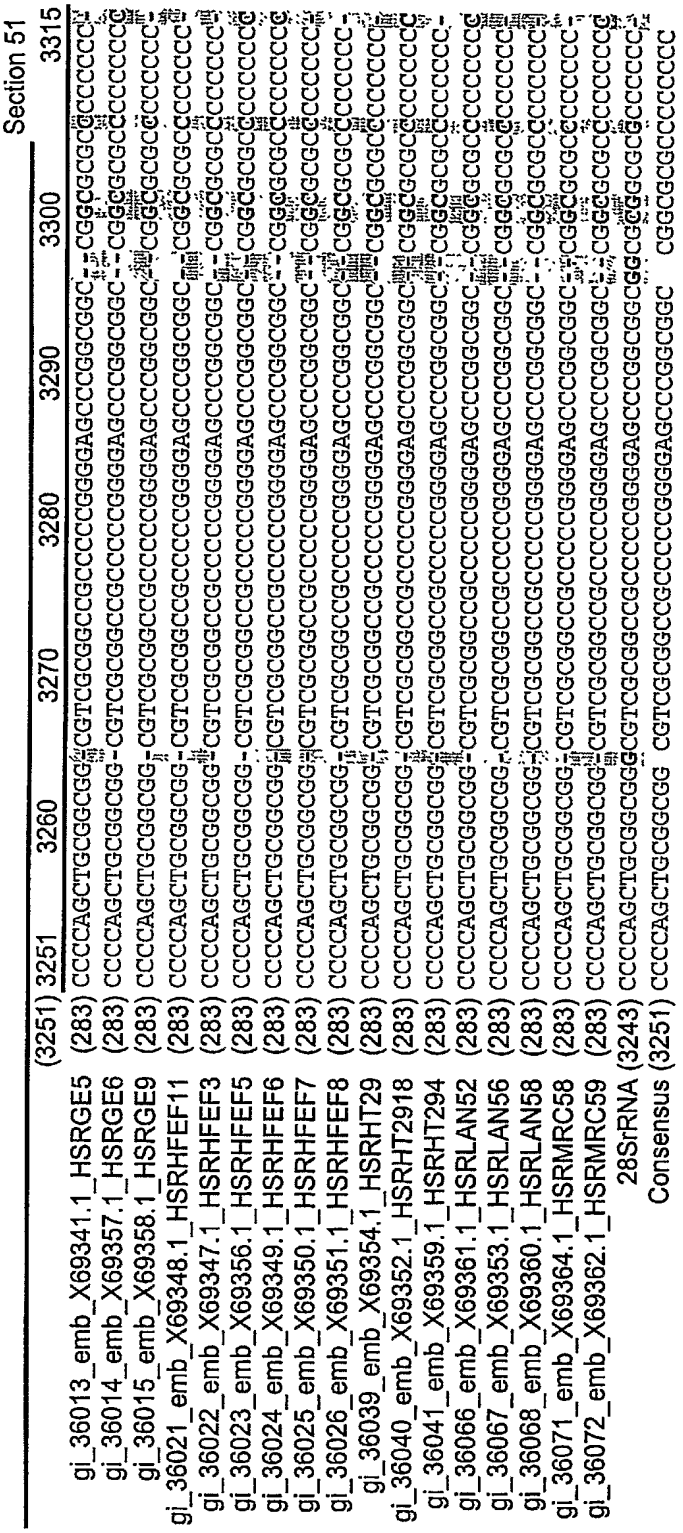

Section 51
(3251) 3251 3260 3270 3280 3290 3300 3315
gi_36013_emb_X69341.1_HSRGE5 (283) CCCCAGCTGCGGCGGCGTCGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCGCCCCCCC
gi_36014_emb_X69357.1_HSRGE6 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36015_emb_X69358.1_HSRGE9 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36021_emb_X69348.1_HSRHFEF11 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36022_emb_X69347.1_HSRHFEF3 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36023_emb_X69356.1_HSRHFEF5 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36024_emb_X69349.1_HSRHFEF6 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36025_emb_X69350.1_HSRHFEF7 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36026_emb_X69351.1_HSRHFEF8 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36039_emb_X69354.1_HSRHT29 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36040_emb_X69352.1_HSRHT2918 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36041_emb_X69359.1_HSRHT294 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36066_emb_X69361.1_HSRLAN52 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36067_emb_X69353.1_HSRLAN56 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36068_emb_X69360.1_HSRLAN58 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36071_emb_X69364.1_HSRMRC58 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
gi_36072_emb_X69362.1_HSRMRC59 (283) CCCCAGCTGCGGCGG-CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC
28SrRNA (3243) CCCCAGCTGCGGCGGCGTCGCGGCCGCCCCGGGAGCCCGGCGGCGGCGCGCCCCCCC
Consensus (3251) CCCCAGCTGCGGCGG CGTCGCGGCCGCCCCGGGAGCCCGGCGGC CGGCGCGCCCCCCCC

FIG. 4H

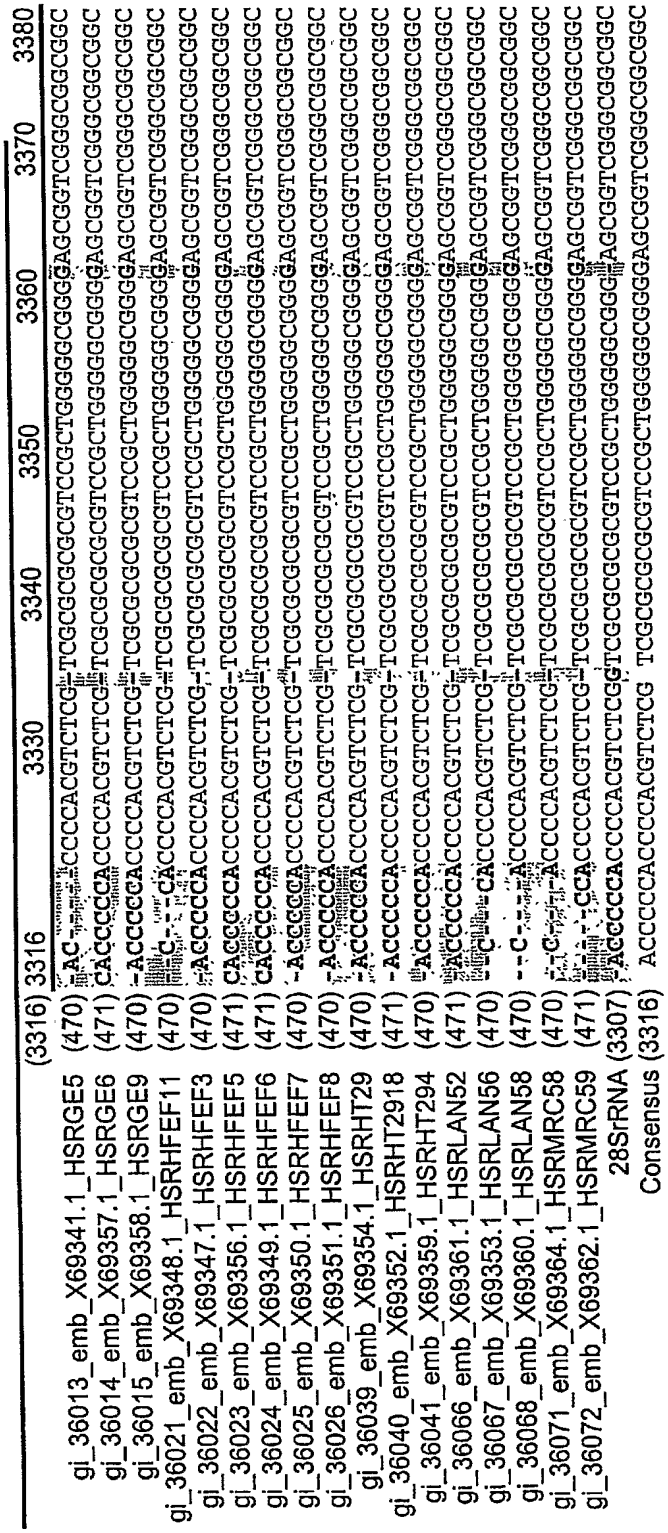

Section 52
(3316) 3316 3330 3340 3350 3360 3370 3380
gi_36013_emb_X69341.1_HSRGE5 (470)
gi_36014_emb_X69357.1_HSRGE6 (471)
gi_36015_emb_X69358.1_HSRGE9 (470)
gi_36021_emb_X69348.1_HSRHFEF11 (470)
gi_36022_emb_X69347.1_HSRHFEF3 (470)
gi_36023_emb_X69356.1_HSRHFEF5 (471)
gi_36024_emb_X69349.1_HSRHFEF6 (471)
gi_36025_emb_X69350.1_HSRHFEF7 (470)
gi_36026_emb_X69351.1_HSRHFEF8 (470)
gi_36039_emb_X69354.1_HSRHT29 (470)
gi_36040_emb_X69352.1_HSRHT2918 (471)
gi_36041_emb_X69359.1_HSRHT294 (470)
gi_36066_emb_X69361.1_HSRLAN52 (471)
gi_36067_emb_X69353.1_HSRLAN56 (470)
gi_36068_emb_X69360.1_HSRLAN58 (470)
gi_36071_emb_X69364.1_HSRMRC58 (470)
gi_36072_emb_X69362.1_HSRMRC59 (471)
28SrRNA (3307)
Consensus (3316)

FIG. 4I

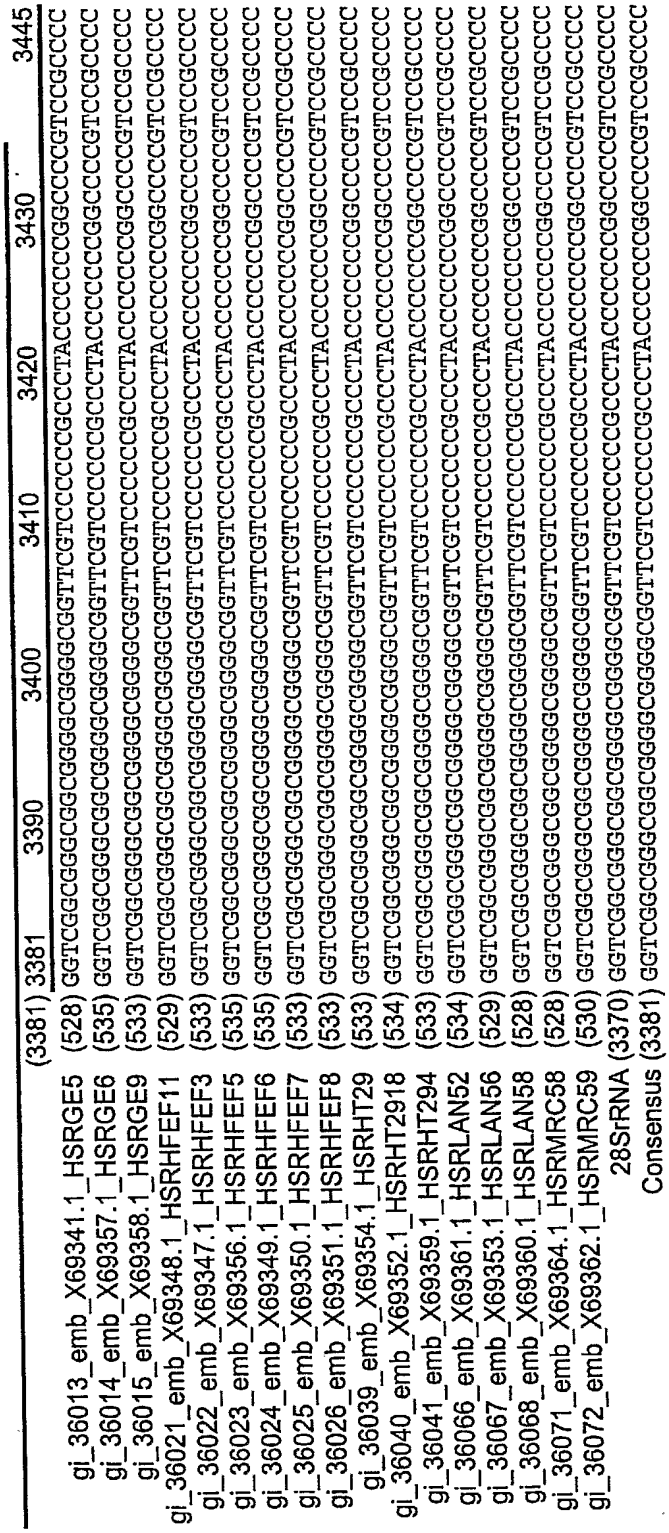

Section 53
3381
3390
3400
3410
3420
3430
3445
gi_36013_emb_X69341.1_HSRGE5 (528) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36014_emb_X69357.1_HSRGE6 (535) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36015_emb_X69358.1_HSRGE9 (533) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36021_emb_X69348.1_HSRHFEF11 (529) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36022_emb_X69347.1_HSRHFEF3 (533) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36023_emb_X69356.1_HSRHFEF5 (535) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36024_emb_X69349.1_HSRHFEF6 (535) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36025_emb_X69350.1_HSRHFEF7 (533) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36026_emb_X69351.1_HSRHFEF8 (533) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36039_emb_X69354.1_HSRHT29 (533) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36040_emb_X69352.1_HSRHT2918 (534) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36041_emb_X69359.1_HSRHT294 (533) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36066_emb_X69361.1_HSRLAN52 (534) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36067_emb_X69353.1_HSRLAN56 (529) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36068_emb_X69360.1_HSRLAN58 (528) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36071_emb_X69364.1_HSRMRC58 (528) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
gi_36072_emb_X69362.1_HSRMRC59 (530) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
28SrRNA (3370) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC
Consensus (3381) GGTCGGCGGGCGGGGCGGTTCGTCCCCCCGCCCTACCCCCCCGGCCCCGTCCGCCCC

FIG. 4J

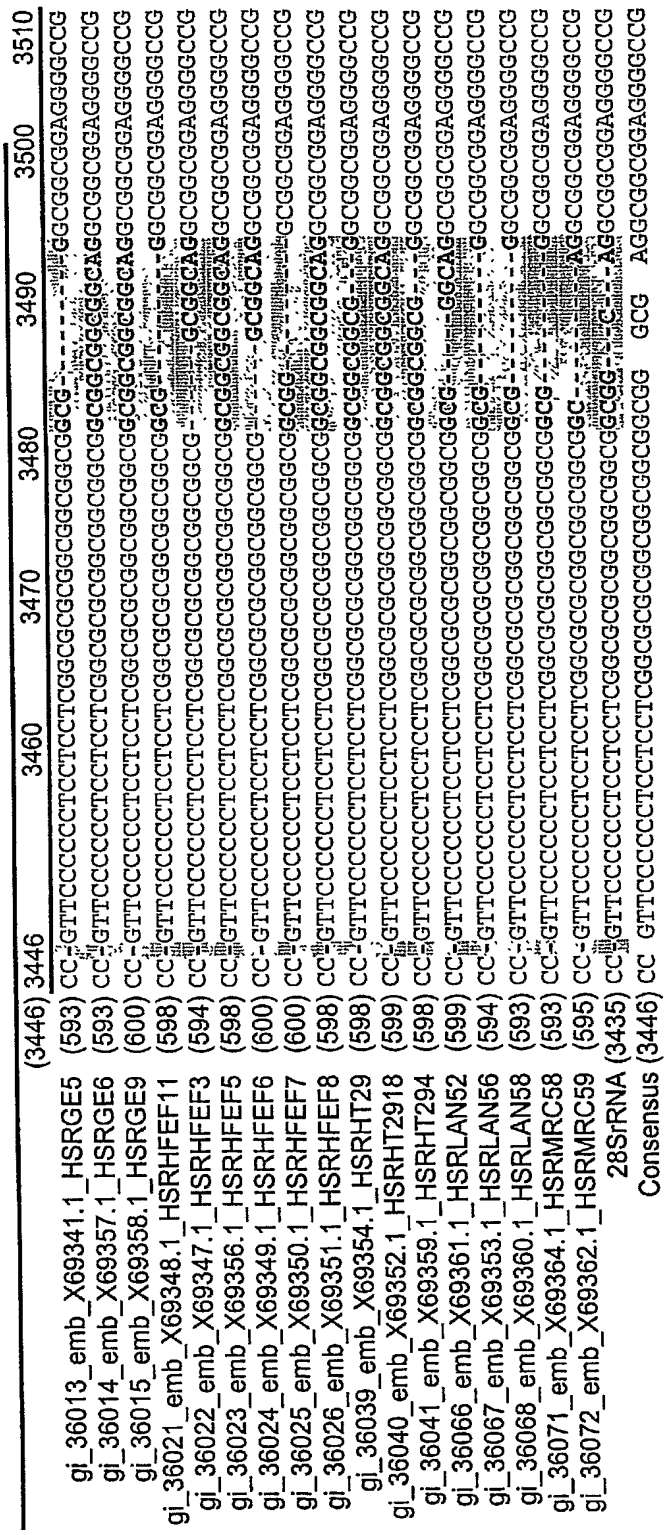

Section 54
(3446) 3446 3460 3470 3480 3490 3500 3510
gi_36013_emb_X69341.1_HSRGE5 (593)
gi_36014_emb_X69357.1_HSRGE6 (593)
gi_36015_emb_X69358.1_HSRGE9 (600)
gi_36021_emb_X69348.1_HSRHFEF11 (598)
gi_36022_emb_X69347.1_HSRHFEF3 (594)
gi_36023_emb_X69356.1_HSRHFEF5 (598)
gi_36024_emb_X69349.1_HSRHFEF6 (600)
gi_36025_emb_X69350.1_HSRHFEF7 (600)
gi_36026_emb_X69351.1_HSRHFEF8 (598)
gi_36039_emb_X69354.1_HSRHT29 (598)
gi_36040_emb_X69352.1_HSRHT2918 (599)
gi_36041_emb_X69359.1_HSRHT294 (598)
gi_36066_emb_X69361.1_HSRLAN52 (599)
gi_36067_emb_X69353.1_HSRLAN56 (594)
gi_36068_emb_X69360.1_HSRLAN58 (593)
gi_36071_emb_X69364.1_HSRMRC58 (593)
gi_36072_emb_X69362.1_HSRMRC59 (595)
28SrRNA (3435)
Consensus (3446)

FIG. 4K

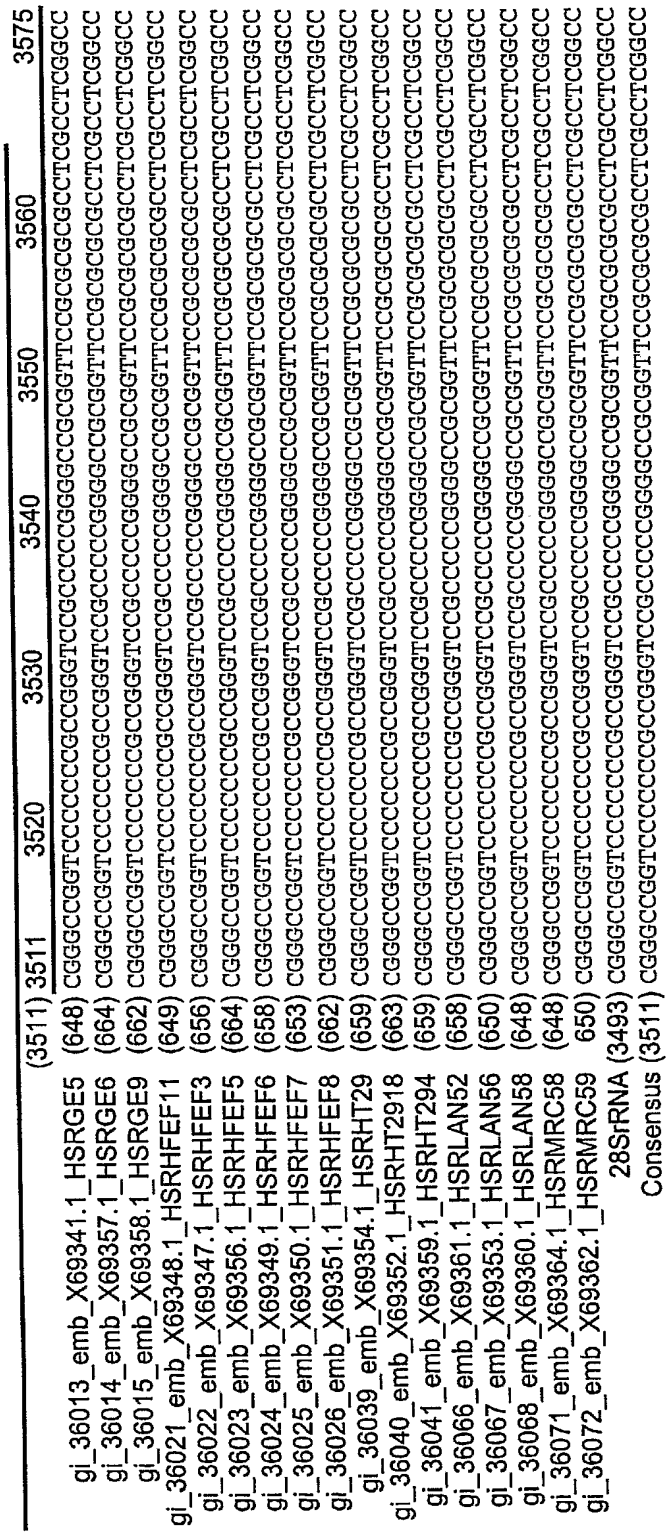

Section 55
(3511) 3511 3520 3530 3540 3550 3560 3575
gi_36013_emb_X69341.1_HSRGE5 (648) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCGCCTCGCCTCGGCC
gi_36014_emb_X69357.1_HSRGE6 (664) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36015_emb_X69358.1_HSRGE9 (662) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36021_emb_X69348.1_HSRHFEF11 (649) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36022_emb_X69347.1_HSRHFEF3 (656) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36023_emb_X69356.1_HSRHFEF5 (664) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36024_emb_X69349.1_HSRHFEF6 (658) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36025_emb_X69350.1_HSRHFEF7 (653) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36026_emb_X69351.1_HSRHFEF8 (662) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36039_emb_X69354.1_HSRHT29 (659) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36040_emb_X69352.1_HSRHT2918 (663) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36041_emb_X69359.1_HSRHT294 (659) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36066_emb_X69361.1_HSRLAN52 (658) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36067_emb_X69353.1_HSRLAN56 (650) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36068_emb_X69360.1_HSRLAN58 (648) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36071_emb_X69364.1_HSRMRC58 (648) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
gi_36072_emb_X69362.1_HSRMRC59 650) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
28SrRNA (3493) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC
Consensus (3511) CGGGCCGGTCCCCCCCGCCGGTCCGCCCCCGGGGCCGCGGTTCCGCGCGCCTCGCCTCGGCC

FIG. 4L

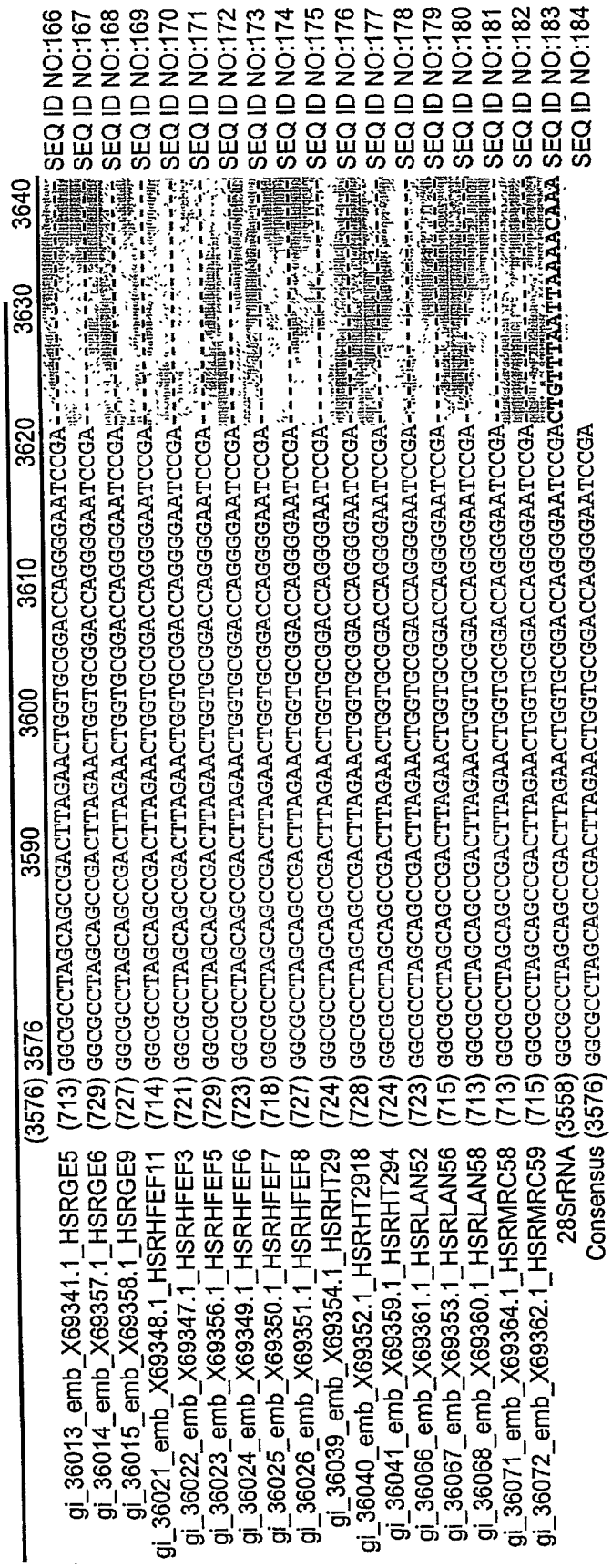
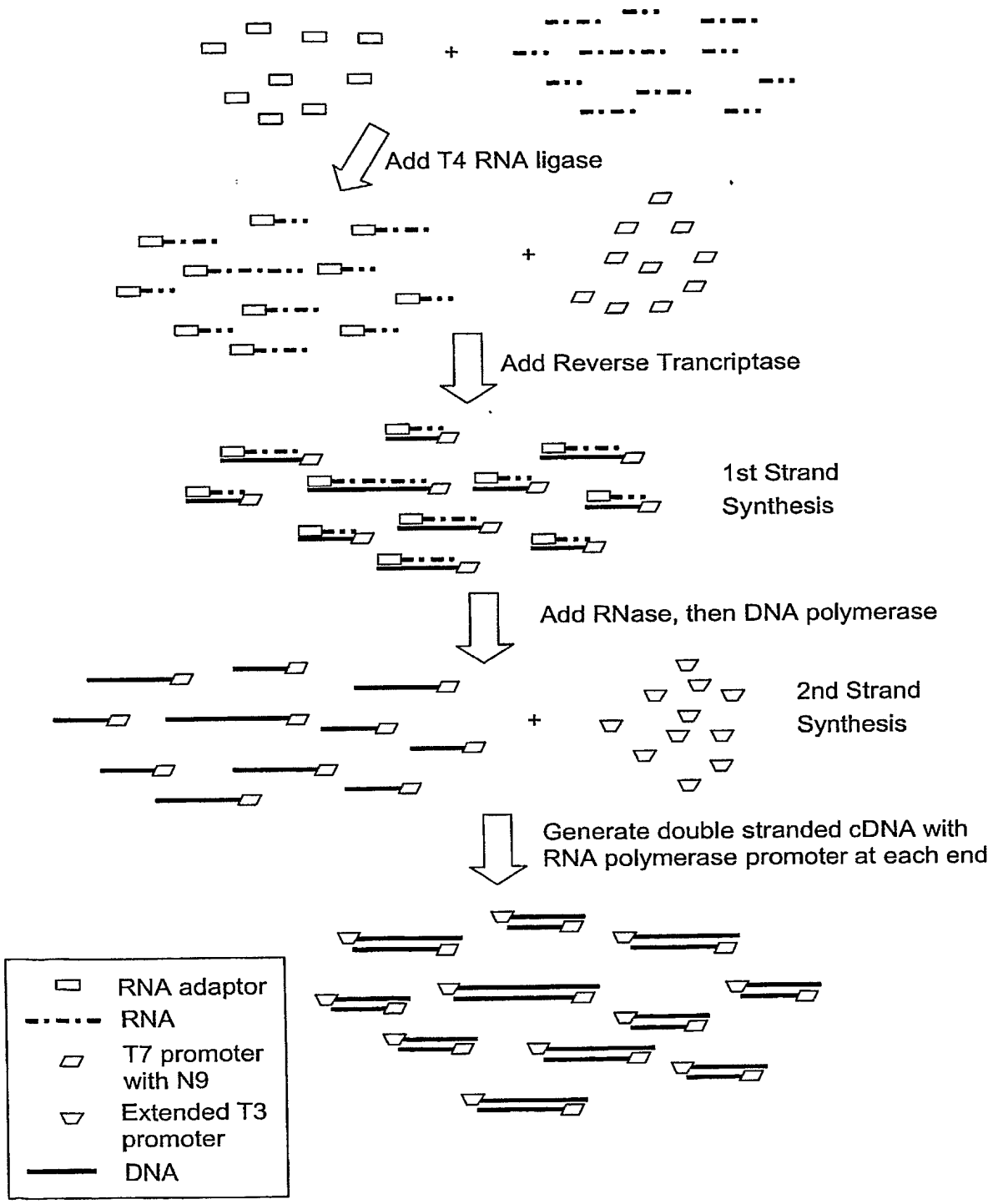

Section 56
3576
3590
3600
3610
3620
3630
3640
(3576)
gi_36013_emb_X69341.1_HSRGE5 (713) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:166
gi_36014_emb_X69357.1_HSRGE6 (729) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:167
gi_36015_emb_X69358.1_HSRGE9 (727) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:168
gi_36021_emb_X69348.1_HSRHFEF11 (714) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:169
gi_36022_emb_X69347.1_HSRHFEF3 (721) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:170
gi_36023_emb_X69356.1_HSRHFEF5 (729) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:171
gi_36024_emb_X69349.1_HSRHFEF6 (723) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:172
gi_36025_emb_X69350.1_HSRHFEF7 (718) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:173
gi_36026_emb_X69351.1_HSRHFEF8 (727) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:174
gi_36039_emb_X69354.1_HSRHT29 (724) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:175
gi_36040_emb_X69352.1_HSRHT2918 (728) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:176
gi_36041_emb_X69359.1_HSRHT294 (724) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:177
gi_36066_emb_X69361.1_HSRLAN52 (723) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:178
gi_36067_emb_X69353.1_HSRLAN56 (715) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:179
gi_36068_emb_X69360.1_HSRLAN58 (713) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:180
gi_36071_emb_X69364.1_HSRMRC58 (713) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:181
gi_36072_emb_X69362.1_HSRMRC59 (715) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:182
28SrRNA (3558) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGACTGTTTAATTAAAACAAA SEQ ID NO:183
Consensus (3576) GGCGCCTAGCAGCCGACTTAGAACTGGTGCGGACCAGGGGAATCCGA SEQ ID NO:184

FIG. 4M

FIG. 5



IDT Exhibit 1020.4200

SEQUENCE LISTING

<110> FREIJE, WADIHA
BRIKUN, IGOR

<120> SUBTRACTIVE SEPARATION AND AMPLIFICATION OF NON-RIBOSOMAL TRANSCRIBED RNA (nrRNA)

<130> 701046

<150> US 60/705,964
<151> 2005-08-05

<160> 248

<170> PatentIn version 3.3

<210> 1
<211> 5025
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 1
cgcgacctca gatcagacgt ggcgacccgc tgaatttaag catattagtc agcggaggaa 60

aagaaactaa ccaggattcc ctcagtaacg gcgagtgaac agggaagagc ccagcgccga 120

atccccgccc cgcggggcgc gggacatgtg gcgtacggaa gacccgctcc ccggcgccgc 180

tcgtgggggg cccaagtcct tctgatcgag gcccagcccg tggacggtgt gaggccggta 240

gcggccggcg cgcgcccggg tcttcccgga gtcgggttgc ttgggaatgc agcccaaagc 300

gggtggtaaa ctccatctaa ggctaaatac cggcacgaga ccgatagtca acaagtaccg 360

taagggaaag ttgaaaagaa ctttgaagag agagttcaag agggcgtgaa accgttaaga 420

ggtaaacggg tggggtccgc gcagtccgcc cggaggattc aacccggcgg cgggtccggc 480

cgtgtcggcg gcccggcgga tctttcccgc cccccgttcc tcccgacccc tccacccgcc 540

ctcccttccc ccgccgcccc tcctcctcct ccccggaggg ggcgggctcc ggcgggtgcg 600

ggggtgggcg ggcggggccg ggggtggggt cggcggggga ccgtcccccg gaccggcgac 660

cggccgccgc cgggcgcatt tccaggcggt gcgccgcgac cggctccggg acggctggga 720

aggcccggcg gggaaggtgg ctcggggggc cccgtccgtc cgtccgtcct cctcctcccc 780

cgtctccgcc ccccggcccc gcgtcctccc tcgggagggc gcgcgggtcg gggcggcggc 840

ggcggcggcg gtggcggcgg cggcgggggc ggcgggaccg aaaccccccc cgagtgttac 900

agcccccccg gcagcagcac tcgccgaatc ccggggccga gggagcgaga cccgtcgccg 960

cgctctcccc cctcccggcg cccacccccg cgggaatccc cgcgaggggg gtctcccccg 1020

gcgcggcgcc ggcgtctcct cgtgggggcg ccgggccacc cctcccacgg cgcgaccgct 1080

ctcccacccc tcctccccgc gcccccgccc cggcgacggg gggggtgccg cgcgcgggtc 1140

IDT Exhibit 1020.4201

gggggggcggg gcggactgtc cccagtgcgc cccgggcggg tcgcgccgtc gggcccgggg 1200
gaggttctct cggggccacg cgcgcgtccc ccgaagaggg ggacggcgga gcgagcgcac 1260
ggggtcggcg gcgacgtcgg ctacccaccc gacccgtctt gaaacacgga ccaaggagtc 1320
taacacgtgc gcgagtcggg ggctcgcacg aaagccgccg tggcgcaatg aaggtgaagg 1380
ccggcgcgct cgccggccga ggtgggatcc cgaggcctct ccagtccgcc gaggggcacc 1440
accggcccgt ctcgcccgcc gcgccgggga ggtggagcac gagcgcacgt gttaggaccc 1500
gaaagatggt gaactatgcc tgggcagggc gaagccagag gaaactctgg tggaggtccg 1560
tagcggtcct gacgtgcaaa tcggtcgtcc gacctgggta taggggcgaa agactaatcg 1620
aaccatctag tagctggttc cctccgaagt ttccctcagg atagctggcg ctctcgcaga 1680
cccgacgcac ccccgccacg cagttttatc cggtaaagcg aatgattaga ggtcttgggg 1740
ccgaaacgat ctcaacctat tctcaaactt taaatgggta agaagcccgg ctcgctggcg 1800
tggagccggg gtggaatgcg agtgcctagt gggccacttt tggtaagcag aactggcgct 1860
gcgggatgaa ccgaacgccg ggttaaggcg cccgatgccg acgctcatca gaccccagaa 1920
aaggtgttgg ttgatataga cagcaggacg gtggccatgg aagtcggaat ccgctaagga 1980
gtgtgtaaca actcacctgc cgaatcaact agccctgaaa atggatggcg ctggagcgtc 2040
gggcccatac ccggccgtcg ccggcagtcg agagtggacg ggagcggcgg gggcggcggc 2100
gcgcgcgcgc gtgtggtgtg cgtcggaggg cggcggcggc ggcggcggcg ggggtgtggg 2160
gtccttcccc cgccccccgc cccacgcctc ctcccctcct cccgcccacg ccccgctccc 2220
cgcccccgga gccccgcgga gctacgccgc gacgagtagg agggccgctg cggtgagcct 2280
tgaagcctag ggcgcgggcc cgggtggagg ccgccgcagg tgcagatctt ggtggtagta 2340
gcaaatattc aaacgagaac tttgaaggcc gaagtggaga agggttccat gtgaacagca 2400
gttgaacatg ggtcagtcgg tcctgagaga tgggcgagcg ccgttccgaa gggacgggcg 2460
atggcctccg ttgccctcgg ccgatcgaaa gggagtcggg ttcagatccc cgaatccgga 2520
gtggcggaga tgggcgccgc gaggcgtcca gtgcggtaac gcgaccgatc ccggagaagc 2580
cggcgggagc cccggggaga gttctctttt ctttgtgaag ggcagggcgc cctggaatgg 2640
gttcgccccg agagaggggc ccgtgccttg gaaagcgtcg cggttccggc ggcgtccggt 2700
gagctctcgc tggcccttga aaatccgggg gagagggtgt aaatctcgcg ccgggccgta 2760
cccatatccg cagcaggtct ccaaggtgaa cagcctctgg catgttggaa caatgtaggt 2820
aagggaagtc ggcaagccgg atccgtaact tcgggataag gattggctct aagggctggg 2880
tcggtcgggc tggggcgcga agcggggctg ggcgcgcgcc gcggctggac gaggcgcgcg 2940
ccccccccac gcccggggca ccccccctcgc ggccctcccc cgccccaccc gcgcgcgccg 3000

IDT Exhibit 1020.4202

ctcgctccct ccccaccccg cgccctctct ctctctctct cccccgctcc ccgtcctccc 3060
ccctccccgg gggagcgccg cgtgggggcg cggcgggggg agaagggtcg gggcggcagg 3120
ggccgcgcgg cggccgccgg ggcggccggc gggggcaggt ccccgcgagg ggggccccgg 3180
ggacccgggg ggccggcggc ggcgcggact ctggacgcga gccgggccct tcccgtggat 3240
cgccccagct gcggcgggcg tcgcggccgc ccccggggag cccggcggcg gcgcggcgcg 3300
ccccccaccc ccaccccacg tctcggtcgc gcgcgcgtcc gctgggggcg ggagcggtcg 3360
ggcggcggcg gtcggcgggc ggcggggcgg ggcggttcgt ccccccgccc tacccccccg 3420
gccccgtccg ccccccgttc cccctcctc ctcggcgcgc ggcggcggcg gcggcaggcg 3480
gcggaggggc cgcgggccgg tccccccccgc cgggtccgcc cccggggccg cggttccgcg 3540
cgcgcctcgc ctcggccggc gcctagcagc cgacttagaa ctggtgcgga ccaggggaat 3600
ccgactgttt aattaaaaca aagcatcgcg aaggcccgcg gcgggtgttg acgcgatgtg 3660
atttctgccc agtgctctga atgtcaaagt gaagaaattc aatgaagcgc gggtaaacgg 3720
cgggagtaac tatgactctc ttaaggtagc caaatgcctc gtcatctaat tagtgacgcg 3780
catgaatgga tgaacgagat tcccactgtc cctacctact atccagcgaa accacagcca 3840
agggaacggg cttggcggaa tcagcgggga aagaagaccc tgttgagctt gactctagtc 3900
tggcacggtg aagagacatg agaggtgtag aataagtggg aggcccccgg cgcccccccg 3960
gtgtccccgc gaggggcccg gggcggggtc cgcggccctg cgggccgccg gtgaaatacc 4020
actactctga tcgttttttc actgacccgg tgaggcgggg gggcgagccc gaggggctct 4080
cgcttctggc gccaagcgcc cgcccggccg ggcgcgaccc gctccgggga cagtgccagg 4140
tggggagttt gactggggcg gtacacctgt caaacggtaa cgcaggtgtc ctaaggcgag 4200
ctcagggagg acagaaacct cccgtggagc agaagggcaa aagctcgctt gatcttgatt 4260
ttcagtacga atacagaccg tgaaagcggg gcctcacgat ccttctgacc ttttgggttt 4320
taagcaggag gtgtcagaaa agttaccaca gggataactg gcttgtggcg gccaagcgtt 4380
catagcgacg tcgcttttg atccttcgat gtcggctctt cctatcattg tgaagcagaa 4440
ttcgccaagc gttggattgt tcacccacta atagggaacg tgagctgggt ttagaccgtc 4500
gtgagacagg ttagttttac cctactgatg atgtgttgtt gccatggtaa tcctgctcag 4560
tacgagagga accgcaggtt cagacatttg gtgtatgtgc ttggctgagg agccaatggg 4620
gcgaagctac catctgtggg attatgactg aacgcctcta agtcagaatc ccgcccaggc 4680
gaacgatacg gcagcgccgc ggagcctcgg ttggcctcgg atagccggtc ccccgcctgt 4740
ccccgccggc gggccgcccc cccctccacg cgccccgccg cgggagggcg cgtgccccgc 4800
cgcgcgccgg gaccggggtc cggtgcggag tgcccttcgt cctgggaaac ggggcgcggc 4860
cggaaaggcg gccgcccccc cgcccgtcac gcaccgcacg ttcgtgggga acctggcgct 4920

IDT Exhibit 1020.4203

aaaccattcg tagacgacct gcttctgggt cggggtttcg tacgtagcag agcagctccc 4980

tcgctgcgat ctattgaaag tcagccctcg acacaagggt ttgtc 5025

<210> 2
<211> 2235
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<220>
<221> misc_feature
<222> (61)..(64)
<223> "n" may be any nucleotide

<220>
<221> misc_feature
<222> (88)..(88)
<223> "n" may be any nucleotide

<220>
<221> misc_feature
<222> (97)..(101)
<223> "n" may be any nucleotide

<220>
<221> misc_feature
<222> (106)..(106)
<223> "n" may be any nucleotide

<400> 2
cgctgctcct cccgtcgccg tccgggcccg tccgtccgtc cgtccgtcgt cctcctcgct 60

nnnncggggc gccgggcccg tcctcacngg cccccgnnnn ngtccnggcc cgtcggggcc 120

tcgccgcgct ctaccttacc tacctggttg atcctgccag tagcatatgc ttgtctcaaa 180

gattaagcca tgcatgtcta agtacgcacg gccggtacag tgaaactgcg aatggctcat 240

taaatcagtt atggttcctt tggtcgctcg ctcctctcct acttggataa ctgtggtaat 300

tctagagcta atacatgccg acgggcgctg acccccttcg cgggggggat gcgtgcattt 360

atcagatcaa aaccaacccg gtcagcccct ctccggcccc ggccgggggg cgggcgccgg 420

cggctttggt gactctagat aacctcgggc cgatcgcacg ccccccgtgg cggcgacgac 480

ccattcgaac gtctgcccta tcaactttcg atggtagtcg ccgtgcctac catggtgacc 540

acgggtgacg gggaatcagg gttcgattcc ggagagggag cctgagaaac ggctaccaca 600

tccaaggaag gcagcaggcg cgcaaattac ccactcccga cccggggagg tagtgacgaa 660

aaataacaat acaggactct ttcgaggccc tgtaattgga atgagtccac tttaaatcct 720

ttaacgagga tccattggag ggcaagtctg gtgccagcag ccgcggtaat tccagctcca 780

atagcgtata ttaaagttgc tgcagttaaa aagctcgtag ttggatcttg ggagcgggcg 840

IDT Exhibit 1020.4204

ggcggtccgc cgcgaggcga gccaccgccc gtccccgccc cttgcctctc ggcgcccccc 900
cgatgctctt agctgagtgt cccgcggggc ccgaagcgtt tactttgaaa aaattagagt 960
gttcaaagca ggcccgagcc gcctggatac cgcagctagg aataatggaa taggaccgcg 1020
gttctatttt gttggttttc ggaactgagg ccatgattaa gagggacggc cgggggcatt 1080
cgtattgcgc cgctagaggt gaaattcctt ggaccggcgc aagacggacc agagcgaaag 1140
catttgccaa gaatgttttc attaatcaag aacgaaagtc ggaggttcga agacgatcag 1200
ataccgtcgt agttccgacc ataaacgatg ccgaccggcg atgcggcggc gttattccca 1260
tgacccgccg ggcagcttcc gggaaaccaa agtctttggg ttccgggggg agtatggttg 1320
caaagctgaa acttaaagga attgacggaa gggcaccacc aggagtggag cctgcggctt 1380
aatttgactc aacacgggaa acctcacccg gcccggacac ggacaggatt gacagattga 1440
tagctctttc tcgattccgt gggtggtggt gcatggccgt tcttagttgg tggagcgatt 1500
tgtctggtta attccgataa cgaacgagac tctggcatgc taactagtta cgcgaccccc 1560
gagcggtcgg cgtcccccaa cttcttagag ggacaagtgg cgttcagcca cccgagattg 1620
agcaataaca ggtctgtgat gcccttagat gtccggggct gcacgcgcgc tacactgact 1680
ggctcagcgt gtgcctaccc tacgccggca ggcgcgggta acccgttgaa ccccattcgt 1740
gatggggatc ggggattgca attattcccc atgaacgagg aattcccagt aagtgcgggt 1800
cataagcttg cgttgattaa gtccctgccc tttgtacaca ccgcccgtcg ctactaccga 1860
ttggatggtt tagtgaggcc ctcggatcgg ccccgccggg gtcggcccac ggccctggcg 1920
gagcgctgag aagacggtcg aacttgacta tctagaggaa gtaaaagtcg taacaaggtt 1980
tccgtaggtg aacctgcgga aggatcatta acggagcccg gacggcggcc cgcggcggcg 2040
ccgcgccgcg cttccctccg cacacccacc cccccaccgc gacggcgcgt gcgggcgggg 2100
ccgtgcccgt tcgttcgctc gctcgttcgt tcgccgcccg gcccggccgc gagagccgag 2160
aactcgggag ggagacgggg gagagagaga gagagagaga gagagagaga gagagagaga 2220
gaaagaaggg cgtgt 2235

<210> 3
<211> 156
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 3
cgactcttag cggtggatca ctcggctcgt gcgtcgatga agaacgcagc tagctgcgag 60
aattaatgtg aattgcagga cacattgatc atcgacactt cgaacgcact tgcggccccg 120

IDT Exhibit 1020.4205

ggttcctccc ggggctacgc ctgtctgagc gtcgct 156

<210> 4
<211> 121
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 4
gtctacggcc ataccaccct gaacgcgccc gatctcgtct gatctcggaa gctaagcagg 60

gtcgggcctg gttagtactt ggatgggaga ccgcctggga ataccgggtg ctgtaggctt 120

t 121

<210> 5
<211> 1558
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 5
gctaaaccta gccccaaacc cactccacct tactaccaga caaccttagc caaaccattt 60

acccaaataa agtataggcg atagaaattg aaacctggcg caatagatat agtaccgcaa 120

gggaaagatg aaaaattata accaagcata atatagcaag gactaacccc tataccttct 180

gcataatgaa ttaactagaa ataactttgc aaggagagcc aaagctaaga cccccgaaac 240

cagacgagct acctaagaac agctaaaaga gcacacccgt ctatgtagca aaatagtggg 300

aagatttata ggtagaggcg acaaacctac cgagcctggt gatagctggt tgtccaagat 360

agaatcttag ttcaacttta aatttgccca cagaaccctc taaatcccct tgtaaattta 420

actgttagtc caaagaggaa cagctctttg gacactagga aaaaaccttg tagagagagt 480

aaaaaattta acacccatag taggcctaaa agcagccacc aattaagaaa gcgttcaagc 540

tcaacaccca ctacctaaaa aatcccaaac atataactga actcctcaca cccaattgga 600

ccaatctatc accctataga agaactaatg ttagtataag taacatgaaa acattctcct 660

ccgcataagc ctgcgtcaga ttaaaacact gaactgacaa ttaacagccc aatatctaca 720

atcaaccaac aagtcattat taccctcact gtcaacccaa cacaggcatg ctcataagga 780

aaggttaaaa aaagtaaaag gaactcggca aatcttaccc cgcctgttta ccaaaaacat 840

cacctctagc atcaccagta ttagaggcac cgcctgccca gtgacacatg tttaacggcc 900

gcggtaccct aaccgtgcaa aggtagcata atcacttgtt ccttaaatag ggacctgtat 960

gaatggctcc acgagggttc agctgtctct tacttttaac cagtgaaatt gacctgcccg 1020

tgaagaggcg ggcataacac agcaagacga gaagacccta tggagcttta atttattaat 1080

IDT Exhibit 1020.4206

gcaaacagta cctaacaaac ccacaggtcc taaactacca aacctgcatt aaaaatttcg 1140

gttggggcga cctcggagca gaacccaacc tccgagcagt acatgctaag acttcaccag 1200

tcaaagcgaa ctactatact caattgatcc aataacttga ccaacggaac aagttaccct 1260

agggataaca gcgcaatcct attctagagt ccatatcaac aatagggttt acgacctcga 1320

tgttggatca ggacatcccg atggtgcagc cgctattaaa ggttcgtttg ttcaacgatt 1380

aaagtcctac gtgatctgag ttcagaccgg agtaatccag gtcggtttct atctaccttc 1440

aaattcctcc ctgtacgaaa ggacaagaga aataaggcct acttcacaaa gcgccttccc 1500

ccgtaaatga tatcatctca acttagtatt atacccacac ccacccaaga acagggtt 1558

<210> 6
<211> 953
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 6
aataggtttg gtcctagcct ttctattagc tcttagtaag attacacatg caagcatccc 60

cgttccagtg agttcaccct ctaaatcacc acgatcaaaa ggaacaagca tcaagcacgc 120

agcaatgcag ctcaaaacgc ttagcctagc cacaccccca cgggaaacag cagtgattaa 180

cctttagcaa taaacgaaag tttaactaag ctatactaac cccagggttg gtcaatttcg 240

tgccagccac cgcggtcaca cgattaaccc aagtcaatag aagccggcgt aaagagtgtt 300

ttagatcacc ccctccccaa taaagctaaa actcacctga gttgtaaaaa actccagttg 360

acacaaaata gactacgaaa gtggctttaa catatctgaa cacacaatag ctaagaccca 420

aactgggatt agataccca ctatgcttag ccctaaacct caacagttaa atcaacaaaa 480

ctgctcgcca gaacactacg agccacagct taaaactcaa aggacctggc ggtgcttcat 540

atccctctag aggagcctgt tctgtaatcg ataaaccccg atcaacctca ccacctcttg 600

ctcagcctat ataccgccat cttcagcaaa ccctgatgaa ggctacaaag taagcgcaag 660

tacccacgta aagacgttag gtcaaggtgt agcccatgag gtggcaagaa atgggctaca 720

ttttctaccc cagaaaacta cgatagccct tatgaaactt aagggtcgaa ggtggattta 780

gcagtaaact aagagtagag tgcttagttg aacagggccc tgaagcgcgt acacaccgcc 840

cgtcaccctc ctcaagtata cttcaaagga catttaacta aaaccctac gcatttatat 900

agaggagaca agtcgtaaca tggtaagtgt actggaaagt gcacttggac gaa 953

<210> 7
<211> 21
<212> DNA

IDT Exhibit 1020.4207

<213> Artificial

<220>
<223> Synthetic

<400> 7
acgacggacg gacggacgga c 21

<210> 8
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 8
taggtaaggt agagcgcggc ga 22

<210> 9
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 9
ttcactgtac cggccgtgcg ta 22

<210> 10
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 10
gcccgtcggc atgtattagc tc 22

<210> 11
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 11
tctagagtca ccaaagccgc cg 22

<210> 12
<211> 21
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4208

<223> Synthetic

<400> 12
tggtcaccat ggtaggcacg g 21

<210> 13
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 13
ggtcgggagt gggtaatttg cg 22

<210> 14
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 14
gccctccaat ggatcctcgt ta 22

<210> 15
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 15
ccgctcccaa gatccaacta cg 22

<210> 16
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 16
gcgggacact cagctaagag ca 22

<210> 17
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 17

IDT Exhibit 1020.4209

ttcctagctg cggtatccag gc 22

<210> 18
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 18
ccggtccaag gaatttcacc tc 22

<210> 19
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 19
tctgatcgtc ttcgaacctc cg 22

<210> 20
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 20
gactttggtt tcccggaagc tg 22

<210> 21
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 21
ttaagccgca ggctccactc ct 22

<210> 22
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 22
ctaagaacgg ccatgcacca cc 22

IDT Exhibit 1020.4210

<210> 23
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 23
tctaagaagt tgggggacgc cg 22

<210> 24
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 24
taggcacacg ctgagccagt ca 22

<210> 25
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 25
ctgggaattc ctcgttcatg gg 22

<210> 26
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 26
ccgatccgag ggcctcacta aa 22

<210> 27
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 27
gctccgttaa tgatccttcc gc 22

<210> 28
<211> 21
<212> DNA

IDT Exhibit 1020.4211

<213> Artificial

<220>
<223> Synthetic

<400> 28
cgaacgaacg agcgagcgaa c 21

<210> 29
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 29
gtcgccacgt ctgatctgag 20

<210> 30
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 30
ctcttccctg ttcactcgcc gt 22

<210> 31
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 31
cctcgatcag aaggacttgg gc 22

<210> 32
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 32
tttaccaccc gctttgggct g 21

<210> 33
<211> 22
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4212

<223> Synthetic

<400> 33
acggtttcac gccctcttga ac 22

<210> 34
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 34
cgccgggttg aatcctcc 18

<210> 35
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 35
cctccgggga ggaggaggag g 21

<210> 36
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 36
gcaccgcctg gaaatgcgcc 20

<210> 37
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 37
gattcggcga gtgctgct 18

<210> 38
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 38

IDT Exhibit 1020.4213

agagcggtcg cgccgtggga g 21

<210> 39
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 39
cgtggccccg agagaacctc 20

<210> 40
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 40
gtccgtgttt caagacgggt cg 22

<210> 41
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 41
cgccggcctt caccttcatt 20

<210> 42
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 42
tcgggtccta acacgtgcgc t 21

<210> 43
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 43
tttcgcccct atacccaggt cg 22

IDT Exhibit 1020.4214

<210> 44
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 44
tcggccccaa gacctctaat ca 22

<210> 45
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 45
actaggcact cgcattccac cc 22

<210> 46
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 46
gggtctgatg agcgtcggca t 21

<210> 47
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 47
gggctagttg attcggcagg tg 22

<210> 48
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 48
ccctccgacg cacaccacac 20

<210> 49
<211> 21
<212> DNA

IDT Exhibit 1020.4215

<213> Artificial

<220>
<223> Synthetic

<400> 49
aggcttcaag gctcaccgca g 21

<210> 50
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 50
cccttctcca cttcggcctt ca 22

<210> 51
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 51
tcgcccatct ctcaggaccg a 21

<210> 52
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 52
ttcggggatc tgaacccgac t 21

<210> 53
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 53
cgccctgccc ttcacaaaga aa 22

<210> 54
<211> 21
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4216

<223> Synthetic

<400> 54
ttttcaaggg ccagcgagag c 21

<210> 55
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 55
tgttccaaca tgccagaggc tg 22

<210> 56
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 56
cgacccagcc cttagagcca a 21

<210> 57
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 57
gtggggaggg agcgagcgg 19

<210> 58
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 58
ctgccgcccc gacccttctc 20

<210> 59
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 59

IDT Exhibit 1020.4217

cggctcgcgt ccagagtcc 19

<210> 60
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 60
ccgaggagga gggggggaac 19

<210> 61
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 61
cgcaccagtt ctaagtcggc tg 22

<210> 62
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 62
ttgacattca gagcactggg ca 22

<210> 63
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 63
cgttcatcca ttcatgcgcg tc 22

<210> 64
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 64
tgattccgcc aagcccgttc 20

IDT Exhibit 1020.4218

<210> 65
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 65
cgggggcctc ccacttattc ta 22

<210> 66
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 66
cgcctcaccg ggtcagtgaa a 21

<210> 67
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 67
tcaaactccc cacctggcac tg 22

<210> 68
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 68
ccttctgctc cacgggaggt tt 22

<210> 69
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 69
atcgtgaggc cccgctttca c 21

<210> 70
<211> 22
<212> DNA

IDT Exhibit 1020.4219

<213> Artificial

<220>
<223> Synthetic

<400> 70
cgacgtcgct atgaacgctt gg 22

<210> 71
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 71
gtgaacaatc caacgcttgg cg 22

<210> 72
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 72
tctgaacctg cggttcctct cg 22

<210> 73
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 73
gcctgggcgg gattctgact ta 22

<210> 74
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 74
gtttcccagg acgaagggca 20

<210> 75
<211> 22
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4220

<223> Synthetic

<400> 75
cgacccagaa gcaggtcgtc ta 22

<210> 76
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 76
caaacccttg tgtcgagggc tg 22

<210> 77
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 77
tcatcgacgc acgagccgag 20

<210> 78
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 78
gcaagtgcgt tcgaagtgtc g 21

<210> 79
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 79
agcgacgctc agacaggcgt a 21

<210> 80
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 80

IDT Exhibit 1020.4221

ctcttagcgg tggatcactc gc 22

<210> 81
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 81
gtcgatgatc aatgtgtcct gcaa 24

<210> 82
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 82
agcgacgctc agacaggcgt ag 22

<210> 83
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 83
actcactgga acggggatgc ttg 23

<210> 84
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 84
tgcgtgcttg atgcttgttc ct 22

<210> 85
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 85
ggcacgaaat tgaccaaccc tg 22

IDT Exhibit 1020.4222

<210> 86
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 86
acgccggctt ctattgactt gg 22

<210> 87
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 87
ggtcttagct attgtgtgtt cag 23

<210> 88
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 88
tgtggctcgt agtgttctgg cg 22

<210> 89
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 89
agagggatat gaagcaccgc ca 22

<210> 90
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 90
agaggtggtg aggttgatcg gg 22

<210> 91
<211> 22
<212> DNA

IDT Exhibit 1020.4223

<213> Artificial

<220>
<223> Synthetic

<400> 91
tcagggtttg ctgaagatgg cg 22

<210> 92
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 92
ttgccacctc atgggctaca cc 22

<210> 93
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 93
ggcggtgtgt acgcgcttca gg 22

<210> 94
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 94
ccatgttacg acttgtctcc tct 23

<210> 95
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 95
gtctggtagt aaggtggagt gg 22

<210> 96
<211> 22
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4224

<223> Synthetic

<400> 96
ttcatctttc ccttgcggta ct 22

<210> 97
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 97
ctgttcttag gtagctcgtc tgg 23

<210> 98
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 98
aaccagctat caccaggctc gg 22

<210> 99
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 99
cctagtgtcc aaagagctgt tcc 23

<210> 100
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 100
tgggtgttga gcttgaacgc tt 22

<210> 101
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 101

IDT Exhibit 1020.4225

tctgacgcag gcttatgcgg ag 22

<210> 102
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 102
tgagcatgcc tgtgttgggt tg 22

<210> 103
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 103
ccctcgtgga gccattcata ca 22

<210> 104
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 104
tcgtcttgct gtgttatgcc cg 22

<210> 105
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 105
actgctcgga ggttgggttc tg 22

<210> 106
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 106
agggtaactt gttccgttgg tc 22

IDT Exhibit 1020.4226

<210> 107
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 107
ttaatagcgg ctgcaccatc gg 22

<210> 108
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 108
ctccggtctg aactcagatc ac 22

<210> 109
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 109
ggggaaggcg ctttgtgaag ta 22

<210> 110
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 110
accctgttct tgggtgggtg tg 22

<210> 111
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 111
gacctcagat cagacgtggc ga 22

<210> 112
<211> 20
<212> DNA

IDT Exhibit 1020.4227

<213> Artificial

<220>
<223> Synthetic

<400> 112
gtcgggttgc ttgggaatgc 20

<210> 113
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 113
tccgcccgga ggattcaa 18

<210> 114
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 114
gcatttccag gcggtgcg 18

<210> 115
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 115
agcagcactc gccgaatc 18

<210> 116
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 116
acccctccca cggcgcga 18

<210> 117
<211> 21
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4228

<223> Synthetic

<400> 117
cacggaccaa ggagtctaac a 21

<210> 118
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 118
cgaaagatgg tgaactatgc ct 22

<210> 119
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 119
ccgaaacgat ctcaacctat tc 22

<210> 120
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 120
cacctgccga atcaactagc cc 22

<210> 121
<211> 19
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 121
ccacgcctcc tcccctcct 19

<210> 122
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 122

IDT Exhibit 1020.4229

gaacagcagt tgaacatggg t 21

<210> 123
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 123
cggggagagt tctcttttct tt 22

<210> 124
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 124
caatgtaggt aagggaagtc gg 22

<210> 125
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 125
ctctctctct cccccgctcc 20

<210> 126
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 126
tcccgtggat cgccccagct 20

<210> 127
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 127
ggcggggcgg ggcggttcgt 20

IDT Exhibit 1020.4230

<210> 128
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 128
cctagcagcc gacttagaac tg 22

<210> 129
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 129
tctcttaagg tagccaaatg cc 22

<210> 130
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 130
ctactatcca gcgaaaccac ag 22

<210> 131
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 131
cggggaaaga agaccctgtt ga 22

<210> 132
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 132
ctcagggagg acagaaacct c 21

<210> 133
<211> 22
<212> DNA

IDT Exhibit 1020.4231

<213> Artificial

<220>
<223> Synthetic

<400> 133
gatgtcggct cttcctatca tt 22

<210> 134
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 134
actaataggg aacgtgagct gg 22

<210> 135
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 135
ctaagtcaga atcccgccca gg 22

<210> 136
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 136
tcgccgcgct ctaccttacc ta 22

<210> 137
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 137
gagctaatac atgccgacgg gc 22

<210> 138
<211> 22
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4232

<223> Synthetic

<400> 138
aacgtctgcc ctatcaactt tc 22

<210> 139
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 139
ggccctgtaa ttggaatgag t 21

<210> 140
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 140
tgctcttagc tgagtgtccc gc 22

<210> 141
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 141
tagaggtgaa attccttgga cc 22

<210> 142
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 142
tgaaacttaa aggaattgac gga 23

<210> 143
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 143

IDT Exhibit 1020.4233

cacggacagg attgacagat t 21

<210> 144
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 144
cccaacttct tagagggaca agt 23

<210> 145
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 145
ctgtgatgcc cttagatgtc c 21

<210> 146
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 146
atgaacgagg aattcccagt aag 23

<210> 147
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 147
acctgcggaa ggatcattaa c 21

<210> 148
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 148
cccactccac cttactacca ga 22

IDT Exhibit 1020.4234

<210> 149
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 149
caaggagagc caaagctaag ac 22

<210> 150
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 150
agaaccctct aaatcccctt gt 22

<210> 151
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 151
caaagaggaa cagctctttg ga 22

<210> 152
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 152
taactgaact cctcacaccc aa 22

<210> 153
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 153
acaggcatgc tcataaggaa ag 22

<210> 154
<211> 22
<212> DNA

IDT Exhibit 1020.4235

<213> Artificial

<220>
<223> Synthetic

<400> 154
accgtgcaaa ggtagcataa tc 22

<210> 155
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 155
agcaagacga gaagacccta tg 22

<210> 156
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 156
ttcaccagtc aaagcgaact ac 22

<210> 157
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 157
gatccaataa cttgaccaac gg 22

<210> 158
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 158
aagattacac atgcaagcat cc 22

<210> 159
<211> 22
<212> DNA
<213> Artificial

<220>

IDT Exhibit 1020.4236

<223> Synthetic

<400> 159
gtcacacgat taacccaagt ca 22

<210> 160
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 160
gataccccac tatgcttagc cc 22

<210> 161
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 161
ttctgtaatc gataaacccc ga 22

<210> 162
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 162
gatcaacctc accacctctt g 21

<210> 163
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 163
ggctacattt tctaccccag aa 22

<210> 164
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 164

IDT Exhibit 1020.4237

gactcttagc ggtggatcac tc 22

<210> 165
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 165
gtctacggcc ataccaccct 20

<210> 166
<211> 759
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 166
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccа ccccccacgtc 480

tcgtcgcgcg cgcgtccgct gggggcgggg agcggtcggg cggcggcggt cggcgggcgg 540

cggggcgggg cggttcgtcc ccccgcccta ccccccccggc cccgtccgcc ccccccttccc 600

ccctcctcct cggcgcgcgg cggcggcggc gggcggcgga ggggccgcgg gccggtcccc 660

cccgccgggt ccgcccccgg ggccgcggtt ccgcgcgcgc ctcgcctcgg ccggcgccta 720

gcagccgact tagaactggt gcggaccagg ggaatccga 759

<210> 167
<211> 775
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 167
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

IDT Exhibit 1020.4238

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc cacccccacc 480

ccacgtctcg tcgcgcgcgc gtccgctggg ggcggggagc ggtcgggcgg cggcggtcgg 540

cgggcggcgg ggcggggcgg ttcgtccccc cgccctaccc ccccggcccc gtccgccccc 600

ccttcccccc tcctcctcgg cgcgcggcgg cggcggcggc ggcggcaggc ggcggagggg 660

ccgcgggccg gtcccccccg ccgggtccgc ccccggggcc gcggttccgc gcgcgcctcg 720

cctcggccgg cgcctagcag ccgacttaga actggtgcgg accaggggaa tccga 775

<210> 168
<211> 773
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 168
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccca cccccacccc 480

acgtctcgtc gcgcgcgcgt ccgctggggg cggggagcgg tcgggcggcg gcggtcggcg 540

ggcggcgggg cggggcggtt cgtccccccg ccctaccccc ccggccccgt ccgccccccc 600

ttccccccctc ctcctcggcg cgcggcggcg gcggcggcgg cggcaggcgg cggaggggcc 660

gcgggccggt ccccccccgcc gggtccgccc ccggggccgc ggttccgcgc gcgcctcgcc 720

tcggccggcg cctagcagcc gacttagaac tggtgcggac caggggaatc cga 773

<210> 169
<211> 760
<212> DNA

IDT Exhibit 1020.4239

<213> Artificial

<220>
<223> Synthetic

<400> 169
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctccccg 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc caccccacgt 480

ctcgtcgcgc gcgcgtccgc tgggggcggg gagcggtcgg gcggcggcgg tcggcgggcg 540

gcggggcggg gcggttcgtc cccccgccct acccccccgg ccccgtccgc ccccccttcc 600

cccctcctcc tcggcgcgcg gcggcggcgg cgggcggcgg aggggccgcg ggccggtccc 660

ccccgccggg tccgcccccg gggccgcggt tccgcgcgcg cctcgcctcg gccggcgcct 720

agcagccgac ttagaactgg tgcggaccag gggaatccga 760

<210> 170
<211> 767
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 170
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctccccа 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccа ccccccacccc 480

acgtctcgtc gcgcgcgcgt ccgctggggg cggggagcgg tcgggcggcg gcggtcggcg 540

ggcggcgggg cggggcggtt cgtcccccg ccctaccccc ccggccccgt ccgcccccccc 600

ttcccccctc ctcctcggcg cgcggcggcg gcggcggcag gcggcggagg ggccgcgggc 660

IDT Exhibit 1020.4240

cggtcccccc cgccgggtcc gcccccgggg ccgcggttcc gcgcgcgcct cgcctcggcc 720

ggcgcctagc agccgactta gaactggtgc ggaccagggg aatccga 767

<210> 171
<211> 775
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 171
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc caccccccacc 480

ccacgtctcg tcgcgcgcgc gtccgctggg ggcggggagc ggtcgggcgg cggcggtcgg 540

cgggcggcgg ggcggggcgg ttcgtccccc cgccctaccc ccccggcccc gtccgccccc 600

ccttcccccc tcctcctcgg cgcgcggcgg cggcggcggc ggcggcaggc ggcggagggg 660

ccgcgggccg gtcccccccg ccgggtccgc ccccggggcc gcggttccgc gcgcgcctcg 720

cctcggccgg cgcctagcag ccgacttaga actggtgcgg accaggggaa tccga 775

<210> 172
<211> 769
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 172
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

IDT Exhibit 1020.4241

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc caccccccacc 480

ccacgtctcg tcgcgcgcgc gtccgctggg ggcggggagc ggtcgggcgg cggcggtcgg 540

cgggcggcgg ggcggggcgg ttcgtccccc cgccctaccc ccccggcccc gtccgccccc 600

ccttcccccc tcctcctcgg cgcgcggcgg cggcggcggc aggcggcgga ggggccgcgg 660

gccggtcccc cccgccgggt ccgcccccgg ggccgcggtt ccgcgcgcgc ctcgcctcgg 720

ccggcgccta gcagccgact tagaactggt gcggaccagg ggaatccga 769

<210> 173
<211> 764
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 173
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctccccg 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccca cccccacccc 480

acgtctcgtc gcgcgcgcgt ccgctggggg cggggagcgg tcgggcggcg gcggtcggcg 540

ggcggcgggg cggggcggtt cgtccccccg ccctaccccc ccggccccgt ccgccccccc 600

ttccccccctc ctcctcggcg cgcggcggcg gcggcgggcg gcggaggggc cgcgggccgg 660

tccccccccgc cgggtccgcc cccggggccg cggttccgcg cgcgcctcgc ctcggccggc 720

gcctagcagc cgacttagaa ctggtgcgga ccaggggaat ccga 764

<210> 174
<211> 773
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 174
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

IDT Exhibit 1020.4242

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctccccg 180
ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240
cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300
cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360
ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420
cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccа cccccacccc 480
acgtctcgtc gcgcgcgcgt ccgctggggg cggggagcgg tcgggcggcg gcggtcggcg 540
ggcggcgggg cggggcggtt cgtccccccg ccctaccccc ccggccccgt ccgccccccc 600
ttccccccctc ctcctcggcg cgcggcggcg gcggcggcgg cggcaggcgg cggaggggcc 660
gcgggccggt cccccccgcc gggtccgccc ccggggccgc ggttccgcgc gcgcctcgcc 720
tcggccggcg cctagcagcc gacttagaac tggtgcggac caggggaatc cga 773

<210> 175
<211> 770
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 175
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60
gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120
ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180
ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240
cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300
cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360
ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420
cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccа cccccacccc 480
acgtctcgtc gcgcgcgcgt ccgctggggg cggggagcgg tcgggcggcg gcggtcggcg 540
ggcggcgggg cggggcggtt cgtccccccg ccctaccccc ccggccccgt ccgccccccc 600
ttccccccctc ctcctcggcg cgcggcggcg gcggcggcgg cgggcggcgg aggggccgcg 660
ggccggtccc ccccgccggg tccgcccccg gggccgcggt tccgcgcgcg cctcgcctcg 720
gccggcgcct agcagccgac ttagaactgg tgcggaccag gggaatccga 770

<210> 176
<211> 774
<212> DNA

IDT Exhibit 1020.4243

<213> Artificial

<220>
<223> Synthetic

<400> 176
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctccccg 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc acccccaccc 480

cacgtctcgt cgcgcgcgcg tccgctgggg gcggggagcg gtcgggcggc ggcggtcggc 540

gggcggcggg gcggggcggt tcgtcccccc gccctacccc cccggccccg tccgcccccc 600

cttcccccct cctcctcggc gcgcggcggc ggcggcggcg gcggcaggcg gcggaggggc 660

cgcgggccgg tccccccccgc cgggtccgcc cccggggccg cggttccgcg cgcgcctcgc 720

ctcggccggc gcctagcagc cgacttagaa ctggtgcgga ccaggggaat ccga 774

<210> 177
<211> 770
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 177
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctccccg 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccca cccccaccc 480

acgtctcgtc gcgcgcgcgt ccgctggggg cggggagcgg tcgggcggcg gcggtcggcg 540

ggcggcgggg cggggcggtt cgtcccccg ccctacccc ccggccccgt ccgcccccc 600

ttcccccctc ctcctcggcg cgcggcggcg gcggcggcgg cgggcggcgg aggggccgcg 660

IDT Exhibit 1020.4244

ggccggtccc ccccgccggg tccgcccccg gggccgcggt tccgcgcgcg cctcgcctcg 720

gccggcgcct agcagccgac ttagaactgg tgcggaccag gggaatccga 770

<210> 178
<211> 769
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 178
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct ccccccgccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc acccccaccc 480

cacgtctcgt cgcgcgcgcg tccgctgggg gcggggagcg gtcgggcggc ggcggtcggc 540

gggcggcggg gcggggcggt tcgtcccccc gccctacccc cccggccccg tccgcccccc 600

cttcccccct cctcctcggc gcgcggcggc ggcggcgggc aggcggcgga ggggccgcgg 660

gccggtcccc cccgccgggt ccgcccccgg ggccgcggtt ccgcgcgcgc ctcgcctcgg 720

ccggcgccta gcagccgact tagaactggt gcggaccagg ggaatccga 769

<210> 179
<211> 761
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 179
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct ccccccgccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360

IDT Exhibit 1020.4245

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc caccccacgt 480

ctcgtcgcgc gcgcgtccgc tgggggcggg gagcggtcgg gcggcggcgg tcggcgggcg 540

gcggggcggg gcggttcgtc cccccgccct acccccccgg ccccgtccgc ccccccccttc 600

ccccctcctc ctcggcgcgc ggcggcggcg gcgggcggcg gaggggccgc gggccggtcc 660

cccccgccgg gtccgccccc ggggccgcgg ttccgcgcgc gcctcgcctc ggccggcgcc 720

tagcagccga cttagaactg gtgcggacca ggggaatccg a 761

<210> 180
<211> 759
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 180
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240

cgccgcgtgg gggcggcggc ggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc accccacgtc 480

tcgtcgcgcg cgcgtccgct gggggcgggg agcggtcggg cggcggcggt cggcgggcgg 540

cggggcgggg cggttcgtcc ccccgcccta ccccccccggc cccgtccgcc ccccccttccc 600

ccctcctcct cggcgcgcgg cggcggcggc gggcggcgga ggggccgcgg gccggtcccc 660

cccgccgggt ccgcccccgg ggccgcggtt ccgcgcgcgc ctcgcctcgg ccggcgccta 720

gcagccgact tagaactggt gcggaccagg ggaatccga 759

<210> 181
<211> 759
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 181
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

IDT Exhibit 1020.4246

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180
ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240
cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300
cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360
ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420
cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc accccacgtc 480
tcgtcgcgcg cgcgtccgct gggggcgggg agcggtcggg cggcggcggt cggcgggcgg 540
cggggcgggg cggttcgtcc ccccgcccta ccccccccggc cccgtccgcc ccccccttccc 600
ccctcctcct cggcgcgcgg cggcggcggc gggcggcgga ggggccgcgg gccggtcccc 660
cccgccgggt ccgcccccgg ggccgcggtt ccgcgcgcgc ctcgcctcgg ccggcgccta 720
gcagccgact tagaactggt gcggaccagg ggaatccga 759

<210> 182
<211> 761
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 182
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60
gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120
ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180
ccccgcgccc tctctctctc tctctccccc gctccccgtc ctccccccctc cccgggggag 240
cgccgcgtgg gggcggcggc gggggggagaa gggtcggggc gccaggggcc ggcggcggcc 300
cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggggcc 360
ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420
cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gccccccccc ccaccccacg 480
tctcgtcgcg cgcgcgtccg ctgggggcgg ggagcggtcg ggcggcggcg gtcggcgggc 540
ggcggggcgg ggcggttcgt ccccccgccc tacccccccg gccccgtccg ccccccccttc 600
ccccctcctc ctcggcgcgc ggcggcggcg gcaggcggcg gaggggccgc gggccggtcc 660
cccccgccgg gtccgccccc ggggccgcgg ttccgcgcgc gcctcgcctc ggccggcgcc 720
tagcagccga cttagaactg gtgcggacca ggggaatccg a 761

<210> 183
<211> 827
<212> DNA

IDT Exhibit 1020.4247

<213> Artificial

<220>
<223> Synthetic

<400> 183
tctggcatgt tggaacaatg taggtaaggg aagtcggcaa gccggatccg taacttcggg 60

ataaggattg gctctaaggg ctgggtcggt cgggctgggg cgcgaagcgg ggctgggcgc 120

gcgccgcggc tggacgaggc gcgcgccccc cccacgcccg gggcaccccc ctcgcggccc 180

tcccccgccc cacccgcgcg cgccgctcgc tccctcccca ccccgcgccc tctctctctc 240

tctctccccc gctccccgtc ctcccccctc cccgggggag cgccgcgtgg gggcgcggcg 300

gggggagaag ggtcggggcg ccaggggccg cgcggcggcc gccggggcgg ccggcggggg 360

caggtccccg cgaggggggc cccgggcacc cgggggccg gcggcggcgc ggactctgga 420

cgcgagccgg gcccttcccg tggatcgccc cagctgcggc gggcgtcgcg gccgcccccg 480

gggagcccgg cggcggcgcg gcgcgccccc caccccccacc ccacgtctcg gtcgcgcgcg 540

cgtccgctgg gggcgggagc ggtcgggcgg cggcggtcgg cgggcggcgg ggcggggcgg 600

ttcgtccccc cgccctaccc ccccggcccc gtccgccccc ccttcccccc tcctcctcgg 660

cgcgcggcgg cggcggcggc aggcggcgga ggggccgcgg gccggtcccc cccgccgggt 720

ccgcccccgg ggccgcggtt ccgcgcgcgc ctcgcctcgg ccggcgccta gcagccgact 780

tagaactggt gcggaccagg ggaatccgac tgtttaatta aaacaaa 827

<210> 184
<211> 771
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 184
ggatccgtaa cttcgggata aggattggct ctaagggctg ggtcggtcgg gctggggcgc 60

gaagcggggc tgggcgcgcg ccgcggctgg acgaggcgcc gccgcccccc ccacgcccgg 120

ggcacccccc tcgcggccct cccccgcccc accccacgcg cgccgctcgc tccctcccca 180

ccccgcgccc tctctctctc tctctccccc gctccccgtc ctcccccctc cccgggggag 240

cgccgcgtgg gggcggcggc gggggagaa gggtcggggc gccaggggcc ggcggcggcc 300

cgccgggggc cccggcggcg gggcacggtc ccccggaggg gcccgggcac ccgggggccc 360

ggcggcggcg gcgactctgg acgcgagccg ggcccttccc gtggatcgcc ccagctgcgg 420

cggcgtcgcg gccgcccccg gggagcccgg cggccggcgc gcccccccccc acccccccacc 480

ccacgtctcg tcgcgcgcgc gtccgctggg ggcggggagc ggtcgggcgg cggcggtcgg 540

cgggcggcgg ggcggggcgg ttcgtccccc cgccctaccc ccccggcccc gtccgccccc 600

IDT Exhibit 1020.4248

ccttcccccc tcctcctcgg cgcgcggcgg cggcggcggg cgaggcggcg gaggggccgc 660

gggccggtcc cccccgccgg gtccgccccc ggggccgcgg ttccgcgcgc gcctcgcctc 720

ggccggcgcc tagcagccga cttagaactg gtgcggacca ggggaatccg a 771

<210> 185
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 185
cttgccgtcc caagcaatgg atg 23

<210> 186
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 186
caataggtgt gcgtcagaag cacc 24

<210> 187
<211> 24
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 187
ggtctcacag tgttgcccag gctg 24

<210> 188
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 188
ggaugaacgu aggaagcuug 20

<210> 189
<211> 26
<212> DNA
<213> Artificial

<220>
<223> Synthetic

IDT Exhibit 1020.4249

<220>
<221> misc_feature
<222> (26)..(26)
<223> "n" may be between 6 and 12 of any nucleotide

<400> 189
ggaattaata cgactcacta tagggn 26

<210> 190
<211> 44
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 190
gcgcgaaatt aaccctcact aaagggatga acgtaggaag cttg 44

<210> 191
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 191
agcaacccga ctccgggaag a 21

<210> 192
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 192
gacccgccgc cgggttgaat 20

<210> 193
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 193
cgcaccgcct ggaaatgc 18

<210> 194
<211> 18
<212> DNA
<213> Artificial

IDT Exhibit 1020.4250

<220>
<223> Synthetic

<400> 194
gattcggcga gtgctgct 18

<210> 195
<211> 18
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 195
tggccccgag agaacctc 18

<210> 196
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 196
cgtgtttcaa gacgggtcgg 20

<210> 197
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 197
ccaggcatag ttcaccatct tt 22

<210> 198
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 198
gccccaagac ctctaatcat tc 22

<210> 199
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

IDT Exhibit 1020.4251

<400> 199
tgttacacac tccttagcgg att 23

<210> 200
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 200
aggcttcaag gctcaccgca 20

<210> 201
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 201
caggaccgac tgacccatgt tc 22

<210> 202
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 202
ctgcccttca caaagaaaag ag 22

<210> 203
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 203
gccgacttcc cttacctaca tt 22

<210> 204
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 204
gagagagaga gagagagggc g 21

IDT Exhibit 1020.4252

<210> 205
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 205
agctggggcg atccacggga 20

<210> 206
<211> 20
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 206
ccgaggagga ggggggaacg 20

<210> 207
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 207
cgcaccagtt ctaagtcggc tg 22

<210> 208
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 208
gacagtggga atctcgttca tc 22

<210> 209
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 209
gggcctccca cttattctac a 21

<210> 210

IDT Exhibit 1020.4253

<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 210
agagtagtgg tatttcaccg gc 22

<210> 211
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 211
ctcgccttag gacacctgcg t 21

<210> 212
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 212
aatgatagga agagccgaca tc 22

<210> 213
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 213
cgttcagtca taatcccaca ga 22

<210> 214
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 214
ggattctgac ttagaggcgt tc 22

<210> 215
<211> 21
<212> DNA
<213> Artificial

IDT Exhibit 1020.4254

<220>
<223> Synthetic

<400> 215
gtttagcgcc aggttcccca c 21

<210> 216
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 216
gcccgtcggc atgtattagc tc 22

<210> 217
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 217
gaatcgaacc ctgattcccc gt 22

<210> 218
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 218
tggactcatt ccaattacag gg 22

<210> 219
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 219
gggacactca gctaagagca tc 22

<210> 220
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

IDT Exhibit 1020.4255

<400> 220
gccggtccaa ggaatttcac c 21

<210> 221
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 221
gtttcagctt tgcaaccata ctc 23

<210> 222
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 222
agttagcatg ccagagtctc gt 22

<210> 223
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 223
catctaaggg catcacagac ct 22

<210> 224
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 224
agcttatgac ccgcacttac tg 22

<210> 225
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 225
ctcactaaac catccaatcg gt 22

IDT Exhibit 1020.4256

<210> 226
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 226
cctacggaaa ccttgttacg ac 22

<210> 227
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 227
acgcccttct ttctctctct ct 22

<210> 228
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 228
ggtcttagct ttggctctcc tt 22

<210> 229
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 229
agtgtccaaa gagctgttcc tc 22

<210> 230
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 230
attgggtgtg aggagttcag tt 22

<210> 231

IDT Exhibit 1020.4257

<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 231
cggaggagaa tgttttcatg tt 22

<210> 232
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 232
ctttccttat gagcatgcct gt 22

<210> 233
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 233
taagagacag ctgaaccctc gt 22

<210> 234
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 234
gcaggtttgg tagtttagga cc 22

<210> 235
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 235
ctagggtaac ttgttccgtt gg 22

<210> 236
<211> 22
<212> DNA
<213> Artificial

IDT Exhibit 1020.4258

<220>
<223> Synthetic

<400> 236
ggtagataga aaccgacctg ga 22

<210> 237
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 237
ctcttgtcct ttcgtacagg ga 22

<210> 238
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 238
tgacttgggt taatcgtgtg ac 22

<210> 239
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 239
gtagtgttct ggcgagcagt t 21

<210> 240
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 240
cttgcgctta ctttgtagcc tt 22

<210> 241
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

IDT Exhibit 1020.4259

<400> 241
atccaccttc gacccttaag tt 22

<210> 242
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 242
gccctgttca actaagcact ct 22

<210> 243
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 243
tgcactttcc agtacactta cca 23

<210> 244
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 244
ttcgaagtgt cgatgatcaa tgt 23

<210> 245
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 245
cggtctccca tccaagtact aa 22

<210> 246
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 246
ggagcttcat ctggacctgg gtc 23

IDT Exhibit 1020.4260

<210> 247
<211> 25
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 247
caaaacacca gtgcaggcca acttg 25

<210> 248
<211> 23
<212> DNA
<213> Artificial

<220>
<223> Synthetic

<400> 248
ttgtaatccc agcactctgg gag 23

IDT Exhibit 1020.4261

Electronic Acknowledgement Receipt

EFS ID:	46340587
Application Number:	17381888
International Application Number:	
Confirmation Number:	6188
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Kelley Warren
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/04US 34079/550
Receipt Date:	04-AUG-2022
Filing Date:	21-JUL-2021
Time Stamp:	21:48:45
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Foreign Reference	WO2007019444A2.pdf	5420718 aacf3171a5e0817117d4252ebc5f8f418362bb3f	no	121

Warnings:

Information:					
2	Foreign Reference	WO2007030759A2.pdf	5545072 07de9b157f4fcb317bd1fc7f74cc21ae0a6dce4b	no	95
Warnings:					
Information:					
3	Foreign Reference	WO2007037678.pdf	2207352 8d134902d9766f17c25715736f34d579fff22f9f	no	47
Warnings:					
Information:					
4	Foreign Reference	WO2007052006A1.pdf	3966832 e9aaf8f191c15d5806d76de3aad1b124eb87ab20	no	94
Warnings:					
Information:					
5	Foreign Reference	WO2007057652A1.pdf	2154914 6210df94e8c3c7fe77b09fe99397fcd372a98995	no	54
Warnings:					
Information:					
6	Foreign Reference	WO2007073165.pdf	3527095 7376b619e60236160e26101db75fce3ec5f2914f	no	57
Warnings:					
Information:					
7	Foreign Reference	WO2007136717A1.pdf	1916540 f56c247513c768973d34e55f677d1eb95a11129d	no	47
Warnings:					
Information:					
8	Foreign Reference	WO2008005459A2.pdf	5062602 04bc72ada8a0e435ff816e46e04257538ec97db6	no	95
Warnings:					
Information:					

9	Foreign Reference	WO2008015396A2.pdf	3548242 f1ef2ba78309a20c7cafea62e66dd4a6070e1ef5	no	81
Warnings:					
Information:					
10	Foreign Reference	WO2008033442A2.pdf	3614326 792fc35503aa8646d7ff21b48677e7e6ad069ed6	no	64
Warnings:					
Information:					
11	Foreign Reference	WO2008093098.pdf	4946576 25fb20d0035089d1cdb75fe09bd8b0bc70d53a56	no	111
Warnings:					
Information:					
12	Foreign Reference	WO2008115185A2.pdf	890258 d4781bc00fe816ac07dd765908ef9020f600cd47	no	21
Warnings:					
Information:					
13	Foreign Reference	34079_12US_-_WO2008150432.pdf	5241093 b677cbabd678730feea1a307154627dd261f25f6	no	93
Warnings:					
Information:					
14	Foreign Reference	WO2009053039A1.pdf	4396998 2fbe911c76ceccc94fb033759e6d044cf9bf781e	no	82
Warnings:					
Information:					
15	Foreign Reference	WO2009102878A2.pdf	12217128 0e4a02287416ba8fb30655de3163ca7a1dc20922	no	198
Warnings:					
Information:					

16	Foreign Reference	WO2009102896A2.pdf	5754041 f647569b5c8ed0cfcf71da85fe35f2df7456a136	no	92
Warnings:					
Information:					
17	Foreign Reference	WO2009112844A1.pdf	2638836 b64570d75c3266e7a57db99e4418eb8d5610c916	no	58
Warnings:					
Information:					
18	Foreign Reference	WO2009117698A2.pdf	4771142 765f82bf813df1f1878d1ccefe060510c78132b5	no	75
Warnings:					
Information:					
19	Foreign Reference	WO2009120372A2.pdf	3871825 72d3e687717801b3a2512d3608bccbc2885fa29f	no	75
Warnings:					
Information:					
20	Foreign Reference	WO2009120374A2.pdf	3542666 4852bd248d6af92e59b564222c9795870dac4419	no	69
Warnings:					
Information:					
21	Foreign Reference	WO2010003153A2.pdf	4503955 317d7352a5aa21164be5ce14496cfd7b449b78ec	no	117
Warnings:					
Information:					
22	Foreign Reference	WO2010030683A1.pdf	5949369 cb170ae386b452f195b3609242ff99918db05e05	no	131
Warnings:					
Information:					

23	Foreign Reference	WO2010039991A2.pdf	2983721 269dbf0eeae527ddb915196d41095192822f78db	no	58
Warnings:					
Information:					
24	Foreign Reference	WO2010063711A1.pdf	1705819 2e52f6643eab416286af0177caeb10d4d55a7b7e	no	37
Warnings:					
Information:					
25	Foreign Reference	WO2010064893A1.pdf	1441788 d859b3d76dfbe419162a5eeaa2f8a5f474572494	no	25
Warnings:					
Information:					
26	Foreign Reference	WO2010085715A1.pdf	1790389 21d0c55647deb77fa34a09aebb007d54b290402a	no	31
Warnings:					
Information:					
27	Foreign Reference	WO2010091246.pdf	3870711 61ffd620e6eb96ae64c6c106ef1752148efcd417	no	65
Warnings:					
Information:					
28	Foreign Reference	WO2010115154A1.pdf	8459556 06b64f200fd928c4f124042e17dc72db56e057c6	no	152
Warnings:					
Information:					
29	Foreign Reference	WO2010129937A2.pdf	7097949 2d1fdf7e40630e1dc2bb24d9d528d172e269baa6	no	155
Warnings:					
Information:					

30	Foreign Reference	WO2011003630A1.pdf	3994085 5149a3908ca6eaa7a23e37e66b6eca6957e2773e	no	84
Warnings:					
Information:					
31	Foreign Reference	WO2011009941A1.pdf	3591522 b5f60f13c549f2d2a1838daf493200f3f9db156a	no	66
Warnings:					
Information:					
32	Foreign Reference	WO2011019964A1.pdf	5173093 97c9dc61a06fbf4c03501acbf0f5c41b4e8f20c4	no	74
Warnings:					
Information:					
33	Foreign Reference	34079_135US_-_WO2011032053.pdf	2561860 ec3a42bfcafc61f7592c1e380345f6096d3952c5	no	48
Warnings:					
Information:					
34	Foreign Reference	WO2011053987A1.pdf	5966984 6501f728116b7d01b19151d8dad0e861f47db0f3	no	47
Warnings:					
Information:					
35	Foreign Reference	WO2011151777A1.pdf	1401642 8bd2ab2d9c414b717d7d252c3497999d733e31cd	no	28
Warnings:					
Information:					
36	Foreign Reference	WO2011156529A2.pdf	3778247 60e4fccf815a5bc3f28f5f6c075056df8b2eef92	no	52
Warnings:					
Information:					

37	Foreign Reference	WO2011032040.pdf	3585399 eb6df021cc34d57891583dd1128dc752a564145b	no	50
Warnings:					
Information:					
38	Foreign Reference	WO2012013932A1.pdf	1217956 a68f925f83172766a1de8b2e142c525d7bb919f2	no	34
Warnings:					
Information:					
39	Foreign Reference	WO2012054873.pdf	3565148 a37233ffa031499552c42a69ab378e53211e2633	no	79
Warnings:					
Information:					
40	Foreign Reference	WO2012061832A1.pdf	4268818 c200e7c5b1d189096e9557ec4e9e4d90986ab71f	no	107
Warnings:					
Information:					
41	Foreign Reference	WO2012103154A1.pdf	4232720 6983a57600ae719f7ba2f6ddab628f5d6d4b9375	no	74
Warnings:					
Information:					
42	Foreign Reference	WO2013059740A1.pdf	7762687 f7fc88dd1d6806594a4f62b72fe7fe9827908be9	no	158
Warnings:					
Information:					
43	Foreign Reference	WO2013059746A1.pdf	3138922 ad52aebe1bfdab1cf73f2c950c59f8a6ad8064e6	no	47
Warnings:					
Information:					

44	Foreign Reference	WO2013112923A1.pdf	3568013 d8829499be952ef8ea13dddf0b6c72895fe3cfa2	no	65
Warnings:					
Information:					
45	Foreign Reference	WO2013130512.pdf	5819514 dffe40d040afab1a697d7c35cf9b6b05d157e881	no	107
Warnings:					
Information:					
46	Foreign Reference	WO2013138510.pdf	5364974 175f9167f85388060c2ae995ffea86f20a9ce916	no	101
Warnings:					
Information:					
47	Foreign Reference	WO2013177220A1.pdf	5855787 72f049aa56106918523937b92ae2bbc5ece6f320	no	109
Warnings:					
Information:					
48	Foreign Reference	WO2013190441.pdf	5567579 a6eeefef4889cfe6ef37cd55939eee1bbf997693	no	114
Warnings:					
Information:					
49	Foreign Reference	WO2013191775A2.pdf	6434915 c259c1ed75328c02bf8d1c8bf394084fe7297ce3	no	106
Warnings:					
Information:					
50	Foreign Reference	34079_135US_-_WO2014039556A1.pdf	23155952 8b3857563d721dcfb368303bea80f85a0045103f	no	136
Warnings:					
Information:					

51	Foreign Reference	WO2014082032.pdf	4223908 312d96c4764dd8e61f89403e5390c151ef892420	no	91
Warnings:					
Information:					
52	Foreign Reference	WO2014144092A1.pdf	3968743 ab6af0f7117c9df4f94b14b7b8a4642896776daf	no	71
Warnings:					
Information:					
53	Foreign Reference	WO2014150931A1.pdf	6667644 64806a6c43bf1b84d3662ecafa45a936335c638e	no	106
Warnings:					
Information:					
54	Foreign Reference	WO2015031691.pdf	25414654 deac67a82847985ca75f98e24fff8d939f1fb809	no	379
Warnings:					
Information:					
55	Foreign Reference	WO2015073711.pdf	2413761 d6c5093b6bc5930a67dc260d68d0ab163af5bee0	no	52
Warnings:					
Information:					
56	Foreign Reference	WO2015131107A1.pdf	11451627 0fd0bde04651480fb8208c33698d5627f72e98a5	no	175
Warnings:					
Information:					
57	Foreign Reference	WO8909284A1.pdf	2718382 dc9fc2e2f1ae835e855df6a50e1481e7277eac1e	no	64
Warnings:					
Information:					

58	Foreign Reference	WO9207951A1.pdf	1943305 bda22e272f61d655dd18b89f19b06aa2ea6be141	no	62
Warnings:					
Information:					
59	Foreign Reference	WO9318052A1.pdf	7027932 d2529e755a39748d6e3159e0e688a01b192c126e	no	147
Warnings:					
Information:					
60	Foreign Reference	WO9416090A1.pdf	2653875 b7df10c863f3a5aab2ffdf1968dc91685e6220ec	no	59
Warnings:					
Information:					
			Total Files Size (in bytes):	301527161	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

<u>New Applications Under 35 U.S.C. 111</u>
If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.
<u>National Stage of an International Application under 35 U.S.C. 371</u>
If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.
<u>New International Application Filed with the USPTO as a Receiving Office</u>
If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Espacenet

Bibliographic data: WO2011151777 (A1) — 2011-12-08

METHOD FOR THE PREPARATION AND AMPLIFICATION OF REPRESENTATIVE AND STRAND- SPECIFIC LIBRARIES OF CDNA FOR HIGH THROUGHPUT SEQUENCING, USE THEREOF, KIT AND CARTRIDGES FOR AUTOMATION KIT

Inventor(s): TULLO APOLLONIA [IT]; MANGIULLI MARINA [IT]; SBISA ELISABETTA [IT]; PESOLE GRAZIANO [IT] ± (TULLO, APOLLONIA, ; MANGIULLI, MARINA, ; SBISA', ELISABETTA, ; PESOLE, GRAZIANO)

Applicant(s): CONSIGLIO NAZIONALE RICERCHE [IT]; TULLO APOLLONIA [IT]; MANGIULLI MARINA [IT]; SBISA ELISABETTA [IT]; PESOLE GRAZIANO [IT] ± (CONSIGLIO NAZIONALE DELLE RICERCHE, ; TULLO, APOLLONIA, ; MANGIULLI, MARINA, ; SBISA', ELISABETTA, ; PESOLE, GRAZIANO)

Classification: **- international:** ***C12N15/10; C12Q1/68; C40B40/08; C40B50/06***
- cooperative: **C12N15/1003**; **C12N15/1096**; **C40B40/08**; **C40B50/06**

Application number: WO2011IB52369 20110530 Global Dossier

Priority number(s): IT2010RM00293 20100531

Also published as: EP2576780 (A1) EP2576780 (B1) ITRM20100293 (A1)

Abstract of WO2011151777 (A1)

The present invention relates to a rapid, simple, effective and economical method for preparing and amplifying representative cDNA libraries for next-generation high throughput sequencing, kit and cartridges for automation kits which use such method manually or in an automatic manner.

IDT Exhibit 1020.4272

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
8 December 2011 (08.12.2011)

PCT

(10) International Publication Number
WO 2011/151777 A1

(51) International Patent Classification:
C12N 15/10 (2006.01) *C40B 40/08* (2006.01)
C12Q 1/68 (2006.01) *C40B 50/06* (2006.01)

(21) International Application Number:
PCT/IB2011/052369

(22) International Filing Date:
30 May 2011 (30.05.2011)

(25) Filing Language: Italian

(26) Publication Language: English

(30) Priority Data:
RM2010A000293 31 May 2010 (31.05.2010) IT

(71) Applicant *(for all designated States except US)*: **CONSIGLIO NAZIONALE DELLE RICERCHE** [IT/IT]; Piazzale Aldo Moro, 7, I-00185 Roma (IT).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **TULLO, Apollonia** [IT/IT]; c/o Consiglio Nazionale delle Ricerche, Piazzale Aldo Moro, 7, I-00185 Roma (IT). **MANGIULLI, Marina** [IT/IT]; c/o Consiglio Nazionale delle Ricerche, Piazzale Aldo Moro, 7, I-00185 Roma (IT). **SBISA', Elisabetta** [IT/IT]; c/o Consiglio Nazionale delle Ricerche, Piazzale Aldo Moro, 7, I-00185 Roma (IT). **PESOLE, Graziano** [IT/IT]; c/o Consiglio Nazionale delle Ricerche, Piazzale Aldo Moro, 7, I-00185 Roma (IT).

(74) Agents: **CAPASSO, Olga** et al.; C/o De Simone and Partners S.p.A., Via Vincenzo Bellini 20, I-00198 Roma (IT).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2011/151777 A1

(54) **Title**: METHOD FOR THE PREPARATION AND AMPLIFICATION OF REPRESENTATIVE AND STRAND- SPECIFIC LIBRARIES OF CDNA FOR HIGH THROUGHPUT SEQUENCING, USE THEREOF, KIT AND CARTRIDGES FOR AUTOMATION KIT

(57) **Abstract**: The present invention relates to a rapid, simple, effective and economical method for preparing and amplifying representative cDNA libraries for next-generation high throughput sequencing, kit and cartridges for automation kits which use such method manually or in an automatic manner.

IDT Exhibit 1020.4273

METHOD FOR THE PREPARATION AND AMPLIFICATION OF REPRESENTATIVE AND STRAND-SPECIFIC LIBRARIES OF cDNA FOR HIGH THROUGHPUT SEQUENCING, USE THEREOF, KIT AND CARTRIDGES FOR AUTOMATION KIT

FIELD OF THE INVENTION

The present invention relates to a rapid, simple, effective and economical method for preparing and amplifying representative and strand-specific cDNA libraries for next-generation high throughput sequencing, kit and cartridges for automation kit which use such method manually or in an automatic manner.

PRIOR ART

Over the past few years new next-generation sequencing (NGS) platforms have been developed, including the "454 GS FLX" by Roche, "SOLiD" by Applied Biosystems and "Genome Analyzer" by Illumina. Initially used almost exclusively for sequencing and/or resequencing whole genomes and for determining single nucleotide polymorphisms (SNP), today they are increasingly employed in an effective manner for many other applications, including identification of biological markers and pathogenic agents in biodiversity studies and metagenomics in the biomedical, agri-food, environmental and industrial sectors.

Among the most recent and promising applications, the sequencing of whole transcriptome is becoming fundamental for studying the regulation of gene expression. In fact, this type of analysis enables the identification and quantification not only of mRNAs, but also of small RNAs, such as micro-RNAs and other non-coding RNAs, whose biological meaning is still largely unknown. Among next-generation sequencing platforms, the Roche 454 GS FLX pyrosequencer makes it possible to obtain in each run over a million reads, each with a length of about 400 bases, even up to about 800 bases. In addition to enabling the identification and relative quantification of mRNAs, sequences of this length also permit the identification of new splicing sites and variants, expressed in different physiopathological conditions. The shorter length of the reads, but of higher number, of the platforms produced by Illumina (100-150 bp) and Applied Biosystems (35-75 bp), on the other hand, makes them particularly suitable for sequencing small RNAs and quantifying the levels of transcripts.

IDT Exhibit 1020.4274

Irrespective of the NGS platform used, the first step required for transcriptome analysis is the construction of cDNA libraries for high throughput sequencing.

One of the main problems is represented by the fact that the starting amount of RNA is often very exiguous and that some transcripts, despite having considerable biological significance, are very scarcely abundant. To this it may be added that around 90-95% of the total RNA extracted consists of ribosomal RNA (rRNA) which, for a successful experiment, should be removed. This means starting off with a higher amount of total RNA, which is not always available.

Existing protocols for cDNA library construction involve the following steps: RNA retrotranscription, synthesis of the second strand and in some cases amplification of the retrotranscribed RNA species. MRNA retrotranscription can be done using oligo-dT or small random primers. In general it is preferred to use oligo-dT because rRNAs are not polyadenylated and as a result are excluded in the retrotranscription reaction. The use of oligo-dT has the large disadvantage of excluding not only rRNAs from the retrotranscription reaction, but also every non-polyadenylated species of RNA, such as the RNA species with a regulatory function or bacterial or viral RNA species (metatranscriptomics) that it is desirable to identify in human cells in some physiological or pathological conditions, such as, for example, in the case of multiple sclerosis. Moreover, the long T or A homopolymer tails can give rise to sequences of poor quality, in particular for the Roche 454 GS FLX sequencer. In order to remedy this technical problem, it is possible to use oligo-dT primers with a sequence at 5' containing a restriction site, which cleaves 14-16 nt downstream. In this manner the T tail can be removed after retrotranscription; otherwise, it is possible to use for retrotranscription oligo-dT primers in which another nucleotide is inserted in every 4 or 5 T, so as to interrupt the long T tail.

The synthesis of the second strand complementary to the retrotranscribed strand is in some cases followed by an amplification reaction using adapters, which are ligated to the ends of the retrotranscribed fragments. This makes it possible to amplify the starting material and in particular the least represented species of RNA which are otherwise hardly detected. For the amplification reaction, Taq polymerase is generally used. However, this protocol has the drawback of producing a population of fragments smaller than 2Kb; as a result, nebulization, which is the fragmentation process used in the standard procedure for preparing shotgun libraries for sequencing, might not give optimal results, since it is more suitable for molecules with a high molecular weight.

Alternative protocols provide for the isolation of mRNAs with oligo-dT, fragmentation of the RNA molecules, gel purification of low molecular weights and then retrotranscription. In this case the nebulization step is omitted. Amplification of the retrotranscribed fragments is not envisaged, with the drawback that the least represented RNA species cannot be identified.

At the same time, another important element for the preparation of cDNA libraries for massive sequencing is the preservation of the information relating to which of the two DNA strands was actually transcribed and thus belongs to the original pool of mRNA; this characteristic is hereinafter called strand specificity. In recent years it has been seen that many genomic regions give rise to transcripts originating from both DNA strands. Antisense transcription plays an important regulatory role. Moreover, overlapping genes are common in the compact genomes of prokaryotes and lower eukaryotes.

It follows that distinguishing from which DNA strand each RNA has been transcribed - RNA that will then be retrotranscribed into cDNA in a library - considerably increases the value of a transcriptome experiment because, during the mapping of sequences on the reference genome, it renders unambiguous the attribution of reads to the DNA strand from which the RNA was transcribed. Therefore, information about strand specificity makes it possible to distinguish coding transcripts from antisense ones, to determine from which strand other non-coding RNAs are synthesized, to define more precisely the exact boundaries of adjacent genes, transcribed from opposite strands, and to determine the correct levels of expression of coding and non-coding RNA molecules, transcribed from overlapping sequences.

WO2006/086668 describes a method for obtaining a representative library in which the possibility of verifying the strand specificity of the cDNAs contained therein is not guaranteed. The primers of the invention are not described.

WO2004/070053 also describes a method for obtaining a representative library in which the possibility of verifying the strand specificity of the cDNAs contained therein is not guaranteed. In that document as well, the primers of the invention are not described.

Disclosed in the state of the art are numerous methods for preserving information about the strand that fall into two groups: one involves the ligation of distinct adapters, with known orientation, at 5' and 3' of the transcript, prior to retrotranscription, whereas in the other one, a strand is marked by introducing chemical modifications allowing it to be distinguished from the complementary strand. Therefore, it doesn't exist a method for

preparing representative and strand-specific cDNA libraries which solves the problems of the prior art.

DESCRIPTION OF THE INVENTION

The present invention enables the preparation of cDNA libraries which can be sequenced by means of the principal massive sequencing platforms today available on the market. The method of the invention solves the described problems providing representative and strand-specific libraries in a rapid, simple and economical manner.

The innovative aspects of the procedure regard the study of the order of the various steps for preparing the sample and optimization of the performance of massive sequencing. In particular the new process is able to work with:

1. use of exiguous starting amounts of total RNA (even 400-500 ng);
2. considerable reduction in rRNA (about 90%);
3. retrotranscription starting from 30-40 ng of polyadenylated and non-polyadenylated transcripts with specific primers capable of preserving strand information, designed for the purpose;
4. ligation of the retrotranscripts;
5. amplification with DNA polymerase with high fidelity and processivity, such as for example Phi29;
6. high yields of amplified DNA (from around 30 μg).

With the strategy devised it is possible to subject to massive sequencing even samples of precious RNAs available in small amounts and otherwise hardly analysable, above all as regards poorly represented transcripts. The fact that the starting amounts of RNA (both total and devoid of rRNA) are halved compared to many protocols and that with the high yields achieved through the procedure of the invention it is possible to use the cDNA library for further subsequent validation experiments, and the strand specificity, are also extremely relevant.

The innovative aspects of the procedure regard the definition of the order of the various steps for preparing the sample and optimization of the performance of massive sequencing both for the production of data and for analysis.

In this context the authors have designed a rapid, simple, effective, economical procedure which makes it possible to start off with halved amounts of total RNA (400-500 ng), to

considerably reduce rRNA (about 90%), to construct a library of polyadenylated and non-polyadenylated transcripts that is also representative of the least abundant species as it includes an amplification step after retrotranscription, and to preserve strand specificity, which, at the time of analysis, allows to identify which DNA strand was actually transcribed.

A cDNA library represents the set of transcripts expressed by a specific cellular system. Highly expressed genes will be represented by many transcripts in the library, whereas genes expressed at low levels will be less represented, or even not at all. A cDNA library is said to be representative when it contains, in at least one copy, all of the transcripts, including the least abundant ones. A cDNA library is said to be strand specific when it enables identification of the DNA strand from which the RNA, later retrotranscribed to cDNA, was transcribed.

DNA polymerases differ from one another in a number of properties, such as processivity, proofreading activity, strand displacement activity, i.e. the capacity to displace the DNA encountered downstream during synthesis, and yields. Among the available DNA polymerases, the following table indicates which have high levels of proofreading and strand displacement activity, and the associated error rate.

	3'—>5' Exonuclease	Error rate ($x10^{-6}$)	Strand Displacement
Bst DNA Polymerase, Large Fragment	–		++++
Bst DNA Polymerase, Full Length	–		–
Taq DNA Polymerase	–	285	–
Vent$_R$® DNA Polymerase	++	57	++
Vent$_R$® (exo–) DNA Polymerase	–	190	+++
Deep Vent$_R$® DNA Polymerase	+++		++
Deep Vent$_R$® (exo–) DNA Polymerase	–		+++
9°N$_m$™ DNA Polymerase	+		+++
T7 DNA Polymerase	++++	15	–

E. coli DNA Polymerase I	++	9	–
T4 DNA Polymerase	++++	<1	–
phi29 DNA Polymerase	++++	9,5	+++++
Phusion® High-Fidelity DNA Polymerase	++++	0.44	
Phusion® High-Fidelity Hot Start DNA Polymerase	++++	0.44	
Phusion® High-Fidelity Hot Start II DNA Polymerase	++++	0.44	
Phire® Hot Start DNA Polymerase	+		–
Phire® Hot Start II DNA Polymerase	+		–
Phusion® Flash High-Fidelity DNA Polymerase	++++		
LongAmp™ *Taq* DNA Polymerase	++		–

The authors have used various DNA polymerases, including Phi29 Polymerase, which are distinguished by the following characteristics: i) strand displacement activity, which enables effective and uniform amplification also in the presence of secondary structures and sequences hardly amplifiable; ii) proof-reading activity among the highest today available in the realm of high-fidelity Taq Polymerases; iii) high processivity, which allows the generation of very long synthesized products, ideal for the nebulization step, without disassociation from the template; iv) high yields of amplified product starting from small amounts of template. Analogous polymerases can likewise be used.

In fact, the method of the present invention makes it possible to obtain, starting from an exiguous amount of total RNA (about 400-500 ng), about 30-40 ng of RNA (the rRNA component considerably reduced) to be retrotranscribed and amplified with a final yield greater than about 30 μg of high molecular weight DNA. Therefore, the method is particularly suitable for achieving excellent fragmentation results with nebulization (see Figure 2). This makes it possible to have material that is more than sufficient not only for the massive sequencing procedure, but also for further analyses and validations subsequent to the result of sequencing, such as determination of the gene expression profile by Real Time PCR. In fact, the method of the invention enables a uniform amplification of all

transcripts present in the starting RNA, thanks to the intrinsic properties of Phi29 polymerase and the entire amplification procedure. In the present invention it is possible to subject to massive sequencing and further validation even precious samples of RNA and available in small amounts, otherwise hardly analysable, especially as regards poorly represented transcripts. The present invention also enables to preserve information about strand specificity so as to be able to correctly determine the direction of transcription and explore the transcriptome complexity in a more effective and accurate manner.

In the present invention a method has been developed for the preparation of cDNA libraries for massive sequencing defining the order of protocols to be used. Therefore, object of the invention is a method for obtaining a representative and strand-specific cDNA library from an RNA sample of at least about 400 ng, comprising the steps of:

a) removing ribosomal RNA from the total RNA extracted in order to obtain a fraction of rRNA-depleted RNA in which said fraction is present in a amount of about at least 30 ng;

b) retrotranscribing the rRNA-depleted fraction of RNA to cDNA;

c) purifying the synthesized cDNAs;

d) ligating and amplifying the purified cDNA to obtain a cDNA library of at least about 30 μg, in which the retrotranscription step is performed in the presence of 5' phosphorylated primers having one of the following sequences:

5'-TCGCGATCGTCGNNNNNNNN-3' (SEQ ID No. 2) or 5'-GCGGCCGCNNNNNN-3' (SEQ ID No. 1),

where N is any one of the four nucleotides A, T, G and C, and the primers are present in equimolar amounts, and wherein the amplification step is performed in the presence of a DNA polymerase with high strand-displacement and proof-reading activity, high processivity and yields.

In a preferred aspect the method of the invention also comprises a preliminary step of extracting total RNA from a sample, for example a biological, cellular, etc. sample.

Preferably the DNA polymerase is Phi29 DNA polymerase. A person skilled in the art can select other DNA polymerases with activity comparable to that of the DNA polymerase Phi 29. In a preferred aspect, the primers are present in an amount of about 25 to about 50 μM and the rRNA-depleted RNA is present in an amount of about 30 ng to about 300 ng. In a more preferred aspect the primers are present in an amount of about 50 μM.

A further object of the invention is a representative and strand-specific cDNA library obtainable with the above-described method, and the use thereof for high throughput sequencing of the cDNA sequences contained therein.

A further object of the invention is a kit for preparing a representative and strand-specific cDNA library from a sample comprising:

- means for extracting the total RNA from the sample;
- means for removing the ribosomal RNA from the total RNA extracted;
- means for retrotranscribing the RNA to cDNA comprising 5' phosphorylated primers having one of the following sequences :

5'-TCGCGATCGTCGNNNNNNNN-3' (SEQ ID No.2) or 5'-GCGGCCGCNNNNNN-3' (SEQ ID No. 1),

where N is any one of the four nucleotides A, T, G and C, and the primers are present in equimolar amounts;

- means for purifying the synthesized cDNAs;
- means for ligating and amplifying the purified synthesized cDNAs comprising Phi29 polymerase.

Further object of the invention is a cartridge for automation kits for carrying out the above-described method in a manual, automated or semi-automated manner.

In consideration of the high sample sequencing processivity of NGS platforms, the possibility of preparing the libraries in an automated manner constitutes an advantage in terms of time and number of samples. For this purpose the reaction components are prepared in cartridges, i.e. supports containing in micro-containers, in an orderly and sequential manner, all or part of the reagents (solutions, primers etc.) in the amount and order envisaged by the method. The cartridges are advantageously inserted in automation platforms (liquid handlers) which run the reactions of the protocol (or some steps thereof) in an automated and controlled manner (controls on the precision of the aspirated volumes, aerosols, sterility, etc.).

The method of the invention, by means of cartridges usable in an automated or semi-automated system, is capable of:

-handling a large number of samples (also tagged with barcodes)
-optimizing sample preparation times
-eliminating any volume errors and aerosol
-optimizing the amount of the numerous reagents included in the kit to be used manually

-simplifying its execution with a single product.

The present invention will be described through non-limitative examples, with reference to the following figures:

Figure 1. Agilent profile of test RNA subjected to the "RiboMinus Eukaryote Kit for RNA Seq" (Invitrogen) (in blue) compared with the profile of the respective total starting RNA (in red). Note how the peaks corresponding to 28S and 18S ribosomal RNAs, evident in the starting sample, are nearly absent in the post-Ribominus sample, which is thus enriched in the non-ribosomal component of RNA.

Figure 2. Agilent profile of: A. a library after nebulization and B. after removal of low molecular weights. Both profiles correspond to the conditions recommended for preparing shotgun libraries to be subjected to sequencing with the 454 GS FLX.

Figure 3. Results of sequencing, conducted with the Roche 454 GS FLX Titanium platform, of two transcriptome samples prepared with the primer of SEQ ID No.1.

Figure 4. Results of sequencing, conducted with the Roche 454 GS FLX Titanium platform, of two transcriptome samples prepared with the primer of SEQ ID No. 2.

MATERIALS AND METHODS

Extraction of total RNA

The total RNA was extracted from cell lines and tissues using the *RNeasy plus mini* kit (QIAGEN Cat. No. 74134), which assures efficient removal of genomic DNA and enrichment in transcripts more than 200 nucleotides long (a large part of the RNAs shorter than this length are selectively removed).

Removal of ribosomal RNA (rRNA)

The removal of the ribosomal component from the total starting RNA was effected using the *RiboMinus Eukaryote Kit for RNA Seq* (Invitrogen Cat. No. A1083708).

rRNA removal was conducted according to the procedure indicated by the manufacturer, shown here below:

1. Set a thermostatically-controlled water bath to 75°C.
2. To a sterile 1.5 ml tube add:

Component	Volume per reaction
Total RNA	minimum 400 ng (*) in <20µl
Ribominus probe (15pmol/µl)	10µl

Hybridization Buffer	300µl

(*) Though the kit indicates 2µg as the minimum amount of starting material, the authors have verified that, starting from 400-500 ng of total RNA, it is possible obtain around 30-40 ng of RNA depleted of the ribosomal fraction, sufficient for the construction and amplification of a representative and strand-specific cDNA library prepared according to the procedure of the present invention.

3. Incubate the tube at 75°C for 5 minutes to denature RNA.
4. Allow the sample to cool slowly and incubate at 37°C for 30 min. To promote sequence-specific hybridization between rRNA and *Ribominus probes* (eukaryote rRNA specific and 5'-biotin labelled probes) it is important to allow slow cooling.

While the sample is at 37°C, prepare the streptavidin-coated magnetic beads.

5. Resuspend the solution containing the beads by vortexing.
6. Pipet 750µl of the bead suspension into a sterile 1.5ml tube.
7. Place the tube on a magnetic separator for 1 min, waiting for the beads to settle on the tube side closest to the magnet, aspirate and discard the supernatant.
8. Add 750µl of H_2O-DEPC to the beads and resuspend them by gentle vortexing.
9. Again place the tube on the magnetic separator for 1 min, aspirate and discard the supernatant.
10. Repeat steps 8 and 9 once.
11. Resuspend the beads in 750µl of Hybridization Buffer and transfer 250µl of beads to a new tube.
12. Place the tube with the remaining 500µl of beads on the magnetic separator for 1 min, aspirate and discard the supernatant.
13. Resuspend the bead pellet in 200µl of Hybridization Buffer and maintain the tube at 37°C until use.
14. After the incubation at 37°C for 30 minutes of the hybridized sample, briefly centrifuge to collect it at the bottom of the tube.
15. Transfer the sample (about 330µl) to the previously prepared beads (step 13) and mix well.
16. Incubate in the water bath at 37°C for 15 min. During incubation, gently mix about every 2 minutes. At the end, briefly centrifuge to collect the sample at the bottom of the tube.

17. Place the tube on the magnetic separator for 1 min to allow the rRNA-probe complex to adhere to the sides. Collect the about 530µl supernatant containing rRNA-depleted RNA.
18. Place the tube containing 250µl beads (from step 11) on the magnetic separator for 1 min, aspirate and discard the supernatant.
19. To this tube with beads attached to the side, add the about 530µl supernatant containing the RNA of interest (from step 17). Mix well. The RNA already depleted of rRNA is then incubated a second time with the streptavidin beads to assure better removal.
20. Incubate the tube at 37°C for 15 min. During incubation, gently mix about every 2 min. At the end, briefly centrifuge to collect the sample at the bottom of the tube.
21. Transfer the supernatant containing RNA depleted of the ribosomal component, about 530µl, to a new 2 ml tube.

The subsequent steps, necessary in order to purify and concentrate the isolated RNA, require the use of the *Ribominus Concentration Module* (Invitrogen Cat. No. K155005), used according to the manufacturer directions:

22. Add to the tube containing RNA depleted of the ribosomal component 1 volume of Binding Buffer and 1 volume of 100% Ethanol. Mix thoroughly.
23. Transfer 700µl of this mixture to the column supplied with the kit.
24. Centrifuge at ≥ 12000 x g for 1 min at room temperature. Discard the flow-through.
25. Repeat steps 23 and 24 until the entire sample (from step 22) is loaded in the column.
26. Add to the column 200µl of Wash Buffer complemented beforehand with 100% ethanol (6ml 100% ethanol per 1.5ml of Wash Buffer).
27. Centrifuge at ≥ 12000 x g for 1 min at room temperature. Discard the flow-through.
28. Repeat the wash with another 200µl of Wash Buffer complemented with 100% ethanol.
29. Discard the *Collection Tube* supplied with the kit and place the column into the *Wash Tube.*
30. Centrifuge the column at maximum speed for 2-3 min at room temperature to remove any residual Wash Buffer. Discard the *Wash Tube.*
31. Place the column in the *Recovery Tube* supplied with the kit.

32. Add 10-15µl (•) of RNase-free Water to the centre of the column. Incubate at room temperature for 1 min.
33. Centrifuge at maximum speed for 1 min at room temperature.
34. The *Recovery Tube* now contains purified, concentrated RNA depleted of the ribosomal component. Place immediately on ice to proceed with downstream applications or store at -80°C until use. This is the last step provided by the kit.

(•) in the present invention the sample was eluted in 15µl of RNase-free H_2O supplied with the kit.

The concentration of the eluted RNA was determined using a Thermo Scientific Nanodrop 2000c micro-volume spectrophotometer.

The efficiency of removal of ribosomal RNAs from the total starting RNA was evaluated by means of an Agilent 2100 bioanalyzer using the specific chip for RNA. As can be seen in Figure 1, where an example case of the results obtained is shown, the profile for the RNA subjected to the aforesaid procedure shows an efficient removal of 18S and 28S ribosomal RNAs relative to the total starting RNA.

Retrotranscription

The RNA depleted of the ribosomal component was converted into single-stranded cDNA using the "SuperScript III (SSIII) First-strand synthesis system for RT-PCR" (Invitrogen Cat. No. 18080-051).

The associated protocol was modified in relation to some components, times and reaction conditions. Of extreme importance was the use of specific primers suitably designed by the authors so as to render applicable the cDNA amplification strategy, envisaged by the "Quantitect Whole Transcriptome" kit, and included in the present invention as step 5' in the procedure described herein, to the sample preparation for the principal ultra massive sequencing technologies available today. The authors also fine tuned the concentration of such primers in the retrotranscription reaction.

The RNA retrotranscription was performed according to the following protocol defined by the authors, who substituted the first step - single-stranded cDNA synthesis - of the "Quantitect Whole Transcriptome" kit (QIAGEN Cat. No. 207043):

1'. Prepare Mix 1 in a 0.5ml Eppendorf tube, adding the components listed here below:

Component	Amount/Volume per reaction
post-Ribominus RNA (Fig. 1)	minimum about 30 ng (**)

Tag-Lig-hexamers or octamers (50μM)	1μl (***)
dNTPs (10mM)	1μl
H_2O-DEPC	up to 10μl

(**) One could start off from an amount of RNA depleted of the ribosomal component smaller than 30 ng (even 1pg), but that would increase the risk that transcripts with a low number of copies might be only partially represented or absent from the sample at the start of the transcriptome amplification step. This in turn cannot assure high reproducibility in the cDNA amplification. However, an experimenter can assess whether to start off from a smaller amount of RNA based on his own experimental requirements and on the number of reads he intends to obtain for each transcript of the examined sample, in particular for the least represented ones.

(***) The primers, indicated as *Tag-Lig-octamers or hexamers*, are 5'-phosphorylated and have sequences, respectively: 5'-TCGCGATCGTCGNNNNNNNN-3' (SEQ ID No. 2) or 5'-GCGGCCGCNNNNNN-3' (SEQ ID No. 1), where N represents any one of the four nucleotides. They are random octamers or hexamers preceded at 5' by a 12 or 8 bp sequence. In the primer of SEQ ID No. 2, the TCG trinucleotide precedes and follows the restriction site (underlined) of the rare-cutter enzyme PvuI, very rarely present in the human genome, at considerably lower frequencies than other restriction sites considered rare because present approximately once every 10 Mb. The trinucleotide upstream and downstream of the restriction site assures that strand information will be preserved, since its sequence, not being palindromic, unlike the PvuI site, is different depending on whether it is read in a 5'-3' direction on one DNA strand or on its complementary strand (palindromic sequence means a sequence of nucleic acid - DNA or RNA - which is the same whether read in a 5'-3' direction on one strand or its complementary strand). More precisely, the cDNA strand directly complementary to the mRNA from which it was synthesized will contain the sequence 5'-TCG-3' upstream and downstream of the PvuI site, whereas the strand complementary to it, and hence identical to the starting mRNA, will contain the trinucleotide 5'-CGA-3' in the same positions. Accordingly, following ligation of the massive sequencing adapters to the double-stranded cDNA, the information on strand specificity will not be lost. The trinucleotide 5'-TCG-3' was selected precisely because it was the rarest result in the RefSeq *Homo Sapiens* collection, representative of the human transcriptome, as well as the 12 bp sequence formed by the PvuI site preceded

and followed by the trinucleotide 5'-TCG-3' resulted considerably rare. The necessity of inserting a known sequence at 5' is tied not only to maintaining strand information, but also to the nature of the amplification procedure used with the "Quantitect Whole Transcriptome" kit (QIAGEN Cat. No. 207043), which envisages a step of ligating all the single-stranded cDNA molecules to give concatenamers around 20kb long. The presence of the known sequence enables identification of the points where the ligation occurred. The fact that it is a particularly rare sequence makes it more functional as a label of the synthesized cDNAs. In the primer of SEQ ID No. 1 is present the restriction site for NotI, another rare cutter enzyme.

2'. Incubate Mix1 at 75°C for 5 min

3'. Place immediately on ice for 5 min.

4'. During this time interval, prepare Mix2, adding each component in the order indicated, with the modifications introduced by the authors:

Component	Volume per reaction
10X RT Buffer per SSIII	2µl
$MgCl_2$ (25Mm)	4µl
DTT (0.1M)	2µl
RNaseOUT (40U/µl)	1µl
SuperScript III reverse transcriptase (200U/µl)	1µl
DMSO (SIGMA-Aldrich)	0.5µl
Total volume	10.5µl

5'. Add Mix2 to Mix1, mix and collect at the bottom by brief centrifugation.

6'. Incubate the final mix thus obtained as indicated below with the authors' modifications:

Temperature	Duration
25°C	10 min
45°C	90 min
95°C	5 min

7'. After blocking the reaction with ice, add 1µl of RNaseH and incubate at 37°C for 20 min.

8'. The synthesized cDNA can be stored at -20°C or immediately used for the subsequent steps of the procedure.

Purification of synthesized cDNAs

The cDNA synthesis reaction was subjected to purification using the *MinElute PCR Purification* kit (QIAGEN Cat. No. 28004) to eliminate primers, nucleotides, enzymes and salts, following the instructions provided in the kit and eluating the cDNA mixture in 10µl of Buffer EB (10mM Tris·Cl, pH 8.5).

cDNA ligation and amplification

The purified cDNAs resuspended in Buffer EB (§) were ligated together in long concatenamers and subsequently amplified following the steps after the first retrotranscription step of the *QuantiTect Whole Transcriptome* kit (QIAGEN Cat. No. 207043) (see retrotranscription protocol).

(§) It was verified that Buffer EB does not interfere with the subsequent cDNA ligation and amplification procedures envisaged by the *QuantiTect Whole Transcriptome* kit (QIAGEN Cat. No. 207043).

The cDNA ligation reaction was carried out according to the following protocol envisaged by the aforesaid kit, except for the mixture of the purified retrotranscripts (step 2''):

1''.Prepare the ligation Mix, adding the individual components in the following order:

Component	Volume per reaction
Ligation Buffer	6µl
Ligation Reagent	2µl
Ligation Enzyme1	1µl
Ligation Enzyme2	1µl
Total Volume	10µl

2''.Add the 10µl of the ligation Mix to the 10µl of the purified cDNA mixture, mix and briefly centrifuge.

3''.Incubate at 22°C for 2h,

After incubation:

4''.Prepare the amplification Mix, which makes use of Phi29 *REPLI-g* polymerase (from the *QuantiTect Whole Transcriptome* kit, QIAGEN Cat. No. 207043), according to the following scheme:

Component	Volume per reaction
REPLI-g Midi Reaction Buffer	29µl
REPLI-g Midi DNA Polymerase	1µl
Total volume	30µl

5''.Add the 30µl of the amplification Mix to the 20µl of the ligation reaction, mix well and collect at the bottom by brief centrifugation

6''.Incubate at 30°C for about 4 hours (or even for 2h, as indicated in the kit)

7''.Stop the amplification reaction, incubating at 95°C for 5 min

8''.Store the amplified cDNA at -20°C until subsequent use.

The cDNA libraries thus obtained, in the form of concatenamers about 20 kb long, were sequenced with excellent results following the standard protocols for the preparation of shotgun libraries for sequencing with the Roche 454 GS FLX platform.

RESULTS

Examples of Agilent profiles for a representative and strand-specific library after nebulization of the amplified sample and after removal of low molecular weight fragments are shown in Figures 2A and 2B. The sequencing results with respect to two samples are presented in Figures 3 and 4. Following the method of the invention, a number of reads was obtained (see *passed filter* column) which was very close to the values corresponding to the highest performances of the 454 GS FLX pyrosequencer, indicated by the manufacturer.

The protocol developed in the present invention can be used to realize a new kit and/or cartridge for automation kits capable of assuring, in rapid times and at moderate costs, a uniform, faithful amplification of all transcripts in samples of small amounts of RNA to be retrotranscribed (30-40 ng) - depleted beforehand of the ribosomal component - thanks to the singular properties of Phi29 polymerase and information about strand specificity. At the same time the kit and/or cartridge for automation kits enables sequencing for transcriptome and metatranscriptome studies using the main NGS platforms (454 GS FLX by Roche, SOLiD by Applied Biosystems and Genome Analyzer by Illumina). Thanks to the high yields of amplified DNA (~30 µg), moreover, it is possible to use the cDNA library produced for further subsequent validation experiments. Finally, the presence of the sequence at 5' of retrotranscription primers of SEQ ID No. 2, makes it possible to preserve information about the direction of transcription, rendering unambiguous the attribution of reads to the strand the RNA was transcribed from. An unambiguous attribution considerably simplifies the step of bioinformatic analysis downstream of the massive sequencing of the transcriptome. The kit may therefore represent an evolution of the "Quantitect Whole Transcriptome" (QIAGEN), since it contains the same last two steps of

the protocol thereof, i.e. ligation of the single-stranded cDNAs and amplification of the concatenamers by Phi29 polymerase, while substituting the first retrotranscription step with the one described in the Materials and Methods section, paragraph on Retrotranscription, carried out with primers designed by the authors and in the conditions defined in the present invention, such as to make the amplified product suitable for subsequent ultra massive sequencing while at the same time maintaining information about the direction of transcription.

Considering the complexity of the procedures envisaged by NGS platforms and the high cost of a single run, the availability of a single and standardized kit for preparing transcriptome samples and capable of optimizing sequencing performance, the number of reads useful for analysis and information of strand-specificity is of extreme interest for users. It should be considered, in fact, that the number of NGS platforms present in the market is growing constantly and rapidly and in Europe alone there are currently: 200 Roche units, 210 Illumina units and 80 Applied Biosystems units, many of which in service.

CLAIMS

1. Method for obtaining a representative and strand-specific cDNA library from an RNA sample of at least about 400 ng, comprising the steps of:

a) removing the ribosomal RNA from the RNA to obtain an rRNA-depleted fraction of RNA wherein said fraction is present in an amount of at least 30 ng;

b) retrotranscribing the rRNA-depleted fraction of RNA to cDNA;

c) purifying the synthesized cDNAs;

d) ligating and amplifying the purified cDNAs to obtain a cDNA library of at least about 30 μg,

wherein the retrotranscription step is performed in the presence of 5' phosphorylated primers having one of the following sequences:

5'-TCGCGATCGTCGNNNNNNNN-3' (SEQ ID No.2) or 5'-GCGGCCGCNNNNNN-3' (SEQ ID No. 1),

where N is any one of the four nucleotides A, T, G and C, and the primers are present in equimolar amounts, and wherein the amplification step is performed in the presence of a DNA polymerase with high strand-displacement and proof-reading activity, high processivity and yields.

2. Method according to claim 1, comprising a preliminary step of extracting the total RNA from a sample.

3. Method according to one of the preceding claims, wherein the DNA polymerase is Phi29 polymerase.

4. Method according to any of previous claims, wherein the primers are present in an amount of about 25 to about 50 μM and the rRNA-depleted RNA is present in an amount of about 30 ng to about 300 ng.

5. Method according to claim 4, wherein the primers are present in an amount of about 50 μM.

6. Representative and strand-specific cDNA library obtainable with the method of claims 1 to 5.

7. Use of the representative and strand-specific cDNA library according to claim 6 for high throughput sequencing.

8. Kit for preparing a representative and strand-specific cDNA library from a sample comprising:

- means for extracting the total RNA from the sample;
- means for removing the ribosomal RNA from the extracted total RNA;
- means for retrotranscribing the RNA to cDNA comprising 5' phosphorylated primers having one of the following sequences:

5'-TCGCGATCGTCGNNNNNNNN-3' (SEQ ID No.2) or 5'-GCGGCCGCNNNNNN-3' (SEQ ID No. 1),

where N is any one of the four nucleotides A, T, G and C, and the primers are present in equimolar amounts,

- means for purifying the synthesized cDNAs;
- means for ligating and amplifying the purified synthesized cDNAs comprising the polymerase Phi29.

9. Cartridge for automation kit for carrying out the method according to claims 1 to 5.

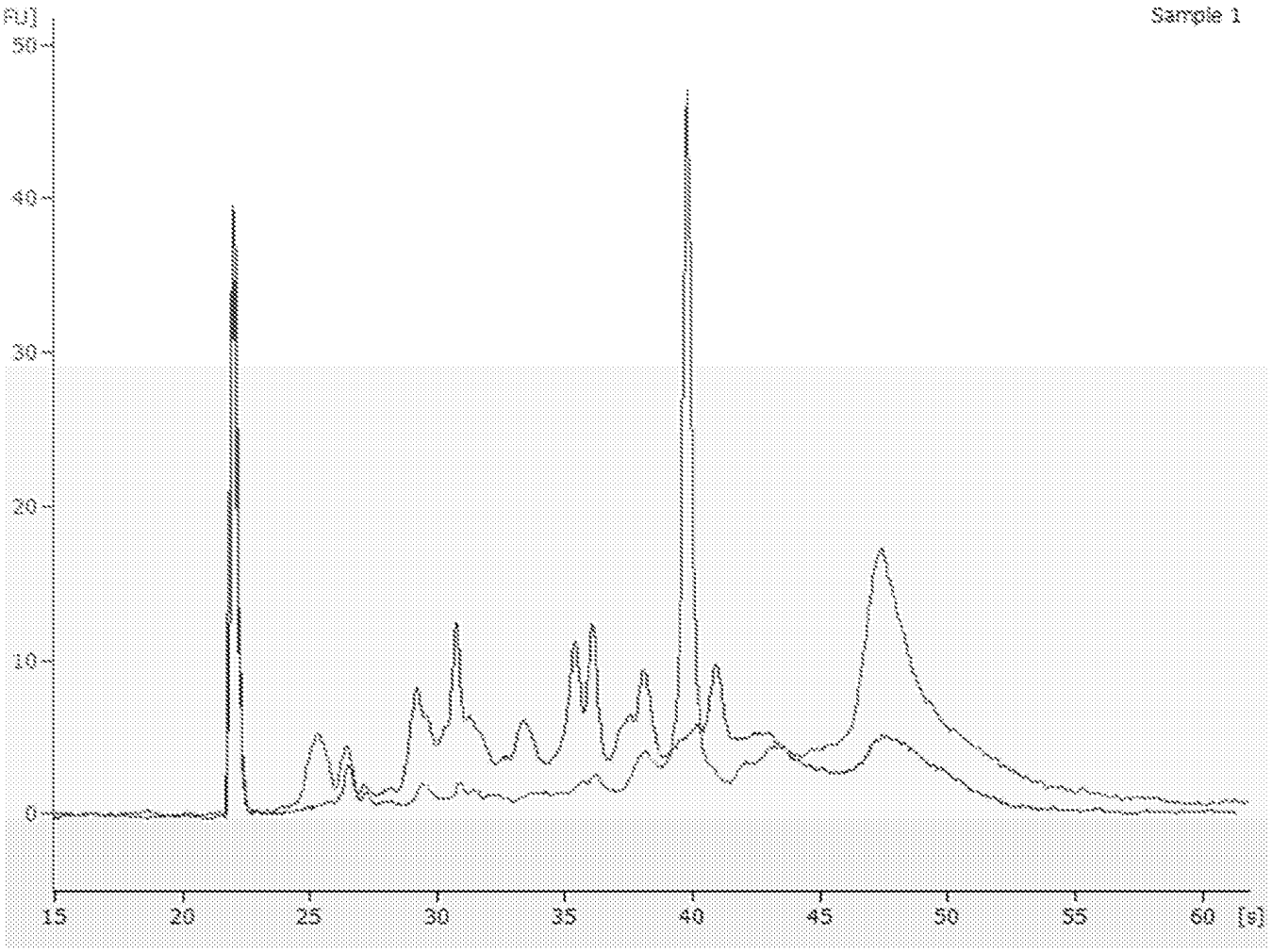
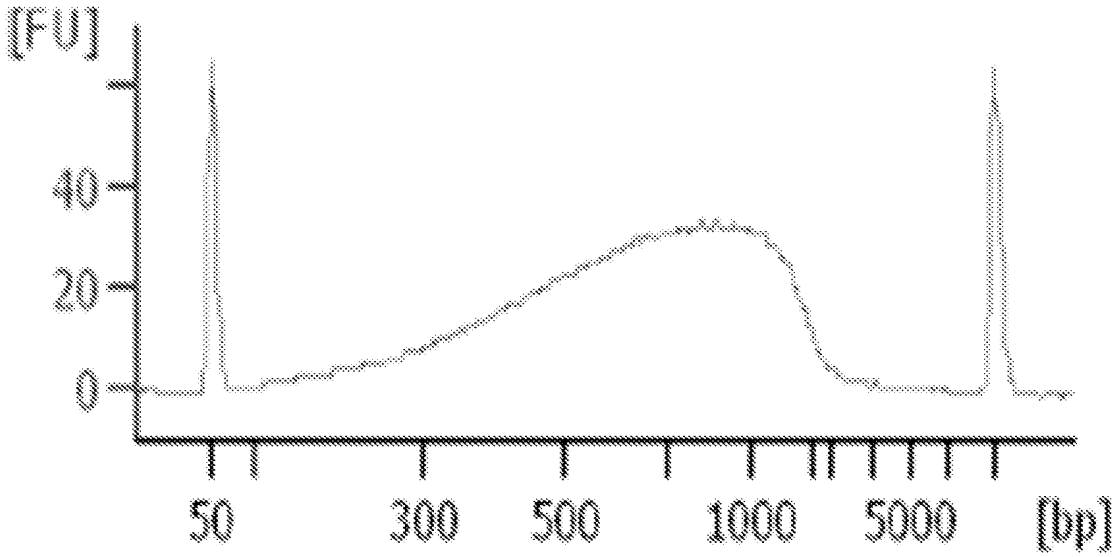


Fig. 1

A.


[FU]
40
20
0
50
300
500
1000
5000
[bp]

B.

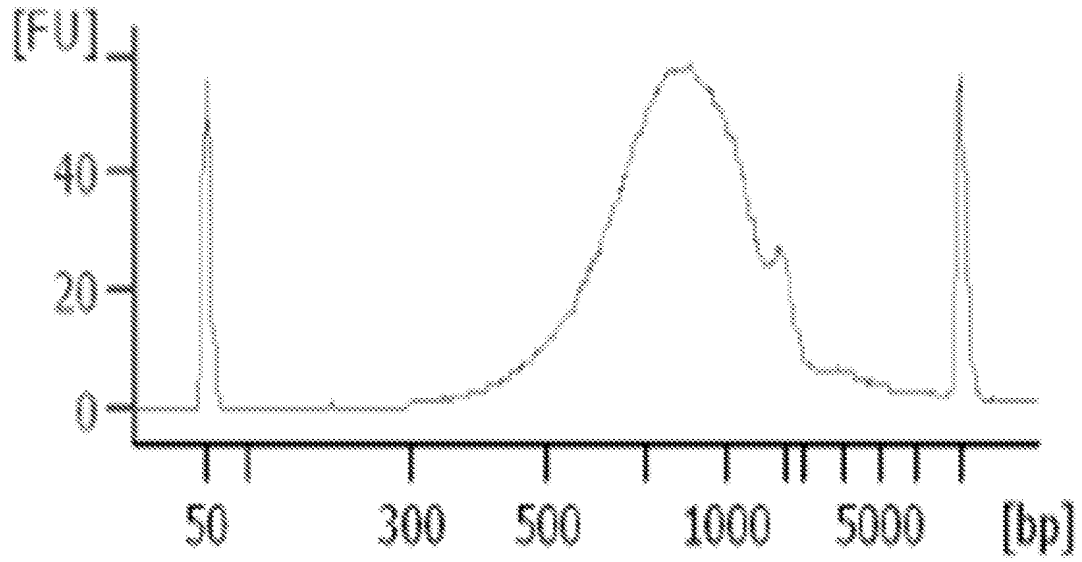
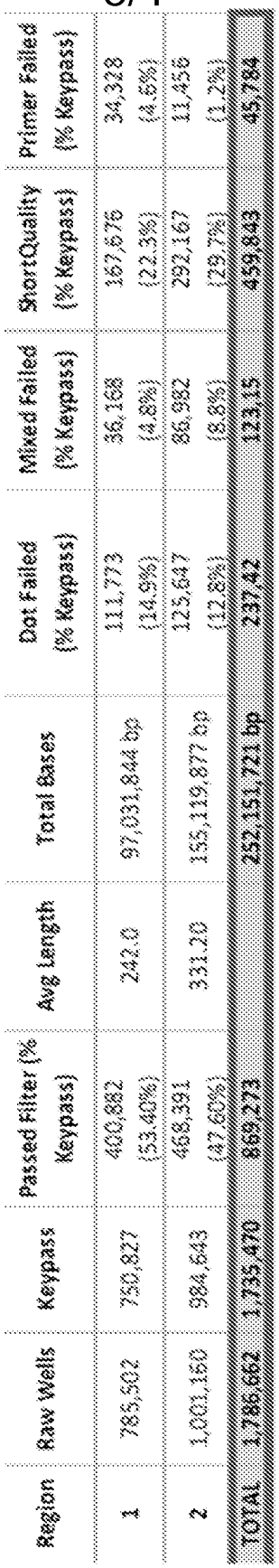

[FU]
40
20
0
50
300
500
1000
5000
[bp]

Fig. 2

3/4



Region	Raw Wells	Keypass	Passed Filter (% Keypass)	Avg Length	Total Bases	Dat Failed (% Keypass)	Mixed Failed (% Keypass)	ShortQuality (% Keypass)	Primer Failed (% Keypass)
1	785,502	750,827	400,882 (53.40%)	242.0	97,031,844 bp	111,773 (14.9%)	36,168 (4.8%)	167,676 (22.3%)	34,328 (4.6%)
2	1,001,160	984,643	468,391 (47.60%)	331.20	155,119,877 bp	125,647 (12.8%)	86,982 (8.8%)	292,167 (29.7%)	11,456 (1.2%)
TOTAL	1,786,662	1,735,470	869,273		252,151,721 bp	237.42	123.15	459,843	45,784

Fig. 3

Region	Raw Wells	Keypass	Passed Filter (% Keypass)	Avg Length	Total Bases	Dot Failed (% Keypass)	Mixed Failed (% Keypass)	ShortQuality (% Keypass)	Primer Failed (% Keypass)
1	942,317	871,141	504,879 (57.96%)	311.83	157,434,969 bp	148,379 (17.03%)	60,238 (6.91%)	143,595 (18.04%)	13,521 (1.55%)
2	993,759	944,715	551,068 (58.33%)	336.83	185,062,182 bp	145,120 (15.36%)	85,487 (9.05%)	153,651 (17.22%)	8,992 (0.95%)
TOTAL	1,936,076	1,815,856	1,055,947	324.35	342,497,151 bp	293,499	145,725	297,246	22,513

Fig. 4

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2011/052369

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/10 C12Q1/68 C40B40/08 C40B50/06
ADD.

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)
C12N C12Q C40B

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

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, CAB Data, Sequence Search, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2006/086668 A2 (EPICT TECHNOLOGIES [US]; JENDRISAK JEROME [US]; MEIS JUDITH [US]; MEIS) 17 August 2006 (2006-08-17) page 36, line 25 - page 37, line 19	5,7,8
X	WO 2004/070053 A2 (AMERSHAM BIOSCIENCES CORP [US]; NELSON JOHN [US]) 19 August 2004 (2004-08-19) the whole document	5,8
X	DE 199 13 934 C1 (DEUTSCHES KREBSFORSCH [DE]) 25 May 2000 (2000-05-25) the whole document	5,7,8
X	WO 2007/019444 A2 (EUCLID DIAGNOSTICS LLC [US]; FREIJE WADIHA [US]; BRIKUN IGOR [US]) 15 February 2007 (2007-02-15) the whole document	5,7,8

-/--

[X] Further documents are listed in the continuation of Box C.

[X] 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 document 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	Date of mailing of the international search report
4 October 2011	11/10/2011

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Hornig, Horst

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2011/052369

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 104 792 A (SILVER JONATHAN [US] ET AL) 14 April 1992 (1992-04-14) the whole document	5
X	WO 2006/110314 A2 (AMBION INC [US]; MENDOZA LEOPOLDO G [US]; MOTURI SHARMILI [US]; SETTER) 19 October 2006 (2006-10-19) the whole document	5
A	US 2003/175709 A1 (MURPHY GEORGE L [US] ET AL) 18 September 2003 (2003-09-18) the whole document	1-8
A	DEAN F B ET AL: "Rapid amplification of plasmid and phage DNA using Phi29 DNA polymerase and multiply-primed rolling circle amplification", GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, WOODBURY, NY, US, vol. 11, no. 6, 1 June 2001 (2001-06-01), pages 1095-1099, XP002223174, ISSN: 1088-9051, DOI: DOI:10.1101/GR.180501 the whole document	1-8
A	WO 02/057447 A2 (GENETICA INC [US]) 25 July 2002 (2002-07-25) the whole document	1-8
A	US 6 613 516 B1 (CHRISTIANS FRED C [US] ET AL) 2 September 2003 (2003-09-02) the whole document	1-8
A	"RiboMinusTM Transcriptome Isolation Kits" In: "New Product Catalog 2006", 2006, INVITROGEN, US, XP002614727, pages 33-34, the whole document	1-8

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2011/052369

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2006086668 A2	17-08-2006	NONE	
WO 2004070053 A2	19-08-2004	EP 1590470 A2	02-11-2005
		JP 2006516410 A	06-07-2006
		US 2004180372 A1	16-09-2004
DE 19913934 C1	25-05-2000	WO 0058458 A2	05-10-2000
WO 2007019444 A2	15-02-2007	US 2009137415 A1	28-05-2009
US 5104792 A	14-04-1992	NONE	
WO 2006110314 A2	19-10-2006	EP 1871913 A2	02-01-2008
		US 2009264635 A1	22-10-2009
US 2003175709 A1	18-09-2003	AU 2002359789 A1	09-07-2003
		CA 2468854 A1	03-07-2003
		EP 1463835 A2	06-10-2004
		WO 03054162 A2	03-07-2003
WO 02057447 A2	25-07-2002	AU 2002247014 A1	30-07-2002
US 6613516 B1	02-09-2003	US 2005059009 A1	17-03-2005

IDT Exhibit 1020.4299

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 August 2010 (12.08.2010)

PCT

(10) International Publication Number
WO 2010/091246 A2

(51) International Patent Classification:
B65B 61/00 (2006.01) *B01J 14/00* (2006.01)

(21) International Application Number:
PCT/US2010/023310

(22) International Filing Date:
5 February 2010 (05.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/150,481 6 February 2009 (06.02.2009) US

(71) Applicant *(for all designated States except US)*: **NORTHWESTERN UNIVERSITY** [US/US]; 633 Clark Street, Evanston, IL 60208 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **KELSO, David, M.** [US/US]; 741 8th Street, Wilmette, IL 60091 (US). **AGARWAL, Abhishek, K.** [US/US]; 607 Hinman Avenue, Evanston, IL 60202 (US). **SUR, Kunal** [IN/US]; 1021 Dempster Street, Evanston, IL 60201 (US). **BEEBE, David, J.** [US/US]; 4647 Tonyawatha Trail, Monona, WI 53716 (US).

(74) Agent: **ARENSON, Tanya, A.**; Casimir Jones, S.C., 2275 Deming Way, Suite 310, Middleton, WI 53562 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

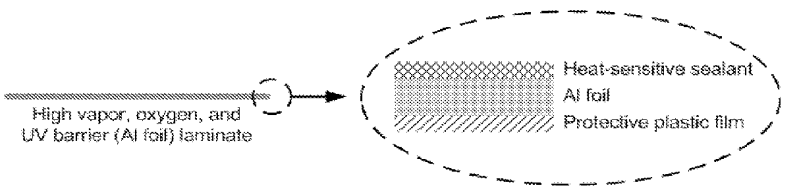
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) **Title**: BURSTABLE LIQUID PACKAGING AND USES THEREOF

Figure 1



(57) **Abstract**: The present invention relates to systems, devices, and methods for performing biological and chemical reactions. In particular, the present invention relates to the use of burstable liquid packaging for delivery of reagents to biological and chemical assays.

WO 2010/091246 A2

BURSTABLE LIQUID PACKAGING AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims priority to provisional application 61/150,481, filed February 6, 2009, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to systems, devices, and methods for performing biological and chemical reactions. In particular, the present invention relates to the use of burstable liquid packaging for delivery of reagents to biological and chemical assays.

BACKGROUND OF THE INVENTION

Many existing methods of storing liquid reagents used in medical diagnostics are done in sterilized plastic bottles that often require cold chain technology for shipping, transportation and storage at the final destination. Although such methods are feasible in most developed nations, such a requirement poses challenges and presents higher costs for developing nations. There may be issues during shipping, customs, and provision of reliable and consistent electricity for the refrigeration equipment at the site for storing the reagents. Any one of these has the potential to expose the reagents to high temperatures, rendering them useless for clinical use. Furthermore, because the reagents are stored and delivered in bulk, a skilled clinical laboratory technician and precision fluid-handling equipment are often required for precision pipetting and aliquoting for the individual medical diagnostic tests. This manual operation increase cross-contamination between samples, takes additional processing time, and increases the cost of administering and processing a diagnostic test.

Depending on how a diagnostic system operates, liquid delivery to a diagnostic test cartridge can be done using precision pipetting, or directly through the stock liquid reagent bottles via tubing, precision pumps, and valves. Such fluidic components add increased cost and complexity to the design of the diagnostic system. Furthermore, they are often prone to contamination, failure (requiring mechanical servicing and/or replacement), and leaks. Additional methods of storing and delivery reagents are needed.

IDT Exhibit 1020.4301

In particular need are compositions and methods for transporting and storing reagents at ambient temperature.

SUMMARY OF THE INVENTION

The present invention relates to systems, devices, and methods for performing biological and chemical reactions. In particular, the present invention relates to the use of burstable liquid packaging for delivery of reagents to biological and chemical assays. In some embodiments, the present invention provides assay systems and methods of their use, comprising: a liquid packaging component comprising one or more liquid storage compartments, wherein the liquid storage compartments comprise a liquid and are covered with a burstable seal; a seal bursting component configured to burst the burstable seal; and an assay device configured to accept liquids from the liquid storage compartments. In some embodiments, the burstable seal is a foil (e.g., aluminum) laminate. In some embodiments, the laminate comprises aluminum foil sandwiched between a protective plastic film and a heat-sensitive sealant. In some embodiments, the seals are peelable or permanent. In some embodiments, the seal bursting component comprises plungers that compress said liquid storage compartment under conditions such that the seals are burst (e.g., by peeling open). In some embodiments, the plungers are driven by one or more motors. In other embodiments, the plungers are manually driven. In some embodiments, the liquid packaging component and chamber within the assay device are connected via fluid conduits or are in direct contact. In some embodiments, the liquid packaging component comprises one or more liquid storage compartments. In some embodiments, the liquid storage compartments comprise less than 60%, and preferably less than 50% air by volume. In some embodiments, the liquid storage compartments comprise less than 400 μl air. In some embodiments, the liquid packaging component further comprises one or more alignment pins to secure the liquid storage compartment to the liquid packaging component.

In some embodiments, the liquid storage compartment comprises a tear-drop clamp that applies uniform pressure across a portion of the burstable seal perimeter. In some embodiments, the tear drop clamp does not apply pressure to the portion of the burstable seal that is in communication with the assay device. In some embodiments, the

IDT Exhibit 1020.4302

liquid storage compartments comprise reagents for performing a biological or chemical assay. In some embodiments, the assay is selected from a diagnostic assay or a research assay (e.g., nucleic acid based assays (e.g., PCR) or protein based assays).

DESCRIPTION OF THE FIGURES

Figure 1. Cross-sectional diagram of a typical (opaque) high vapor and oxygen barrier aluminum (Al) foil laminate. (b) a thin sheet of Al foil that behaves as the barrier, and (c) a thin protective plastic film that prevents the Al foil from being damaged or torn during handling and processing.

Figure 2. Process diagram showing how a blister (here, hemispherical) is made in a high vapor and oxygen barrier laminate that can be pressure formed. (a) The cold forming station comprises a male plug with vent hole, stripper plate with a through-hole to allow the male plug to pass through, and a matching female cavity with vent hole. A corner radius is machined into the female cavity to prevent the laminate film from tearing/pinching during cold-forming. (b) Pressure is applied on the stripper plate to hold the laminate film firmly. Next, pressure is applied on the male plug to create the blister shape. (c) Liquid can be precisely aliquoted into the cold-formed blister (top); a photograph of a blister is also shown (bottom).

Figure 3. Diagram of a cold-formed high vapor, oxygen, and UV barrier (Al foil) laminate blister with liquid. The apex of the liquid droplet is in the same plane as the top of the laminate blister. The cross-section of a typical Al foil laminate is show on the right. The liquid rests on the heat-sensitive sealant side of the blister laminate.

Figure 4. Two types of heat seals – peelable and permanent. Peelable seals (top) are fabricated at lower temperatures using different sealant materials. They have a lower peel strength and are designed to be opened post-manufacturing. Permanent seals (middle) are typically fabricated at higher temperatures using similar sealant materials and have higher peel strengths. The heat sealing can also be extended to bonding laminate to rigid plastics, as shown in the bottom image.

Figure 5. Design schematics of the contour (impulse) heat seal press (left) used to seal the liquids inside the cold-formed blister. The upper platen (top right) holds an interchangeable heat seal (here, a circular geometry) band that is designed to match a

IDT Exhibit 1020.4303

blister's geometry. The lower platen (bottom right) holds an interchangeable silicone/aluminum pressure also designed to match the blister geometry.

Figure 6. General process for heat sealing a cold-formed Al foil laminate blister to store liquids. (a-b) The cold formed-blister is positioned in the lower platen relief. Foil laminate #2 is positioned on top such that the sealants face one another. (c) The upper platen, with the heat seal band, comes down on top of the blister, applying synchronized pressure and heat to heat seal (peelable) the two laminates together. (d) The final packaged and heat sealed blister.

Figure 7. Cross-sectional diagrams of liquid stored in a heat sealed (peelable) Al foil laminate blister. (a) Parameters of the heat sealing process are: x – distance between the edge of the blister and heat seal band; w – heat seal width. (b) Due to the Al foil laminate in both foil laminates, liquid loss does not occur through the film, but only through the heat seals. It is a function of the blister volume, ambient temperature, heat-sensitive sealant material properties (permeability), heat seal surface area (a function of w), and *tseal* (thickness of the final heat seal).

Figure 8. Perspective and cross-sectional diagrams showing how a packaged blister may be integrated with a rigid, disposable (plastic) cartridge using double-sided tape. (a) Conceptual schematics showing a perspective view (solid view – left; transparent view – right) of the cartridge and integrated blister. (b) Double-sided adhesive is bonded to the cartridge, with appropriate slits for the input port(s). The packaged blister is subsequently bonded to the opposite side of the double-sided adhesive. Foil laminate #2 is extended beyond the blister and has a matching slit that is aligned with the input port. (c) Alternative method for (b) where the cold-formed laminate is not adhesively bonded to the cartridge. (d-e) Alternative methods – foil laminate #2 is trimmed to the blister size and the packaged blister is bonded to the cartridge using double-sided.

Figure 9. Cross-sectional diagrams showing how a packaged blister may be integrated with a rigid, disposable (plastic) cartridge using heat seals, with the option of adding double-sided adhesive. (a) The packaged blister is heat sealed to the rigid cartridge. The heat-sensitive sealant on the cold-formed blister laminate is similar to the rigid cartridge material. (b) Alternative method – Foil laminate #2 is adhesively bonded

IDT Exhibit 1020.4304

to the rigid cartridge. The cold-formed blister laminate is subsequently heat sealed to the rigid cartridge.

Figure 10. Mechanical clamping mechanism to burst a packaged blister to deliver the liquid to the cartridge chamber. The initial state of this example cartridge is shown in Figure 8(e). (a-b) Cross-sectional (left) and corresponding top-view (right) diagrams. A mechanical clamp is positioned around the edge of the blister and around the input port of the rigid cartridge to ensure the liquid from the blister will only flow in one direction – from the blister to the input port. A mechanical plunger is used to apply uniform pressure on the blister until the peelable seal breaks, allowing the stored liquid to escape, enter the input port, and flow into the cartridge chamber.

Figure 11. Schematic showing how an exemplary rigid cartridge with three packaged blisters may interface with the mechanical clamping and bursting module. The current design shows three separate linear motors operating each of the mechanical clamp and plunger combination, but has the potential for being driven by a single linear stepper motor.

Figure 12. Cross-sectional diagrams showing an integrated cartridge and packaged blister with both peelable and permanent heat seals. (a) Initial state of the cartridge – same configuration shown in Figure 9(b). (b) A mechanical plunger is aligned with and pressed against the cold-formed blister. Due to its lower peel strength, the peelable seal breaks in a random location, and the liquid flows out of the blister into the cartridge chamber. The permanent heat seal does not burst.

Figure 13. An example of a diagnostic cartridge that has three packaged foil laminate blisters integrated within it on the underside which store three separate aqueous and/or non-aqueous liquids. The diagrams also illustrate how a lyophilized assay bead may be readily integrated into the cartridge for dissolution with the respective liquid. A hydrophobic air permeable membrane provides a point for air to escape as liquid is dispensed into the respective chambers. (a) The initial state of the cartridge. (b) A blister is burst open, dispensing the liquid through the input port and channel into chamber 3. The lyophilized bead dissolves. (c) The second blister is burst open, dispensing the second liquid in a similar fashion. (d) The third blister is burst open similarly. Here, this

IDT Exhibit 1020.4305

blister has a non-aqueous liquid which helps prevent contamination by separating chamber 3 from both chamber 1 and overflow.

Figure 14. Photograph of a blister crusher. (a) Front view of a 3-blister crusher showing the three tear-drop clamps and plungers that mate with the respective blisters. (b) Side view of the 3-blister crusher showing the tear-drop clamps and springs that provide the clamping force. Here, individual stepper motors drive each of the plungers. The cartridge fits between the aluminum plate and mounting plate. (c) Schematic of the cartridge showing the position of the three respective blisters (shown by the concentric gray circles), input ports, channels, and chambers.

Figure 15. Exemplary blister crushing mechanism. (a) A cross-sectional diagram of a fluidic cartridge that has an adhesively bonded blister on the backside. The blister is positioned such that the top of the heat seal perimeter is directly below the input port. (b) A stepper motor is used to apply pressure on most of the heat seal perimeter via the tear-drop clamp. Next, a plunger, also driven by the stepper motor, presses against the blister and bursts the peelable heat seal nearest to the input port. Air escapes into the channel and chambers, followed by the liquid reagent. (c) Top view diagram showing a blister and the position of the tear-drop clamp. The additional clamping force prevents the heat seals in the clamped areas from being burst open. Only at the location marked by *x* does the heat seal peel open, allowing first the air to escape, and subsequently, the liquid reagent.

Figure 16. Example schematic of a modified fluidic cartridge (front and back views) showing three blisters and how they are interfaced with the fluidic cartridge, and and additional plastic pieces (blister guards) to ensure the blisters only burst in one specified location – marked by the *x*.

Figure 17. (a) Photograph of one of the three tear-drop clamps showing the plunger (with relief) inside. An o-ring on the outside periphery of the tear-drop clamp ensures intimate contact with the blister surface. (b) Photograph of a cartridge positioned in the blister crusher. The inset photo shows the position of the blister with respect to the input port on the fluidic cartridge.

Figure 18. Graph showing the forces required to crush the respective blisters – elution, oil, and lysis. The graph shows that both the size of the blister, and liquid

IDT Exhibit 1020.4306

volume seem to affect the force required to burst open the blisters. Vertical bars show the standard deviation.

Figure 19. Data plot showing the effect of blister stroke on the liquid volume fill capacity for two different blister diameters – 0.55" and 0.72".

Figure 20. Schematic of a characterization cartridge used to determine the dead volume for a given blister geometry and fill volume.

Figure 21. Data plot showing the effect of liquid volume in given blister geometry on the load (force) required to burst it open.

Figure 22. Diagram of a cold formed blister.

Figure 23. (a) Top-view schematic of a cold-formed blister in a foil laminate. Two alignment holes are punched through the foil laminate that serve as alignment guides. They are punched along the central axis (dashed line) point of the cold-formed blister and are outside the circular heat seal width (dashed-dotted line). (b) Trimetric schematic view of the cold-formed blister in a foil laminate showing the blister and two alignment holes.

Figure 24. Top-view schematic of the lidstock. Three holes are punched into the lidstock material – two serve as alignment holes for the heat sealing process, and subsequent integration and positioning with the rigid test cartridge; the third hole serves as the liquid thru port.

Figure 25. Brief outline of the heat sealing procedure, which shows how the cold-formed blister (a) is aligned with the lidstock (b). Retractable alignment pins on the lower platen of the impulse heat sealer facilitate the positioning and alignment (c-f).

Figure 26. (a) Schematic of the general heat seal band used for heat sealing the blisters. It has an active area in the geometry of a donut ring with tab extensions on two sides, which overlap with the three punched holes on the cold-formed blister and lidstock. The inactive area is for electrical contact and does not promote heat sealing. (b) Top-view of a heat sealed blister showing the heat seal perimeter (outlined by the dashed-dotted line) and how it overlaps the three punched holes. (c) A heat sealed blister which has been punch-cut to the desired profile shape (i.e., trimmed around the heat seal perimeter).

IDT Exhibit 1020.4307

Figure 27. (Left) Cross-sectional views of the rigid assay cartridge with a double-sided transfer adhesive bonded to it. A packaged blister with the three punched holes can be bonded to the rigid assay cartridge (shown on the right). The bonding is on the same level across the blister and therefore, there are no channels created as before.

Figure 28. (a) Top-view schematic of a cold-formed blister with reagent. (b-c) Two embodiments showing mechanical clamps positioned on the respective blister, aligned using the alignment holes. Pressure from the clamp is applied on the heat seal perimeter of the blister.

DEFINITIONS

To facilitate an understanding of this disclosure, terms are defined below:

"Purified polypeptide" or "purified protein" or "purified nucleic acid" means a polypeptide or nucleic acid of interest or fragment thereof which is essentially free of, e.g., contains less than about 50%, preferably less than about 70%, and more preferably less than about 90%, cellular components with which the polypeptide or polynucleocide of interest is naturally associated.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, which is separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

"Polypeptide" and "protein" are used interchangeably herein and include all polypeptides as described below. The basic structure of polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as

IDT Exhibit 1020.4308

peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types.

A "fragment" of a specified polypeptide refers to an amino acid sequence which comprises at least about 3-5 amino acids, more preferably at least about 8-10 amino acids, and even more preferably at least about 15-20 amino acids derived from the specified polypeptide.

The term "immunologically identifiable with/as" refers to the presence of epitope(s) and polypeptide(s) which also are present in and are unique to the designated polypeptide(s). Immunological identity may be determined by antibody binding and/or competition in binding. The uniqueness of an epitope also can be determined by computer searches of known data banks, such as GenBank, for the polynucleotide sequence which encodes the epitope and by amino acid sequence comparisons with other known proteins.

As used herein, "epitope" means an antigenic determinant of a polypeptide or protein. Conceivably, an epitope can comprise three amino acids in a spatial conformation which is unique to the epitope. Generally, an epitope consists of at least five such amino acids and more usually, it consists of at least eight to ten amino acids. Methods of examining spatial conformation are known in the art and include, for example, x-ray crystallography and two-dimensional nuclear magnetic resonance.

A "conformational epitope" is an epitope that is comprised of a specific juxtaposition of amino acids in an immunologically recognizable structure, such amino acids being present on the same polypeptide in a contiguous or non-contiguous order or present on different polypeptides.

A polypeptide is "immunologically reactive" with an antibody when it binds to an antibody due to antibody recognition of a specific epitope contained within the polypeptide. Immunological reactivity may be determined by antibody binding, more particularly, by the kinetics of antibody binding, and/or by competition in binding using as competitor(s) a known polypeptide(s) containing an epitope against which the antibody is directed. The methods for determining whether a polypeptide is immunologically reactive with an antibody are known in the art.

IDT Exhibit 1020.4309

As used herein, the term "immunogenic polypeptide containing an epitope of interest" means naturally occurring polypeptides of interest or fragments thereof, as well as polypeptides prepared by other means, for example, by chemical synthesis or the expression of the polypeptide in a recombinant organism.

"Purified product" refers to a preparation of the product which has been isolated from the cellular constituents with which the product is normally associated and from other types of cells which may be present in the sample of interest.

"Analyte," as used herein, is the substance to be detected which may be present in the test sample, including, biological molecules of interest, small molecules, pathogens, and the like. The analyte can include a protein, a polypeptide, an amino acid, a nucleotide target and the like. The analyte can be soluble in a body fluid such as blood, blood plasma or serum, urine or the like. The analyte can be in a tissue, either on a cell surface or within a cell. The analyte can be on or in a cell dispersed in a body fluid such as blood, urine, breast aspirate, or obtained as a biopsy sample.

A "capture reagent," as used herein, refers to an unlabeled specific binding member which is specific either for the analyte as in a sandwich assay, for the indicator reagent or analyte as in a competitive assay, or for an ancillary specific binding member, which itself is specific for the analyte, as in an indirect assay. The capture reagent can be directly or indirectly bound to a solid phase material before the performance of the assay or during the performance of the assay, thereby enabling the separation of immobilized complexes from the test sample.

The "indicator reagent" comprises a "signal-generating compound" ("label") which is capable of generating and generates a measurable signal detectable by external means. In some embodiments, the indicator reagent is conjugated ("attached") to a specific binding member. In addition to being an antibody member of a specific binding pair, the indicator reagent also can be a member of any specific binding pair, including either hapten-anti-hapten systems such as biotin or anti-biotin, avidin or biotin, a carbohydrate or a lectin, a complementary nucleotide sequence, an effector or a receptor molecule, an enzyme cofactor and an enzyme, an enzyme inhibitor or an enzyme and the like. An immunoreactive specific binding member can be an antibody, an antigen, or an antibody/antigen complex that is capable of binding either to the polypeptide of interest

IDT Exhibit 1020.4310

as in a sandwich assay, to the capture reagent as in a competitive assay, or to the ancillary specific binding member as in an indirect assay. When describing probes and probe assays, the term "reporter molecule" may be used. A reporter molecule comprises a signal generating compound as described hereinabove conjugated to a specific binding member of a specific binding pair, such as carbazole or adamantane.

The various "signal-generating compounds" (labels) contemplated include chromagens, catalysts such as enzymes, luminescent compounds such as fluorescein and rhodamine, chemiluminescent compounds such as dioxetanes, acridiniums, phenanthridiniums and luminol, radioactive elements and direct visual labels. Examples of enzymes include alkaline phosphatase, horseradish peroxidase, beta-galactosidase and the like. The selection of a particular label is not critical, but it should be capable of producing a signal either by itself or in conjunction with one or more additional substances.

"Solid phases" ("solid supports") are known to those in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic or non-magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, and others. The "solid phase" is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of microtiter wells, glass or silicon chips, are all suitable examples. It is contemplated and within the scope of the present invention that the solid phase also can comprise any suitable porous material.

As used herein, the terms “detect”, “detecting”, or “detection” may describe either the general act of discovering or discerning or the specific observation of a detectably labeled composition.

The term "polynucleotide" refers to a polymer of ribonucleic acid (RNA), deoxyribonucleic acid (DNA), modified RNA or DNA, or RNA or DNA mimetics. This term, therefore, includes polynucleotides composed of naturally-occurring nucleobases, sugars and covalent internucleoside (backbone) linkages as well as polynucleotides having non-naturally-occurring portions which function similarly. Such modified or substituted polynucleotides are well-known in the art and for the purposes of the present invention, are referred to as "analogues."

IDT Exhibit 1020.4311

As used herein, the term "nucleic acid molecule" refers to any nucleic acid containing molecule, including but not limited to, DNA or RNA. The term encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxylmethyl) uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-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, and 2,6-diaminopurine.

The term "nucleic acid amplification reagents" includes conventional reagents employed in amplification reactions and includes, but is not limited to, one or more enzymes having polymerase activity, enzyme cofactors (such as magnesium or nicotinamide adenine dinucleotide (NAD)), salts, buffers, deoxynucleotide triphosphates (dNTPs; for example, deoxyadenosine triphosphate, deoxyguanosine triphosphate, deoxycytidine triphosphate and deoxythymidine triphosphate) and other reagents that modulate the activity of the polymerase enzyme or the specificity of the primers.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides such as an oligonucleotide or a target nucleic acid) related by the base-pairing rules. 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

IDT Exhibit 1020.4312

importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The term "homology" refers to a degree of identity. There may be partial homology or complete homology. A partially identical sequence is one that is less than 100% identical to another sequence.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (i.e., the strength of the association between the nucleic acids) is impacted by such factors as the degree of complementary between the nucleic acids, stringency of the conditions involved, the Tm of the formed hybrid, and the G:C ratio within the nucleic acids.

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

As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds, under which nucleic acid hybridizations are conducted. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Thus, conditions of "weak" or "low" stringency are often required when it is desired that nucleic acids which are not completely complementary to one another be hybridized or annealed together.

The term "wild-type" refers to a gene or gene product which has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product which displays

IDT Exhibit 1020.4313

modifications in sequence and or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

The term "oligonucleotide" as used herein is defined as a molecule comprised of two or more deoxyribonucleotides or ribonucleotides, preferably at least 5 nucleotides, more preferably at least about 10-15 nucleotides and more preferably at least about 15 to 30 nucleotides, or longer. The exact size will depend on many factors, which in turn depends on the ultimate function or use of the oligonucleotide. The oligonucleotide may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, or a combination thereof.

Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. A first region along a nucleic acid strand is said to be upstream of another region if the 3' end of the first region is before the 5' end of the second region when moving along a strand of nucleic acid in a 5' to 3' direction.

When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide.

The term "primer" refers to an oligonucleotide which is capable of acting as a point of initiation of synthesis when placed under conditions in which primer extension is initiated. An oligonucleotide "primer" may occur naturally, as in a purified restriction digest or may be produced synthetically.

IDT Exhibit 1020.4314

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to systems, devices, and methods for performing biological and chemical reactions. In particular, the present invention relates to the use of burstable liquid packaging for delivery of reagents to biological and chemical assays.

In some embodiments, the present invention provides a disposable liquid packaging module that stores liquids, both aqueous and nonaqueous, in sealed high vapor, oxygen, and UV barrier laminates (e.g., aluminum foil laminates) blisters, and has the capacity to deliver the fluids by bursting the seals using applied pressure.

In some embodiments, such packaging modules are used to dispense liquid into channels and respective fluidic chambers in an assay device such as, for example, a rigid (e.g., plastic disposable) diagnostic cartridge. In some embodiments, laminates, which have a sealant layer on one side, are cold-formed using pressure to create a hemispherical blister appropriately sized for the necessary liquid volume; liquids are precisely aliquoted into the formed blisters; a secondary flat laminate with a different sealant material is placed on top and a perimeter heat seal is made using a heat sealer (the seal may also be made, for example, using ultrasonic, radio frequency, and laser welding techniques). The packaged blister is aligned and adhered to the rigid cartridge, which contains an input port for fluid entry and connecting channel to the fluidic chamber. By application of a controlled pressure on the blister, the heat seal can be burst open, allowing the fluid to enter the input port, and flow through the channel into the respective chamber in the plastic cartridge. One example use of the diagnostic cartridge is for polymerase chain reaction (PCR) based detection and analysis for infections including, but not limited to HIV, Chlamydia, and Gonorrhea or other pathogens or analytes of interest.

This method of packaging and delivering liquids is designed and developed for any number of diagnostic and clinical uses, although it especially serves point-of-care and resource-limited settings, where refrigeration and cold chain technologies are not consistently available. It enables the medical diagnostic cartridge to be self-sufficient since the appropriate liquid reagents are packaged with the test. The high vapor, oxygen, and UV barrier laminates prevent contamination and evaporation of the small liquid volumes. The method of bursting the pouches and delivering the fluids to a secondary

IDT Exhibit 1020.4315

location removes the necessity of additional fluidic components, such as pumps, valves, and precision liquid metering units.

Embodiments of the invention disclosed herein provide many benefits to overcome challenges associated with existing technology:

- Self-sufficient test cartridge with on-chip liquid reagents
- High vapor, oxygen, and UV barrier storage blisters using low cost Al foil laminates prevent contamination and evaporation of the liquid until its use
- Removal of complex and costly fluid handling components, such as precision pumps, valves, and tubing
- Elimination of cold chain technology when integrated with lyophilized assay beads
- Simple liquid delivery mechanism by applying controlled pressure on the blister and bursting its seal
- Precision aliquoting of the reagents into the blisters can be done at the manufacturing site, reducing the complexity at the clinical setting

I. Liquid Storage Component

As described above, embodiments of the present invention provide liquid storage components comprising one or more blisters with burstable seals. In some exemplary embodiments, the blisters are made with the following steps and utilized in the following exemplary applications. The invention is not limited to these exemplary embodiments. Each of the following steps is described in more detail below.

1. A high vapor, oxygen, and UV barrier laminate is cold-formed using pressure to create a hemispherical blister;
2. Liquid is precisely aliquoted into the blister (e.g., in a laboratory setting, using a manual hand-operated pipet);
3. A perimeter heat seal is created between the cold-formed laminate blister and secondary laminate using one of many available heat sealing technologies (e.g., resistive, laser, radio frequency, ultrasonic);
4. Integrating a packaged blister with a rigid plastic cartridge;

IDT Exhibit 1020.4316

5. Bursting the blister is realized by placing a mechanical clamp around the blister's heat seal and input port in the plastic cartridge and applying uniform pressure on the hemispherical blister until the seal breaks between the blister and input port hole; and

6. Example of an integrated PCR diagnostic cartridge with one or more packaged blisters to realize a self-sufficient diagnostic cartridge

The below discussion describes exemplary methods of manufacture and use of blister packaging. Additional fabrication techniques and applications are within the scope of one of skill in the art.

A. Cold-forming a blister

In some embodiments, pressure (cold) formable high vapor, oxygen, other gases, and UV barrier laminates are chosen and used to create blisters into which liquids are stored. The option of choosing materials that can be cold-formed presents the advantage of lowering the production cost since heat (for thermoforming applications) is not required. These high vapor and oxygen barrier laminates can be manufactured to be transparent or opaque. Transparent laminates offer almost equal barrier protection through numerous methods, such as, for example SiO_x and Al_2O_3, and many can be either cold- or thermo- formed; however, the transparent laminates cost 4-10× as much as opaque laminates that alternatively use a thin sheet of aluminum (Al) foil to serve as a barrier. To reduce the overall cost of the disposable plastic diagnostic cartridge, some embodiments of the present invention use opaque Al or other metal foil laminates. The total thickness of such laminate films typically ranges from 0.002" to 0.012". In general, they are comprised of at least three laminates – heat-sensitive sealant, Al foil film, and plastic film to protect the Al foil from physical damage (tears, scratches) (See e.g., Figure 1).

In some embodiments, the blisters are formed in the laminate using a cold-forming station that consists of a male plug, stripper plate, and female cavity. The heat-sensitive sealant side of the laminate film, which is chosen to be compatible with liquid to be stored within it, faces the male plug for the subsequent heat sealing process.

Figure 2 shows a process diagram of one method of how a blister can be cold-formed in a high vapor and oxygen barrier laminate. The laminate film is held firmly

IDT Exhibit 1020.4317

using applied pressure between the female cavity and stripper plate, both of which are machined to a very flat and uniform surface. A minimum of 0.157" (4.0 mm) from the edge of the female cavity is preferably held firmly flat to allow for the heat sealing process in order to prevent the laminate from wrinkling. Next, with applied pressure, the male plug subsequently comes down on the laminate film, creating a hemispherical blister. The dimensions of the male plug and female cavity are designed according to the amount of liquid that needs to be stored in the blister. The diameter of the female cavity is $\geq \varphi_1 + 2t$, where φ_1 is the diameter of the male plug, and t is the thickness of the foil laminate sheet. The depth of the cold-formed blister (h) is dependent on how far the male plug is pushed into the laminate sheet and affects the total liquid capacity of the blister. The applied pressure for both the stripper plate and male plug can be realized by many different schemes, including, but not limited to, by manual compression using screws, compressed air, and stepper motors.

The shape of the blister is dependent on the shapes of the male plug and female cavity, and is not limited to the hemispherical shape (e.g., oval, square, rectangular, etc.). In some embodiments, a chamfer (corner radius) is machined in the female cavity to prevent tearing/pinching of the laminate film at the edge. In some embodiments, to prevent stiction of the laminate film to the male plug or female cavity during forming, very small vent holes are drilled into both the male plug and female cavity. The vent holes allow air to escape during the cold forming and prevent any vacuum build-up.

In one exemplary embodiments, the final aspect ratio h/Φ_1 of the blister, as shown in Figure 2(c), is ≥ 0.30. No tears or pinholes are visually observed in the laminate sheet after cold-forming.

In some embodiments, blisters are designed with a head space to allow blisters to move along a conveyor, as they would be in a full line production facility using a form/fill/seal (F/F/S) machine. This significantly reduces any chances of spilling the liquid over the brim of the blister edge. In a F/F/S system, a web of blisters will be moving down the line, accelerating and decelerating, which may cause the liquid to potentially spill over. Embodiments of the present invention overcome such issues by providing dead space in blisters.

IDT Exhibit 1020.4318

In some embodiments, the assay cartridge comprises one or more (e.g., two) alignment pins that help to secure the blister to the cartridge.

B. Liquid aliquoting into a cold-formed blister

In some embodiments, once the blister has been cold-formed, liquid can be aliquoted into it manually (e.g., using a pipetting device) or automatically using a liquid dispensing tool (e.g., during manufacturing). In one exemplary embodiment, a manual pipetting device (e.g., PIPETMAN pipetting device (0.1 μL resolution)) is used to deposit the desired liquid volume, which ranges from 10 μL to 1.0 mL. The heat-sensitive sealant film on the laminate sheet is often hydrophobic, causing aqueous liquids to have a relatively high contact angle. Here, liquid is filled in the blister until the apex of the liquid droplet aligns with the top of the laminate sheet as shown in Figure 3.

In is preferred that the apex of the liquid droplet should not be higher than the top of the foil laminate since in the subsequent heat sealing procedure, the liquid may spill/leak out of the blister and into heat sealing areas (i.e., perimeter of the blister). This is especially usefull for non-aqueous liquids since they will tend to wet the surface of the blister, creating a small contact angle compared to aqueous liquids. The heat-sensitive surface of the blister may also be surface treated by various chemical or physical methods, such as, for example, surfactants or plasma, to increase its hydrophilicity (suitable for aqueous liquids). Treatment methods are designed to not affect the heat sealing qualities of the sealant film. In some embodiments, the aqueous liquid contains a chemical component (e.g., surfactant or detergent) that makes it preferentially wet the blister surface by reducing its surface tension.

C. (Heat) sealing the blister to store the liquid

One of any number of technologies to bond/seal materials (e.g., laminate to laminate, laminate to rigid plastic, plastic to plastic) with heat-sensitive sealants may be utilized, including but not limited to constant heat sealers, impulse heat sealers, laser welding, radio frequency, and ultrasonic sealing. The exemplary method described here is based on using an impulse heat sealer, in which a mechanical pressure is first applied to the perimeter of the blister to sandwich both laminate films together, creating a liquid and

IDT Exhibit 1020.4319

vapor tight closure. Next, the power to the heat seal band is turned on, rapidly increasing the heat and melting the heat-sensitive sealants and bonding them together. The heat is turned off, but the mechanical clamping pressure is only released once the heat seal band has cooled, and the seal has set and has good strength and appearance. The advantages of this method, compared to a constant heat sealer where the heat is always on, are:

(1) a stronger seal is created with superior appearance, and (2) the liquid is not exposed to high temperatures during the heat sealing process, which can cause liquid evaporation, and vapor entrapment in the heat seals (poor seal strengths).

A heat sealer is a function of four parameters:

• Time – length of time the heat seal bands maintained at the preset temperature which depends on several parameters, including thickness of the laminate sheets and/or rigid plastics and type of seal strength desired.

• Pressure – the amount of pressure (psi) exerted down on the two materials to be sealed together (e.g., laminate to laminate, laminate to rigid plastics).

• Temperature – the temperature of the heat seal band which generally ranges from 200-500 °F.

• Heat-sensitive sealant material – are the sealant materials for both laminate sheets similar or different.

Two types of seals can be created by manipulating the combination of the aforementioned parameters: (1) peelable seals which have a lower peel strength and are designed to be opened either manually by hand, or with the assistance of an automated machine. They are created at low temperatures with dissimilar heat-sensitive sealant materials. A variety of heat-sensitive sealants are available from manufacturers specifically designed to fabricate peelable seals. (2) permanent seals which have significantly stronger peel strength and are not designed to be opened. These are created at high temperatures with similar heat-sensitive sealant materials. Figure 4 shows a conceptual diagram of how peelable and permanent heat seals are made by choosing the appropriate heat sealing temperature and sealant material. In some embodiments, a laminate sheet can be heat sealed to a rigid plastic material with similar material properties.

IDT Exhibit 1020.4320

In some embodiments, these processes are used to heat seal and store the liquid inside the cold-formed blister. In some embodiments, peelable seals are created to allow subsequent bursting of the blisters when the liquids are required for use. In some embodiments, a contour impulse heat seal press with an upper and lower platen and remote user-operated module (to configure the time, temperature, and pressure synchronization) is designed and developed to make the heat seals (see Figure 5). The upper platen holds an interchangeable circular heat seal band (band width, *w*) that creates a circular perimeter seal around the blister. The heat seal band is designed to match the blister geometry. The larger *w*, the greater the seal strength, but at the cost of requiring more power and force by the press to create the seal. Typical values for *w* for heat sealing applications range from 1/8" to 1/4". The lower platen holds interchangeable silicone/aluminum pressure plates that are machined to a specific blister geometry diameter. A relief is provided in both the pressure plate and platen to accommodate for the blister and prevent it from crushing during the heat sealing process. The general heat sealing process used in exemplary embodiments to create peelable seals is schematically shown in Figure 6.

In some embodiments, the cold-formed blister previously filled with the liquid is positioned on the lower platen, cradled by the silicone/aluminum pressure plate and relief. Foil laminate #2, which has a different sealant specifically designed for peelable seals, is positioned on top so that the heat sensitive sealant faces down (similar to the top image in Figure 4) to facilitate heat sealing. The upper platen comes down with a force, mechanically clamping the blister. Power to the heat seal band is turned on to its preset temperature for a brief time period until a heat seal is realized. The power is turned off to the heat seal band, and when it has cooled, the upper platen is raised to release the mechanical clamping. This results in a packaged and heat sealed blister with peelable seals that is ready to be integrated with a rigid (plastic) cartridge.

Figure 7 shows cross-sectional diagrams of a packaged blister. Heat seal parameters, in addition to the aforementioned blister geometries, are indicated in Figure 7(a). *x* is the distance between the edge of the cold-formed blister and where the heat seal begins. Here, it is minimized as much as possible (≤ 0.01"). As mentioned previously, *w* is the heat seal width created by the heat seal band. Some trapped air (which is

IDT Exhibit 1020.4321

compressible) in the packaged and sealed blister is necessary so that external air pressure changes (e.g., during air shipment) will not cause the blister to collapse or burst. Similarly, it is beneficial to minimize the trapped air to both reduce the overall size of the blister and dead volume spaces where the liquid can become trapped during bursting.

Figure 7(b) shows a similar cross-sectional diagram that describes how and where liquid loss can occur over time. Due to the Al foil in both foil laminates, liquid loss does not occur through the laminate sheets, but only through the heat sealed sealants. This liquid loss is a function of blister volume, ambient environmental temperature, heat-sensitive sealant material properties (permeability), heat seal surface area (a function of w), and t_{seal} – the thickness of the final heat seal.

D. Integration of a packaged blister with a rigid cartridge

In some embodiments, heat sealed blisters are integrated with an assay device (e.g., rigid cartridge). In some embodiments, polypropylene plastic is chosen as the material for the cartridge. Polypropylene is cost effective and safe and (bio)compatible for the diagnostic chemistry and readily disposable. The plastic cartridge is manufactured using one of many existing high-volume techniques, including, but not limited to, injection molding and vacuum forming. The blister is then integrated with the rigid cartridge. In some embodiments, (1) double-sided adhesives are used, while in other embodiments, (2) sealing via impulse/constant heat sealers, laser welding, radio frequency, or ultrasonic methods are used. Both techniques are described here.

In some embodiments, when using double-sided adhesives, one of three methods is selected for bonding the packaged blister to the rigid cartridge. Conceptual perspective and cross-sectional diagrams of the integrated blister and cartridge are shown in Figure 8.

Figure 8(a) shows conceptual solid (left) and semi-transparent (right) perspective views of the blister integrated with the rigid cartridge. The cartridge has an input port, channel, and chamber, which are made accessible to the packaged blister. A hydrophobic air permeable membrane is also integrated into the cartridge to allow air to escape when the blister is burst and liquid fills the cartridge channel and chamber. The blister is positioned just outside the input port to the cartridge channel and chamber. Figure 8(b-e) show corresponding cross-sectional diagrams of four exemplary methods by which the

IDT Exhibit 1020.4322

blister can be bonded to the cartridge using double-sided adhesives (not drawn to scale). Figure 8(b) shows a blister with an extended foil laminate #2 with a slit that is subsequently aligned with the cartridge's input port. Double-sided adhesive, which also has a slit for the input port, is used to bond the entire surface area of the blister to the cartridge. This is advantageous since the liquid will only flow on the foil laminate (minimal contact if any with the adhesive) and also ensure there are no air gaps between the blister and cartridge where the liquid may potentially wick into after bursting. Figure 8(c) shows a slight modification to Figure 8(b) where the cold-formed blister laminate is not adhesively bonded to the rigid cartridge. Figure 8(d-e) shows a blister with foil laminate #2 dimensionally trimmed to the blister geometry. Once the blister is burst, the liquid will again have minimal, if any, contact with the adhesive; however, there may be a potential for the liquid to wick between foil laminate #2 and rigid cartridge. This may either be avoided due to the (flat) pressure that is applied to burst the blister (See below section), or by modifying the design to Figure 8(e).

An alternative method of bonding the blister to the rigid cartridge is primarily using heat seals, with the option of adding some double-sided adhesive. Cross-sectional diagrams of this method are shown in Figure 9.

Figure 9(a) shows how a packaged blister can be bonded to a rigid cartridge using heat seals. Here, the heat-sensitive sealant material has closely matching properties with the cartridge material, and therefore can be designed to be a permanent seal (higher peel strength compared to a peelable seal). Figure 9(b) shows an alternative method by which a combination of heat seals and double-sided adhesive tape is used to bond the packaged blister to the cartridge. This reduces the chance of any air pockets that may trap the liquid once the blister has been burst. However, it also introduces the potential for the liquid to come in contact with the adhesive as it flows into the cartridge chamber via the channel and input port.

II. Embodiments of the Invention in Use

In some embodiments, the present invention provides methods of performing assays using the liquid storage, assay and seal bursting devices described herein. The present invention is not limited to the exemplary systems and methods described below.

IDT Exhibit 1020.4323

A. Bursting the blister and delivering the liquids

When the diagnostic cartridge is ready to be used, the liquid-containing blisters are burst open, and the liquid is directed into the cartridge chamber via the channel and input port. In some embodiments, a seal bursting component is utilized to burst the seals and deliver the liquid to an assay device. Two model integrated cartridge and blister systems are shown in Figure 8(e) and Figure 9(a). There are two possible bursting mechanisms by which the blister can be burst – (1) in conjunction with Figure 8(e), apply a mechanical clamp around the periphery of the blister and input port to specify the burst site on the peelable seal and ensure the liquid only flows towards the input port, and (2) in conjunction with Figure 9(b), there is no mechanical clamp, but only the plunger which bursts the blister; the difference in peelable and permanent peel strengths is leveraged. Both of these bursting mechanisms are described in detail below.

Figure 8(e) shows a model of how the peelable seal is burst with the aid of a mechanical clamp to deliver the liquid inside the cartridge chamber. Figure 10 shows cross-sectional and top-view diagrams of a ruptured peelable seal.

A mechanical clamp is aligned and pressed against the rigid cartridge on the peelable heat seals and around the input port in the cartridge as shown in Figure 10(b). The purposes of the mechanical clamp are: (1) to provide a leak-proof seal so the liquid will only flow in the specified areas, (2) specify a consistent location where the peelable seal will burst. A mechanical plunger subsequently applies uniform and controlled pressure against the packaged blister until the peelable seal bursts. The plunger regulates the amount of pressure against the packaged blister to control the liquid volume that flows out of the burst blister, through the input port, and into the cartridge chamber via the channel – see Figure 10(a). The cartridge is positioned vertically so that the trapped air bubble in the blister will rise to the top, opposite to the input port. This bursting mechanism allows a user to consistently know how and where the blister will burst and liquid will flow. It also allows a user to compensate for the presence of the trapped air bubble in the cold-formed blister, and ensure the location of the burst occurs where there is no air bubble, (bubbles rise to the top) but only liquid. The cartridge is filled only with liquid; any air bubbles that may be introduced into the cartridge will rise to the top and

IDT Exhibit 1020.4324

exit via the hydrophobic air permeable membrane. Furthermore, the clamp minimizes the potential dead air spaces where the liquid could potentially migrate to, minimizing the dead (liquid) volume. This saving on liquid loss can reduce the initial liquid fill volume and blister size and geometry, saving on material cost.

A diagram of one exemplary method of how a cartridge may interface with the mechanical clamp and plunger is shown in Figure 11. In some embodiments, the rigid cartridge has multiple (e.g., two or more, three or more, four or more, etc.) packaged blisters that is held in place by a cartridge holder. Separate mechanical clamps and plungers are appropriately dimensioned to match the blister geometry and are individually driven by a linear stepper motor. In some embodiments, a single linear stepper motor is used to drive multiple clamps and plungers. The mechanism to drive these mechanical modules is not limited to a stepper motor, but also extends to any other mechanism which creates controllable and consistent force outputs. The mechanical clamps and plungers align with the blisters that have been previously bonded to the rigid cartridge and aligned with their respective input ports.

The second method of bursting leverages the peel strength differences between peelable and permanent heat seals. This mode works with the integrated blister and cartridge designs, for example, as shown in Figure 9. Due to the differences in the peel strengths between the peelable seal, which stores the liquid in the blister, and permanent seal that bonds the blister to the rigid cartridge, it is possible to remove the mechanical clamp altogether. Figure 12 shows an exemplary cross-sectional diagram of how the integrated cartridge and blister from Figure 9(b) is aligned and coupled with the mechanical plunger.

When the mechanical plunger is pressed against the cold-formed blister, the peelable seal bursts, allowing the liquid to flow into the cartridge via the input port. The permanent heat seal remains intact, preventing the liquid from leaking outside the cartridge module.

In some embodiments, a blister crusher module is designed to crush multiple blisters that are adhesively bonded to the plastic microfluidic assay cartridge (see Figure 13(c)). The number of blisters can be adjusted according to the assay requirements. The blister crusher crushes the blisters by peeling the peelable seal in a specific location, and

IDT Exhibit 1020.4325

dispensing the liquid into the cartridge channel and chamber via an input port. See Figure 13.

In some embodiments, tear-drop clamps are used to direct the peeling of the peelable heat seal on the blister so that the bursting consistently occurs in a pre-designated location. See Figure 14. In some embodiments, a blister is adhesively bonded to the surface of the cartridge or bonded using other sealing techniques and positioned such that the top of the heat seal perimeter is directly below the input port (Figure 14(c)). Therefore the orientation of the cartridge is important when crushing the blisters (see Figure 16(b)). The tear-drop clamps apply uniform pressure across most of the heat seal perimeter, except for the top. The applied pressure should be large enough to compensate for any differences in the heat seal quality so that the bursting will have a natural tendency to always peel at the top. This ensures that once the heat seal is compromised, as shown by the x in Figure 14(c), the air will escape first, followed by the liquid. In some embodiments, the cartridge is provided with a small exit port hole inside the 'overflow' chamber, which allows liquids from the blisters to enter the cartridge channels. The air-liquid sequence is particularly important since crushing the blister where the liquid comes out first (i.e., the blister is positioned so that the bottom of the heat seal perimeter is above the input port) causes a heavy mixture of air bubbles and liquid to be dispensed into the cartridge.

In some embodiments, the clamping mechanism is integrated directly into the disposable microfluidic cartridge. For example, once the blisters are bonded to the cartridge, smaller pieces of plastic materials (blister guards) mate with the microfluidic cartridge and provide the necessary mechanical clamping pressure to ensure the blister heat seal only bursts in one location (as shown by the *x* in Figure 15).

In some embodiments, plungers for bursting blisters are shaped similar to a tear-drop shape and are sized slightly larger than the blister to ensure it covers the entire surface area for bursting. This ensures complete compression of the blister and minimizes dead (trapped) liquid volume inside the blister that does not get dispensed into the cartridge. In some embodiments, plungers have a small channel relief cut into the top which prevents complete closure of the channel that is formed during crushing (i.e., when

IDT Exhibit 1020.4326

the peelable heat seal peels apart) which allows the liquid to move from the blister into the input port and cartridge.

In embodiments that utilize multiple blisters the liquids are burst in a specific sequence to prevent any cross-contamination, especially between the lysis and elution chambers. Furthermore, the liquid dispensing ensures that no air bubbles are in the channel and chamber network since they could interfere with the subsequent assay processing. There are two exemplary sequence methods that may be used.

Method 1

1. Burst the elution blister and fill the channel and chamber.
2. Burst the lysis blister and fill the chamber, ensuring it does not overflow.
3. Burst the oil blister to fill the channel and chamber gaps between the elution and lysis. Since it is an immiscible liquid, it will prevent cross-contamination between the lysis and elution reagents.

Method 2

1. Burst the elution blister and fill the channel and chamber.
2. Burst the oil blister to fill the channel and part of the chamber. This will create a liquid buffer between the elution and lysis, perchance the lysis blister overfills and flows into the oil chamber and subsequently into the elution chamber.
3. Burst the lysis blister and fill the channel and chamber.
4. Dispense additional oil from the already-crushed oil blister to fill the remaining chamber gaps between the lysis and elution reagents.

In some embodiments, the method by which the cartridge is designed, and its vertical orientation (which facilitates use of gravity), allow for any air bubbles that have entered the cartridge to float up to the top and near or into the overflow chamber.

IDT Exhibit 1020.4327

B. Exemplary System

Figure 13(a) shows an exemplary integrated cartridge that is designed for a diagnostic assay (e.g., PCR assay), with multiple packaged blisters that contain distinct aqueous and/or non-aqueous liquids, in the volume range of 0.10 mL to 1.0 mL. The embodiment also shows integration of lyophilized assay pellets that are dissolved once a blister is burst and the liquid passes through the input port, fluid channel and into the respective chamber. In some embodiments, An overflow chamber which contains the hydrophobic air permeable membrane is also integrated into the cartridge to allow passage of air when filling the chambers with liquid. One by one, or simultaneously, depending on the application, a blister is burst, dispensing its stored liquid into the rigid cartridge – see Figure 13(b-d). Since the cartridge is kept in the upright position, as shown in the diagrams, any air bubbles that may be transferred from the blister to the rigid cartridge easily float up to the top of the channels and chambers, and into the overflow chamber. Here, the liquid filling chamber 2 is a nonaqueous liquid that helps prevent any contamination by separating the liquid in chamber 3 from both chamber 1 and overflow (US 6,103,265; herein incorporated by reference). In some embodiments, the scheme of bursting and filling the cartridge chambers with the appropriate liquids is fully automated, as previously described by Figure 11.

In some embodiments, the cartridge is positioned vertically so that the liquid inside the blister is pulled down by gravity, as shown in Figure 14(a). This also ensures that once the heat seal is burst open, the air will first escape, followed by the liquid. This minimizes the number of air bubbles injected into the cartridge channel and chambers. The inset image in Figure 16(b) shows how the blisters are positioned directly beneath the input port.

C. Applications

The systems and methods of embodiments of the present invention find use in a any number of diagnostic assays. Examples include, but are not limited to, PCR medical diagnostics tests (e.g., for infectious diseases such as HIV). In some embodiments, the systems and methods of the present invention find use in performing assays in resource

IDT Exhibit 1020.4328

limited areas where temperature controlled environments may not be available. In some embodiments, assays are packaged as self-sufficient, individual tests that will have all the necessary (liquid) reagents on-cartridge to complete the patient's analysis. By further integration with lyophilized assay beads, cold chain technology is avoided, saving on cost and making the test more robust and readily available to a larger public.

The systems and methods of embodiments of the present invention have numerous benefits and applications in any lab-on-a-chip technology where relatively small amounts of liquids must be stored with the test cartridge. Examples of research and diagnostic assays suitable for use with the systems and methods described herein are described below.

i. Sample

Any sample suspected of containing the desired material for purification and/or analysis may be tested according to the disclosed methods. In some embodiments, the sample is biological sample. Such a sample may be cells (e.g. cells suspected of being infected with a virus), tissue (*e.g.*, biopsy samples), blood, urine, semen, or a fraction thereof (*e.g.*, plasma, serum, urine supernatant, urine cell pellet or prostate cells), which may be obtained from a patient or other source of biological material, e.g., autopsy sample or forensic material.

Prior to contacting the sample with the device or as a component of the device or automated system, the sample may be processed to isolate or enrich the sample for the desired molecules. A variety of techniques that use standard laboratory practices may be used for this purpose, such as, e.g., centrifugation, immunocapture, cell lysis, and nucleic acid target capture.

In other embodiments, the methods of embodiments of the present invention are utilized to purify and/or analyze intact cells (e.g., prokaryotic or eukaryotic cells).

ii. Nucleic Acid Detection

Examples of nucleic modification/analysis/detection methods include, but are not limited to, nucleic acid sequencing, nucleic acid hybridization, and nucleic acid amplification. Illustrative non-limiting examples of nucleic acid sequencing techniques

IDT Exhibit 1020.4329

include, but are not limited to, chain terminator (Sanger) sequencing and dye terminator sequencing. Those of ordinary skill in the art will recognize that because RNA is less stable in the cell and more prone to nuclease attack experimentally RNA is usually reverse transcribed to DNA before sequencing. Illustrative non-limiting examples of nucleic acid hybridization techniques include, but are not limited to, *in situ* hybridization (ISH), microarray, and Southern or Northern blot. Nucleic acids may be amplified prior to or simultaneous with detection.

Illustrative non-limiting examples of nucleic acid amplification techniques include, but are not limited to, polymerase chain reaction (PCR), reverse transcription polymerase chain reaction (RT-PCR), transcription-mediated amplification (TMA), ligase chain reaction (LCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA). Those of ordinary skill in the art will recognize that certain amplification techniques (*e.g.*, PCR) require that RNA be reversed transcribed to DNA prior to amplification (*e.g.*, RT-PCR), whereas other amplification techniques directly amplify RNA (*e.g.*, TMA and NASBA).

The polymerase chain reaction (U.S. Pat. Nos. 4,683,195, 4,683,202, 4,800,159 and 4,965,188, each of which is herein incorporated by reference in its entirety), commonly referred to as PCR, uses multiple cycles of denaturation, annealing of primer pairs to opposite strands, and primer extension to exponentially increase copy numbers of a target nucleic acid sequence. In a variation called RT-PCR, reverse transcriptase (RT) is used to make a complementary DNA (cDNA) from mRNA, and the cDNA is then amplified by PCR to produce multiple copies of DNA. For other various permutations of PCR *see*, *e.g.*, U.S. Pat. Nos. 4,683,195, 4,683,202 and 4,800,159; Mullis et al., *Meth. Enzymol.* 155: 335 (1987); and, Murakawa et al., *DNA* 7: 287 (1988), each of which is herein incorporated by reference in its entirety.

Transcription mediated amplification (U.S. Pat. Nos. 5,480,784 and 5,399,491, each of which is herein incorporated by reference in its entirety), commonly referred to as TMA, synthesizes multiple copies of a target nucleic acid sequence autocatalytically under conditions of substantially constant temperature, ionic strength, and pH in which multiple RNA copies of the target sequence autocatalytically generate additional copies. *See*, *e.g.*, U.S. Pat. Nos. 5,399,491 and 5,824,518, each of which is herein incorporated

IDT Exhibit 1020.4330

by reference in its entirety. In a variation described in U.S. Publ. No. 20060046265 (herein incorporated by reference in its entirety), TMA optionally incorporates the use of blocking moieties, terminating moieties, and other modifying moieties to improve TMA process sensitivity and accuracy.

The ligase chain reaction (Weiss, R., *Science* 254: 1292 (1991), herein incorporated by reference in its entirety), commonly referred to as LCR, uses two sets of complementary DNA oligonucleotides that hybridize to adjacent regions of the target nucleic acid. The DNA oligonucleotides are covalently linked by a DNA ligase in repeated cycles of thermal denaturation, hybridization and ligation to produce a detectable double-stranded ligated oligonucleotide product.

Strand displacement amplification (Walker, G. et al., *Proc. Natl. Acad. Sci. USA* 89: 392-396 (1992); U.S. Pat. Nos. 5,270,184 and 5,455,166, each of which is herein incorporated by reference in its entirety), commonly referred to as SDA, uses cycles of annealing pairs of primer sequences to opposite strands of a target sequence, primer extension in the presence of a dNTPαS to produce a duplex hemiphosphorothioated primer extension product, endonuclease-mediated nicking of a hemimodified restriction endonuclease recognition site, and polymerase-mediated primer extension from the 3' end of the nick to displace an existing strand and produce a strand for the next round of primer annealing, nicking and strand displacement, resulting in geometric amplification of product. Thermophilic SDA (tSDA) uses thermophilic endonucleases and polymerases at higher temperatures in essentially the same method (EP Pat. No. 0 684 315).

Other amplification methods include, for example: nucleic acid sequence based amplification (U.S. Pat. No. 5,130,238, herein incorporated by reference in its entirety), commonly referred to as NASBA; one that uses an RNA replicase to amplify the probe molecule itself (Lizardi et al., *BioTechnol.* 6: 1197 (1988), herein incorporated by reference in its entirety), commonly referred to as Qβ replicase; a transcription based amplification method (Kwoh et al., *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)); and, self-sustained sequence replication (Guatelli et al., *Proc. Natl. Acad. Sci. USA* 87: 1874 (1990), each of which is herein incorporated by reference in its entirety). For further discussion of known amplification methods *see* Persing, David H., "In Vitro Nucleic

IDT Exhibit 1020.4331

Acid Amplification Techniques" in *Diagnostic Medical Microbiology: Principles and Applications* (Persing et al., Eds.), pp. 51-87 (American Society for Microbiology, Washington, DC (1993)).

Non-amplified or amplified target nucleic acids can be detected by any conventional means. For example, target mRNA can be detected by hybridization with a detectably labeled probe and measurement of the resulting hybrids. Illustrative non-limiting examples of detection methods are described below.

One illustrative detection method, the Hybridization Protection Assay (HPA) involves hybridizing a chemiluminescent oligonucleotide probe (*e.g.*, an acridinium ester-labeled (AE) probe) to the target sequence, selectively hydrolyzing the chemiluminescent label present on unhybridized probe, and measuring the chemiluminescence produced from the remaining probe in a luminometer. *See*, *e.g.*, U.S. Pat. No. 5,283,174 and Norman C. Nelson et al., Nonisotopic Probing, Blotting, and Sequencing, ch. 17 (Larry J. Kricka ed., 2d ed. 1995, each of which is herein incorporated by reference in its entirety).

Another illustrative detection method provides for quantitative evaluation of the amplification process in real-time. Evaluation of an amplification process in "real-time" involves determining the amount of amplicon in the reaction mixture either continuously or periodically during the amplification reaction, and using the determined values to calculate the amount of target sequence initially present in the sample. A variety of methods for determining the amount of initial target sequence present in a sample based on real-time amplification are well known in the art. These include methods disclosed in U.S. Pat. Nos. 6,303,305 and 6,541,205, each of which is herein incorporated by reference in its entirety. Another method for determining the quantity of target sequence initially present in a sample, but which is not based on a real-time amplification, is disclosed in U.S. Pat. No. 5,710,029, herein incorporated by reference in its entirety.

Amplification products may be detected in real-time through the use of various self-hybridizing probes, most of which have a stem-loop structure. Such self-hybridizing probes are labeled so that they emit differently detectable signals, depending on whether the probes are in a self-hybridized state or an altered state through hybridization to a target sequence. By way of non-limiting example, "molecular torches" are a type of self-hybridizing probe that includes distinct regions of self-complementarity (referred to as

IDT Exhibit 1020.4332

"the target binding domain" and "the target closing domain") which are connected by a joining region (*e.g.*, non-nucleotide linker) and which hybridize to each other under predetermined hybridization assay conditions. In a preferred embodiment, molecular torches contain single-stranded base regions in the target binding domain that are from 1 to about 20 bases in length and are accessible for hybridization to a target sequence present in an amplification reaction under strand displacement conditions. Under strand displacement conditions, hybridization of the two complementary regions, which may be fully or partially complementary, of the molecular torch is favored, except in the presence of the target sequence, which will bind to the single-stranded region present in the target binding domain and displace all or a portion of the target closing domain. The target binding domain and the target closing domain of a molecular torch include a detectable label or a pair of interacting labels (*e.g.*, luminescent/quencher) positioned so that a different signal is produced when the molecular torch is self-hybridized than when the molecular torch is hybridized to the target sequence, thereby permitting detection of probe:target duplexes in a test sample in the presence of unhybridized molecular torches. Molecular torches and many types of interacting label pairs are known (e.g., U.S. Pat. No. 6,534,274, herein incorporated by reference in its entirety).

Another example of a detection probe having self-complementarity is a "molecular beacon" (see U.S. Pat. Nos. 5,925,517 and 6,150,097, herein incorporated by reference in entirety). Molecular beacons include nucleic acid molecules having a target complementary sequence, an affinity pair (or nucleic acid arms) holding the probe in a closed conformation in the absence of a target sequence present in an amplification reaction, and a label pair that interacts when the probe is in a closed conformation. Hybridization of the target sequence and the target complementary sequence separates the members of the affinity pair, thereby shifting the probe to an open conformation. The shift to the open conformation is detectable due to reduced interaction of the label pair, which may be, for example, a fluorophore and a quencher (*e.g.*, DABCYL and EDANS).

Other self-hybridizing probes are well known to those of ordinary skill in the art. By way of non-limiting example, probe binding pairs having interacting labels (e.g., *see* U.S. Pat. No. 5,928,862, herein incorporated by reference in its entirety) may be adapted for use in the compositions and methods disclosed herein. Probe systems used to detect

IDT Exhibit 1020.4333

single nucleotide polymorphisms (SNPs) might also be used. Additional detection systems include "molecular switches," (e.g., *see* U.S. Publ. No. 20050042638, herein incorporated by reference in its entirety). Other probes, such as those comprising intercalating dyes and/or fluorochromes, are also useful for detection of amplification products in the methods disclosed herein (e.g., *see* U.S. Pat. No. 5,814,447, herein incorporated by reference in its entirety).

In some embodiments, detection methods are qualitative (e.g., presence or absence of a particular nucleic acid). In other embodiments, they are quantitative (e.g., viral load).

iii. Protein Detection

Examples of protein detection methods include, but are not limited to, enzyme assays, direct visualization, and immunoassays. In some embodiments, immunoassays utilize antibodies to a purified protein. Such antibodies may be polyclonal or monoclonal, chimeric, humanized, single chain or Fab fragments, which may be labeled or unlabeled, all of which may be produced by using well known procedures and standard laboratory practices. *See*, *e.g.*, Burns, ed., *Immunochemical Protocols*, 3rd ed., Humana Press (2005); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory (1988); Kozbor et al., *Immunology Today* 4: 72 (1983); Köhler and Milstein, *Nature* 256: 495 (1975). In some embodiments, commercially available antibodies are utilized.

D. Data Analysis

In some embodiments, following purification and detection, a computer-based analysis program is used to translate the raw data generated by the detection assay (*e.g.*, the presence, absence, or amount of a given target molecule) into data of predictive value for a clinician or researcher. In some embodiments, the software program is integrated into an automated device. In other embodiments, it is remotely located. The clinician can access the data using any suitable means. Thus, in some preferred embodiments, the present invention provides the further benefit that the clinician, who is not likely to be trained in genetics or molecular biology, need not understand the raw data. The data is

IDT Exhibit 1020.4334

presented directly to the clinician in its most useful form. The clinician is then able to immediately utilize the information in order to optimize the care of the subject.

Any method may be used that is capable of receiving, processing, and transmitting the information to and from laboratories conducting the assays, information provides, medical personal, and subjects. For example, in some embodiments of the present invention, a sample (*e.g.*, a biopsy or a serum or urine sample) is obtained from a subject and submitted to a service (*e.g.*, clinical lab at a medical facility, genomic profiling business, etc.), located in any part of the world (*e.g.*, in a country different than the country where the subject resides or where the information is ultimately used) to generate raw data. Where the sample comprises a tissue or other biological sample, the subject may visit a medical center to have the sample obtained and sent to the profiling center, or subjects may collect the sample themselves (*e.g.*, a urine sample) and directly send it to a profiling center. Where the sample comprises previously determined biological information, the information may be directly sent to the profiling service by the subject (*e.g.*, an information card containing the information may be scanned by a computer and the data transmitted to a computer of the profiling center using an electronic communication systems). Once received by the profiling service, the sample is processed and a profile is produced (*i.e.*, expression data), specific for the diagnostic or prognostic information desired for the subject.

The profile data is then prepared in a format suitable for interpretation by a treating clinician. For example, rather than providing raw data, the prepared format may represent a diagnosis or risk assessment (*e.g.*, viral load levels) for the subject, along with recommendations for particular treatment options. The data may be displayed to the clinician by any suitable method. For example, in some embodiments, the profiling service generates a report that can be printed for the clinician (*e.g.*, at the point of care) or displayed to the clinician on a computer monitor.

In some embodiments, the information is first analyzed at the point of care or at a regional facility. The raw data is then sent to a central processing facility for further analysis and/or to convert the raw data to information useful for a clinician or patient. The central processing facility provides the advantage of privacy (all data is stored in a central facility with uniform security protocols), speed, and uniformity of data analysis.

IDT Exhibit 1020.4335

The central processing facility can then control the fate of the data following treatment of the subject. For example, using an electronic communication system, the central facility can provide data to the clinician, the subject, or researchers.

In some embodiments, the subject is able to directly access the data using the electronic communication system. The subject may chose further intervention or counseling based on the results. In some embodiments, the data is used for research use. For example, the data may be used to further optimize the inclusion or elimination of markers as useful indicators of a particular condition or stage of disease.

E. Compositions & Kits

In some embodiments, systems and/or devices of the present invention are shipped containing all components necessary to perform purification and analysis (e.g., blister seals reagents and cartridges for performing assays). In other embodiments, additional reaction components are supplied in separate vessels packaged together into a kit.

Any of these compositions, alone or in combination with other compositions disclosed herein or well known in the art, may be provided in the form of a kit. Kits may further comprise appropriate controls and/or detection reagents. Any one or more reagents that find use in any of the methods described herein may be provided in the kit.

EXPERIMENTAL

The following examples are provided to demonstrate and illustrate certain preferred embodiments and aspects of the compositions and methods disclosed herein, but are not to be construed as limiting the scope of the claimed invention.

Example 1

Bursting force

Experiments were conducted to ascertain how much force is required to burst open blisters. This information is used to aid in determining the correct type of stepper motor (which outputs sufficient force to depress the plunger against the blister). An

IDT Exhibit 1020.4336

Instron compression instrument was used to blindly burst open the blisters (i.e., no tear-drop clamps). The blister parameters are described below.

- Lysis blisters
  - 0.72" diameter
  - 0.20" stroke (depth)
  - Liquid volume = 500 μL
- Oil blisters
  - 0.72" diameter
  - 0.18" stroke (depth)
  - Liquid volume = 400 μL
- Elution blisters
  - 0.55" diameter
  - 0.150" stroke (depth)
  - Liquid volume = 150 μL

The graphical data is shown in Figure 18. The data plot shows the average load (force) required to burst open the elution, oil, and lysis blisters. The relatively consistent forces, as seen by the relatively narrow standard deviations, indicate that the heat sealing quality is consistent. The forces will change if either the blister size, seal width or liquid volume inside the blister is changed.

Example 2

Accelerating aging experiments

Experiments were conducted to quantify the heat sealing process which bonds the two foil laminates together. The quality of the heat seal has a direct impact on both the force required to burst open a blister and any liquid loss via evaporation. The blisters were stored in a forced convective oven at 42-45 °C (1-3% RH) for several weeks. Furthermore, to simulate cold shipment transportation, the blisters were exposed from room temperature (RT) to 0 °C for 16 hours, 0 °C to -20 °C for 8 hours, -20 °C to 0 °C for 16 hours, and back to RT. The liquid loss was measured through periodic weight

IDT Exhibit 1020.4337

measurements. Design of Experiments (DOE) were performed with each liquid reagent (elution, lysis, and oil) to determine the optimal time and temperature regime. It was previously determined that pressure, in the range of 50-90 psi has little or no impact on the quality of the heat seal. Blisters with no liquids were also heat sealed to determine any physical change in the blister material itself that would lead to a change in weight. This serves as the baseline for weight loss observed in blisters with liquids inside.

Lysis DOE

- Time = 2, 5, 8 s
- Temperature = 191, 211, 232 °C

Elution DOE

- Time = 2, 5 s
- Temperature = 191, 211 °C

Oil DOE

- Time = 2, 5 s
- Temperature = 191, 211 °C

While no evaporation is expected for oil blisters, the oil may have wicked onto the heat seal surface, compromising the quality of the heat seal. Also, it is desirable to determine a universal time and temperature that would work for all three blisters and liquids since it would be extremely helpful in large-scale manufacturing. The weight loss data showed the following observations:

Lysis blisters

- 49 days
- Empty blister average weight loss and standard deviation = 0.0006 g ± 0.0001
- Liquid blister weight loss varied between 0.0005 – 0.0013 g

IDT Exhibit 1020.4338

- Observation – taking the empty blister weight loss into consideration, the actual liquid loss was approximately 0.0006 g at most, which corresponds to 0.6 μL and indicates very good heat seals

Elution blisters

- 36 days
- Empty blister average weight loss and standard deviation = 0.0002 g ± 0.0001
- Liquid blister weight loss varied between 0.0001 – 0.0003 g
- Observation – taking the empty blister weight loss into consideration, the actual liquid loss is inconsequential, which indicates very good heat seals

Oil blisters

- 20 days
- Empty blister average weight loss and standard deviation = 0.0001 g ± 0.0001
- Liquid blister weight loss varied between 0.0000 – 0.0005 g
- Observation – taking the empty blister weight loss into consideration, the actual liquid loss was approximately 0.0003 g at most, which corresponds to 0.38 μL and indicates very good heat seals

Example 3

Liquid volume fill capacity

Experiments were conducted to determine the total liquid volume fill capacity for a given blister. This is dependent on the blister diameter and stroke (depth). This information is used to determine the maximum amount of liquid than can be safely heat sealed inside a blister without overflowing (i.e., onto the heat seal perimeter) and compromising the heat seal process. This is especially useful for liquids that preferentially wet the surface of the foil laminate blister. Two blister diameters were characterized: 0.55" and 0.72". Multiple blisters with varying strokes (starting at the maximum stroke where the foil laminate did not tear and decreasing the depth progressively) were cold formed and heat sealed empty. A fine gauge needle was used to pierce the lidstock foil laminate and dispense liquid inside the empty blister until liquid

IDT Exhibit 1020.4339

started to spill out. This was determined to be the liquid volume fill capacity for a blister. See Figure 19.

The data plot shows the liquid volume fill capacity for both blister diameters at several stroke values. The maximum volume is listed above each data point. A linear trend line was also determined for each blister diameter that can offer additional numerical interpretation at other stroke values.

Example 4

Dead volume

When a blister is crushed completely, a percentage of the liquid will always remain inside the blister due to how the blister is crushed – the creases can trap small amounts of liquids. It is useful to characterize the dead volume since it relates directly to the volume that should be inside the blister (blister liquid volume = channel volume + chamber volume + dead volume + estimated evaporation volume). Furthermore, this will also help determine how much liquid is dispensed into the cartridge and if it is sufficient for the assay.

To determine the dead volume, various blisters were crushed on a characterization cartridge. The liquid was dispensed into one long channel that was previously calibrated to correlate length with liquid volume. Therefore, the dead volume can be defined as below. See Figure 20 for a sketch of the cartridge and concept of determining the dead volume.

Dead volume = Total start volume in blister – volume dispensed into channel

The following types of blisters were tested:

- 0.55" diameter
  - 0.150" stroke
  - 150 μL
- 0.72" diameter
  - 0.135", 0.15", 0.18", and 0.20" stroke
  - 300, 350, 400, 500, 550, 575, and 600 μL

The raw data is shown in

IDT Exhibit 1020.4340

Table **1**. The 'total volume fill capacity' value is adapted from Figure 19. The table shows the average dispensed volume, as well as the respective dead volume and its ratio to both the total volume fill capacity and actual dispensed liquid volume. It indicates that a larger dead volume ensues for smaller blisters, and also provides information on the minimum liquid volume that should be stored in the blister (i.e., equal to the dead volume). It also shows that for the liquid volumes and strokes tested, the dead volume is relatively constant.

Table 1. Raw data showing the dispensed volume for each blister type, and its respective dead volume (left behind in the blister). A ratio of the dead volume to both the total liquid volume fill capacity and actual dispensed volume is also reported here.

Blister type	**Elution**	**Oil**			**Lysis**			
Stroke (in)	**0.15**	**0.135**	**0.15**	**0.18**	**0.20**			
Dispensed vol (μL)	**150**	**300**	**350**	**400**	**500**	**550**	**575**	**600**
Total volume fill capacity (μL)	237	519	567	692	823	823	823	823
Ave vol dispensed (μL)	109.3	215.1	245.4	310.2	392.0	423.3	459.3	484.1
S.D. vol dispensed (μL)	11.2	29.4	25.7	13.2	27.0	22.3	20.0	12.7
Ave dead vol (μL)	40.8	84.9	104.6	89.8	108.0	126.7	115.7	115.9
Dead vol : Total vol fill capacity	0.17	0.16	0.18	0.13	0.13	0.15	0.14	0.14
Dead vol : Dispensed blister vol	0.27	0.28	0.30	0.22	0.22	0.23	0.20	0.19

Example 5

Liquid volume vs. force

With varying liquid volumes inside a given blister, its effect on the force required to burst it open will also change. This is due to the amount of air that is present in the blister. While air can be compressed, liquids cannot, and with higher volumes of air, the more force will be required to burst the peelable heat seals. The amount of air in the blister is preferable reduced as much as possible: (1) during transport at lower ATA (atmospheres absolute) where perhaps, the cabin is not pressurized sufficiently, an increased amount of air in a blister can began to expand and begin to peel the heat seal; (2) reducing the amount of air will help realize consistent and uniform forces to burst open the blisters (i.e., more air means higher standard deviation in forces).

IDT Exhibit 1020.4341

Experiments have been done to date to determine how liquid volumes for a given blister geometry affects the force. Here, the following parameters were tested.

- 0.72" diameter blister
- 0.20" stroke
- 100, 200, 400, and 500 μL liquid volume (this corresponds to 12, 24, 49, and 61% total volume fill capacity – 823 μL)

The resulting data is shown in Figure 21. The data plot shows that the force increases at low liquid volumes because of the higher volume of air. Furthermore, the standard deviation is significantly larger for low volumes, which indicates high variability from blister to blister. This is undesirable in actual experiments. When liquid volumes are 400 μL or larger, the force is reduced to more workable values, and the variability across blisters all but disappears.

Example 6

Additional Designs

This example describes additional blister pack designs. However, since the original submission, we have realized issues with this design. In some embodiments, small channels are created during the bonding process between foil laminate and the transfer adhesive (See e.g., Figure 8). These channels are created due to the step difference between the transfer adhesive and foil laminate (caused by the thickness of the lidstock foil laminate). The creation of these channels cause an increase in the dead volume when a blister is burst since liquid can wick into these locations. It places high demand on ensuring the bonding technique minimizes this channeling. In some embodiments, a modified blister packaging design that minimizes dead volume is utilized.

Once a blister has been cold formed, two alignment holes are punched through the blister foil laminate, as shown in **Figure 2**3. Here, the alignment holes, which pass through the central axis point of the blister, are punched after the cold forming.

IDT Exhibit 1020.4342

However, it is feasible to perform this operation simultaneously with the cold forming operation. The alignment holes are positioned such that they are outside the circular heat seal perimeter (shown in **Figure 2**3(a)). The holes serve to align the blister during the heat sealing process (i.e., align foil laminate #1 with foil laminate #2) and manufacturing/assembly of the overall disposable (e.g., integrating and positioning the packaged reagent blister with the rigid test cartridge).

Heat sealing occurs between foil laminate #1 and foil laminate #2 (designated as 'lidstock'). Three holes are punched into the lidstock prior to heat sealing. See **Figure 24**24. Just as with the foil laminate #1, two holes are punched for alignment (both for heat sealing subsequent integration with the rigid test cartridge). The third hole is a liquid port hole which serves as the exit port for the liquid when the blister is crushed. It is positioned just outside the circular heat seal perimeter.

The cold-formed blister and lidstock are prepared for the heat sealing process, which is briefly outlined in **Figure** 25. Retractable pins are used to position and align the cold-formed blister with the lidstock via the punched alignment holes. This method can subsequently be used as registration marks (alignment) during manufacturing.

The heat seal band used for this application is a donut-shaped heat seal band with extensions on the side, as shown in **Figure 26**26(a). This facilitates a vapor and liquid tight heat seal bond, overlapping with the three punched holes.

While only one blister is demonstrated in Figures 23-26, this method can be extended to the design and manufacturing of multiple blisters where, for example, more than one blister is required for a given test cartridge assay.

Furthermore, this overall design facilitates easy attachment to the test cartridge (e.g., using transfer adhesive) since during the adhesive bonding, the blister does not experience any variation in height. The liquid thru port punch hole and design/geometry of the heat seal band facilitates smooth bonding of the lidstock to the cartridge with no chance of liquid leakages or channeling. See **Figure** 27.

Since both the foil laminate #1 and lidstock extend equally across the packaged blister, there is no step difference when bonding the blister to the rigid assay cartridge (via the double-sided transfer adhesive), preventing the creation of any channels.

IDT Exhibit 1020.4343

The purpose of the mechanical clamp in bursting blisters are: (1) to assist in a directional heat seal peeling process to burst the blister and deliver the liquid, (2) to provide uniform pressure along the circular heat seal perimeter at the edge of the cold-formed blister (minimize the gap between the edge of a cold-formed blister and mechanical clamp collar) such that the peeling only occurs in the direction towards the liquid thru port, and (3) to guide a mechanical plunger which applies force on the blister and eventually peels it open. See **Figure** 28. This minimizes the liquid dead volume (e.g., liquid volume left behind in a completely crushed blister) and further ensures that the blister heat seal peels immediately in the direction of interest. If the gap is not minimized, it can cause the peeling can start occurring in a random direction which increases the liquid volume left behind, reducing the actual volume available for the rigid test cartridge, and/or when actively monitoring the forces required to crush and peel open a blister, it may take multiple instances to successfully direct the peeling towards the liquid thru port.

All publications, patents, patent applications and sequences identified by accession numbers mentioned in the above specification are herein incorporated by reference in their entirety. Although the invention has been described in connection with specific embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Modifications and variations of the described compositions and methods of the invention that do not significantly change the functional features of the compositions and methods described herein are intended to be within the scope of the following claims.

IDT Exhibit 1020.4344

CLAIMS

We claim:

1. An assay system, comprising:

a) a liquid packaging component comprising one or more liquid storage compartments, wherein each of said liquid storage compartments comprises a liquid and is covered with a burstable seal;

b) a seal bursting component configured to burst said burstable seal; and

c) an assay device configured to accept liquids from said liquid storage compartments.

2. The system of claim 1, wherein said burstable seal is a foil laminate.

3. The system of claim 2, wherein said foil is aluminum foil.

4. The system of claim 1, wherein said seal bursting component comprises plungers that compress said liquid storage compartment under conditions such that said seal is peeled open.

5. The system of claim 4, wherein said plungers are driven by one or more motors.

6. The system of claim 4, wherein said plungers are driven manually.

7. The system of claim 1, wherein said liquid packaging component and chambers within said assay device are connected via fluid conduits.

8. The system of claim 1, wherein said liquid packaging component and chambers within said assay device are in direct contact.

IDT Exhibit 1020.4345

9. The system of claim 1, wherein said liquid packaging component comprises one or more liquid storage compartments.

10. The system of claim 1, wherein said liquid storage compartments comprise less than 50% air by volume.

11. The system of claim 1, wherein said liquid storage compartments comprise less than 400 μl air.

12. The system of claim 1, wherein said liquid storage compartment comprise a tear-drop clamp that applies uniform pressure across a portion of the burstable seal perimeter.

13. The system of claim 12, wherein said tear drop clamp does not apply pressure to the portion of said burstable seal that is in communication with said assay device.

14. The system of claim 1, wherein said liquid storage compartments comprise reagents for performing an assay.

15. The system of claim 14, wherein said assay is selected from the group consisting of a diagnostic assay and a research assay.

16. The system of claim 1, wherein said liquid packaging component further comprises one or more alignment pins to secure said liquid storage compartment to said liquid packaging component.

17. An assay method, comprising:

contacting a liquid packaging component comprising one or more liquid storage compartments, wherein each of said liquid storage compartments comprises a

IDT Exhibit 1020.4346

liquid and is covered with a burstable seal with a a seal bursting component configured to burst said burstable seal under conditions such that said liquids are transported to an assay device configured to accept liquids from said liquid storage compartments.

18. The method of claim 17, wherein said burstable seal is a foil laminate.

19. The method of claim 18, wherein said foil is aluminum foil.

20. The method of claim 17, wherein said seal bursting component comprises plungers that compress said liquid storage compartment under conditions such that said seals are burst.

21. The method of claim 20, wherein said plungers are driven by one or more motors.

22. The method of claim 17, wherein said plungers are manually driven.

23. The method of claim 17, wherein said liquid packaging component and chambers within said assay device are connected via fluid conduits.

24. The method of claim 17, wherein said liquid packaging component and chambers within said assay device are in direct contact.

25. The method of claim 17, wherein said liquid packaging component comprises one or more liquid storage compartments.

26. The method of claim 17, wherein said liquid storage compartments comprise less than 50% air by volume.

27. The method of claim 17, wherein said liquid storage compartments comprise less than 60% air by volume.

IDT Exhibit 1020.4347

28. The method of claim 17, wherein said liquid storage compartments comprise less than 400 μl air.

29. The method of claim 17, wherein said liquid storage compartment comprise a tear-drop clamp that applies uniform pressure across a portion of the burstable seal perimeter.

30. The method of claim 29, wherein said tear drop clamp does not apply pressure to the portion of said burstable seal that is in communication with said assay device.

31. The method of claim 17, wherein said liquid storage compartments comprise reagents for performing an assay.

32. The method of claim 17, wherein said assay is selected from the group consisting of a diagnostic assay and a research assay.

33. The method of claim 17, wherein said liquid packaging component further comprises one or more alignment pins to secure said liquid storage compartment to said liquid packaging component.

IDT Exhibit 1020.4348

Figure 1

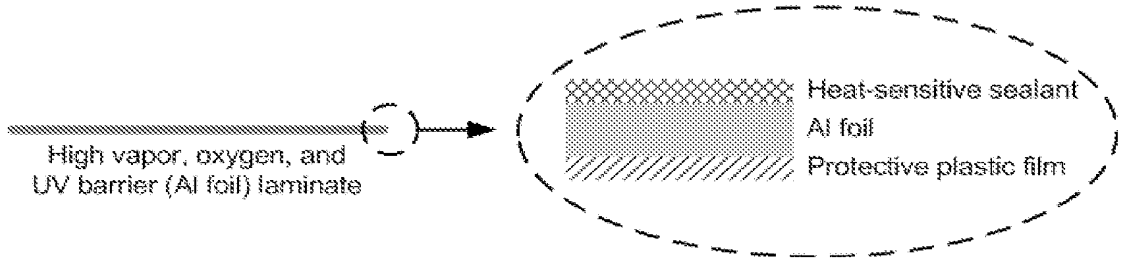
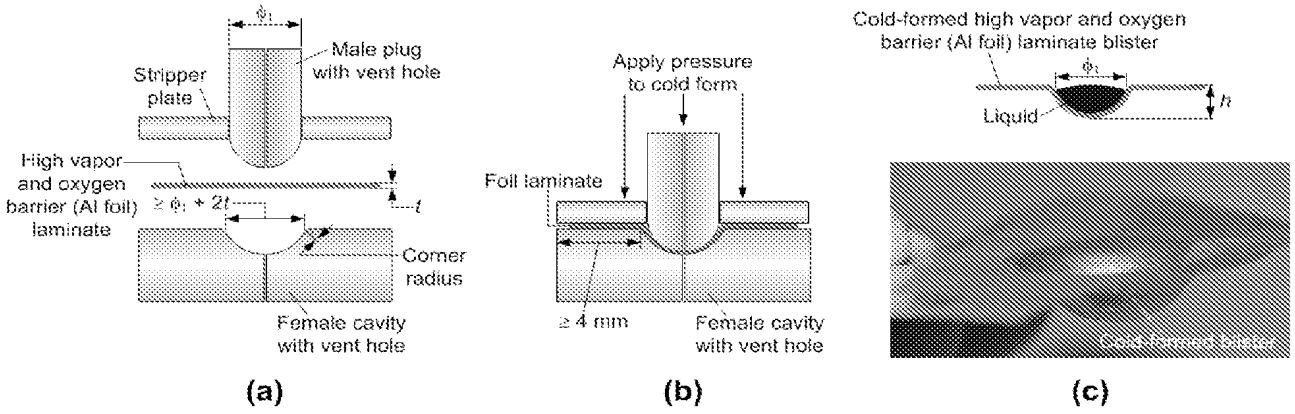


Figure 2



IDT Exhibit 1020.4349

Figure 3

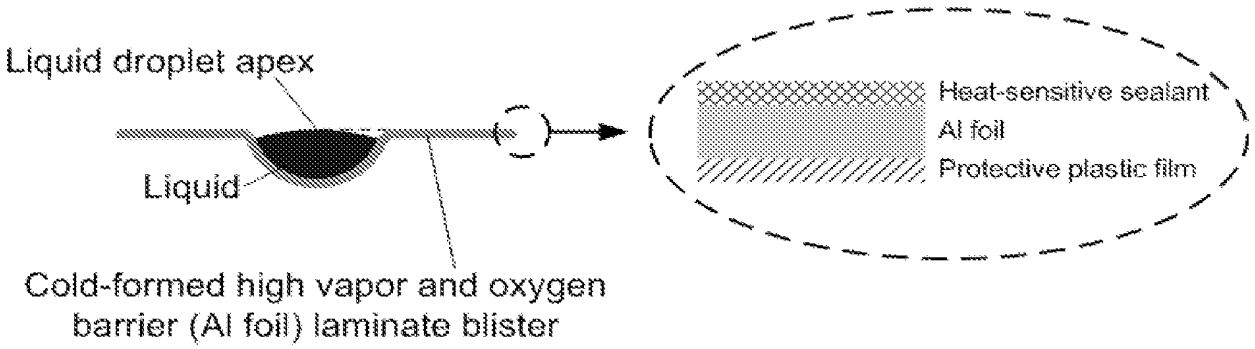
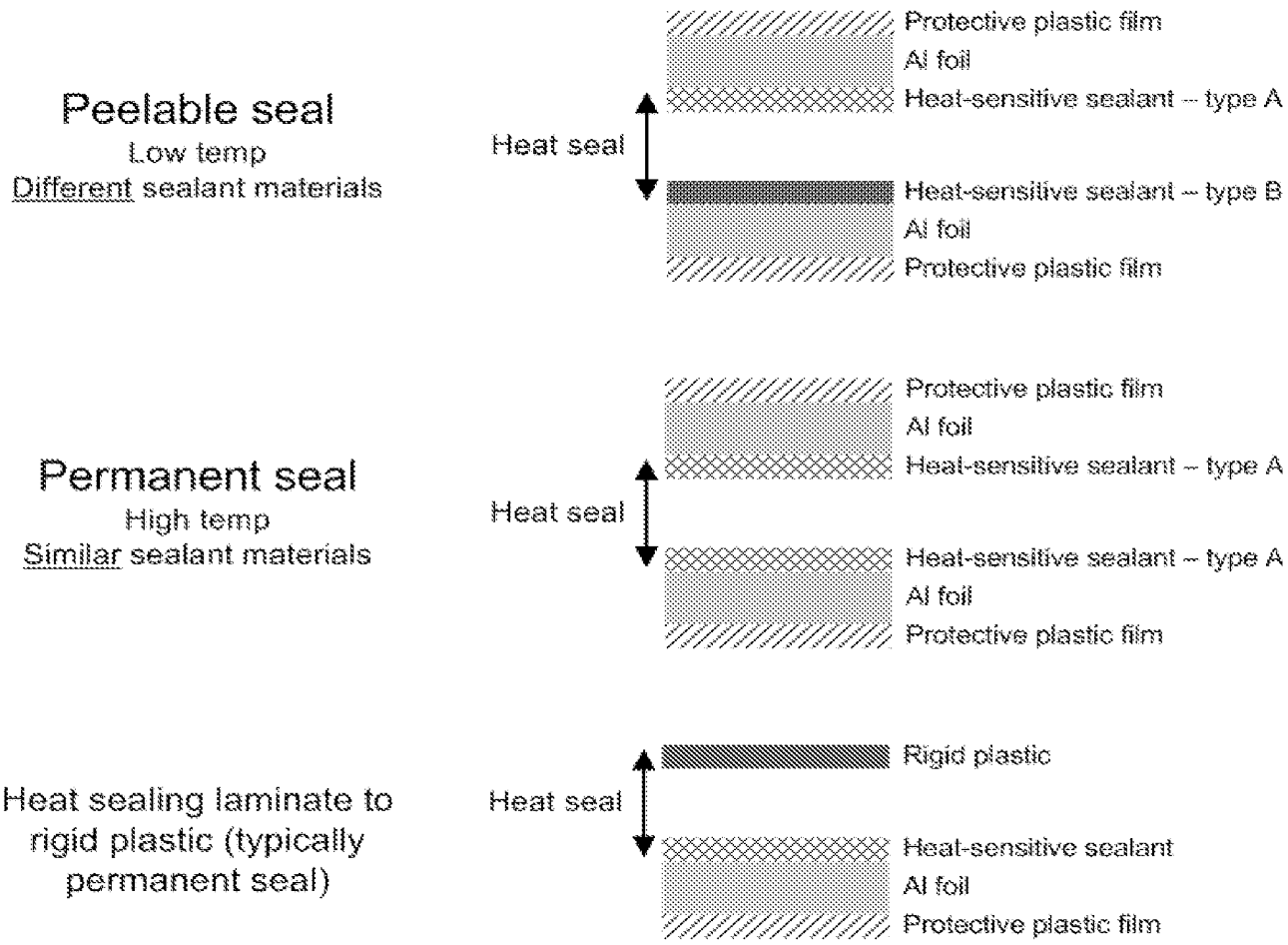
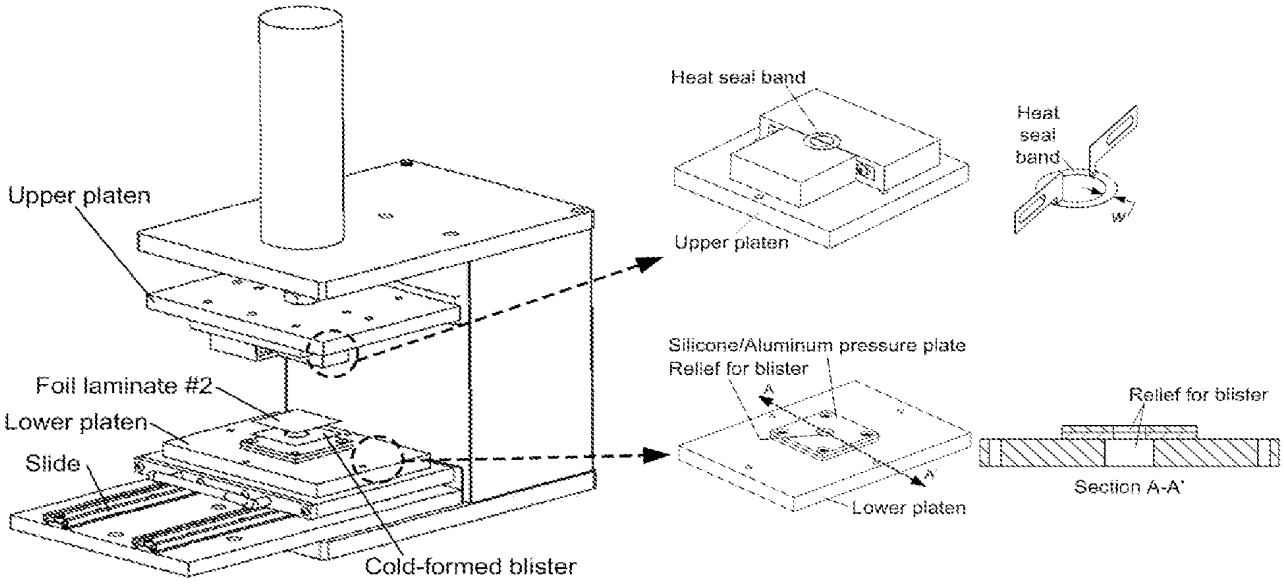


Figure 4



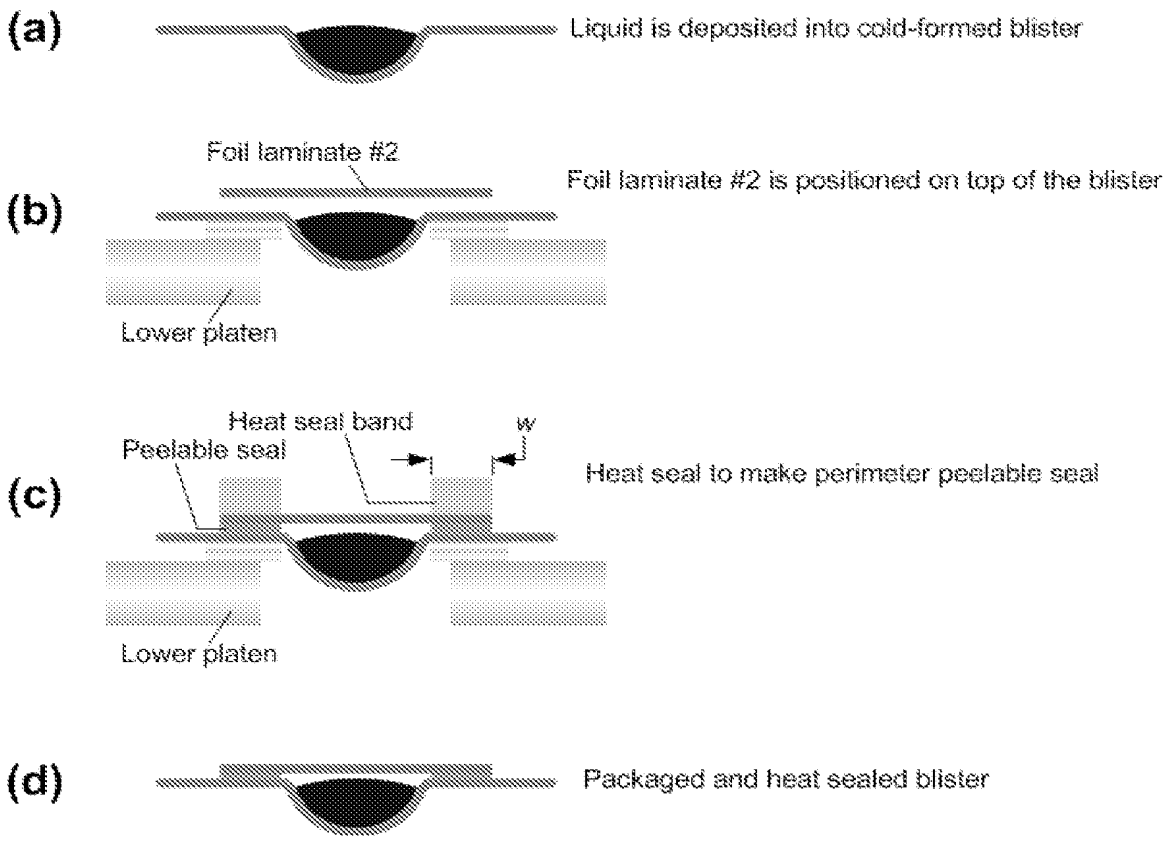
IDT Exhibit 1020.4350

Figure 5



IDT Exhibit 1020.4351

Figure 6



IDT Exhibit 1020.4352

Figure 7

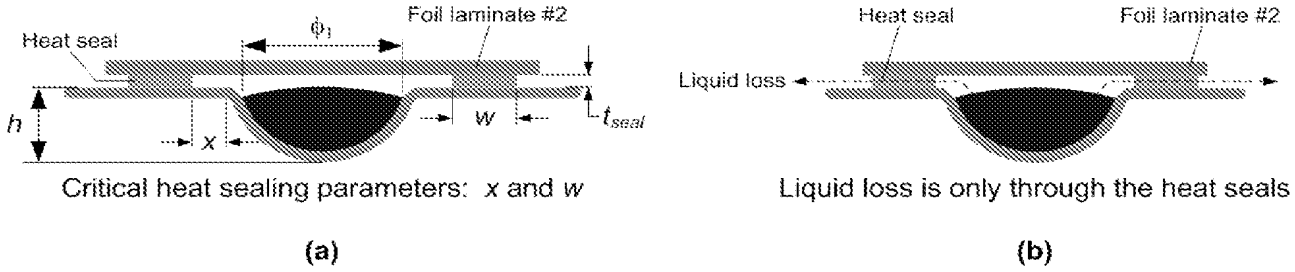
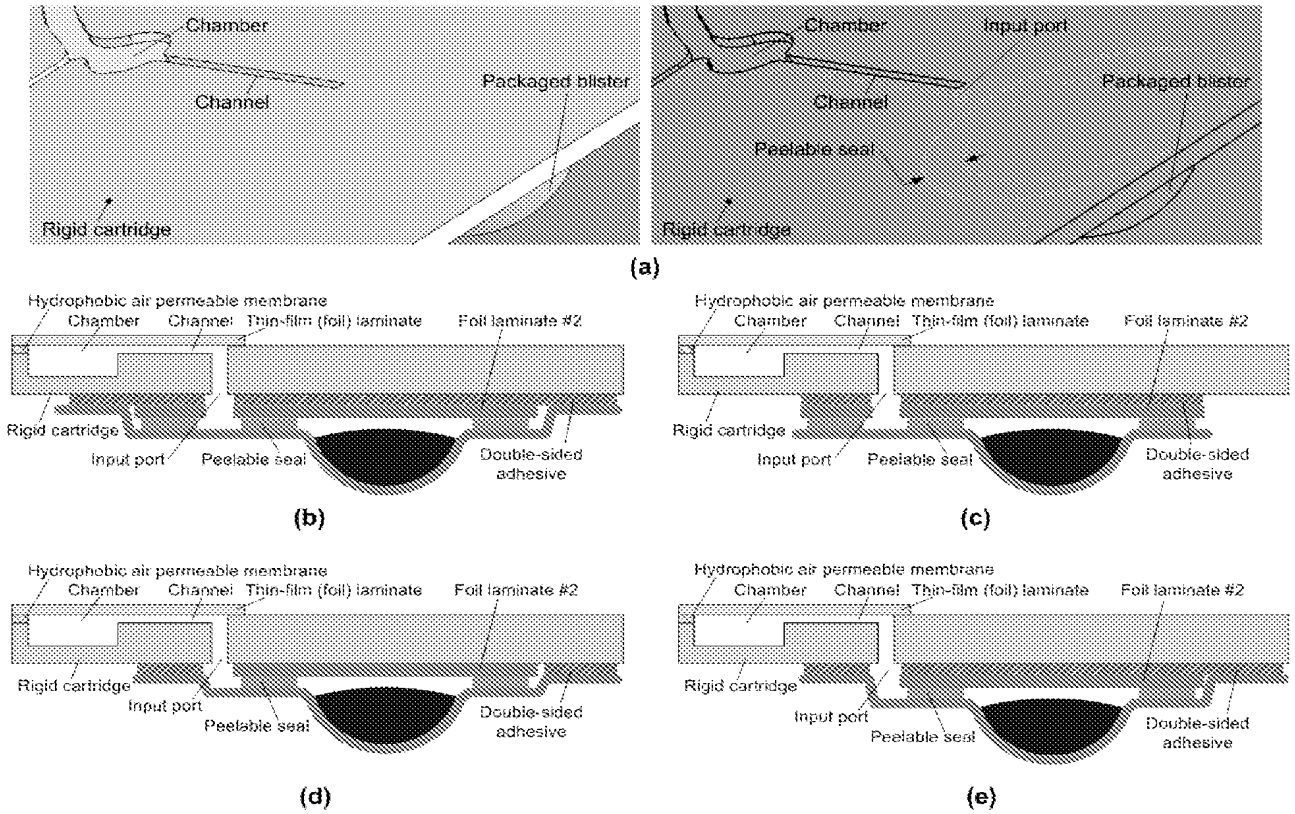


Figure 8



IDT Exhibit 1020.4353

Figure 9

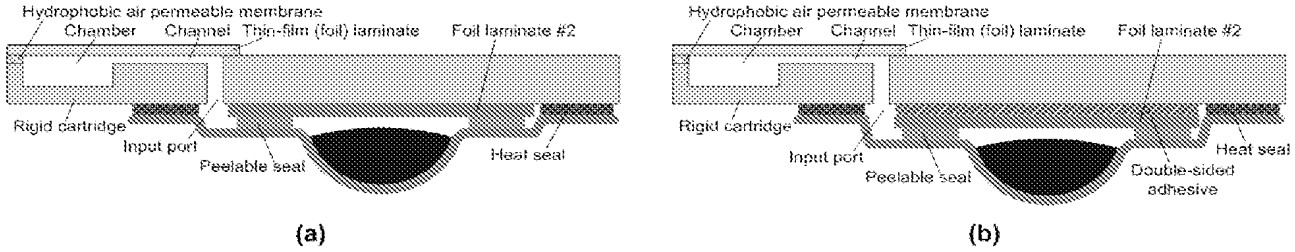
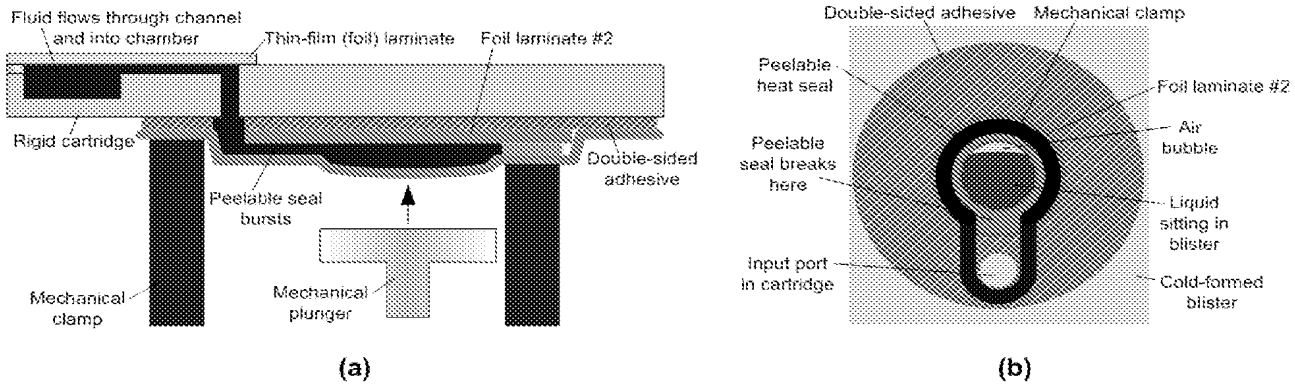
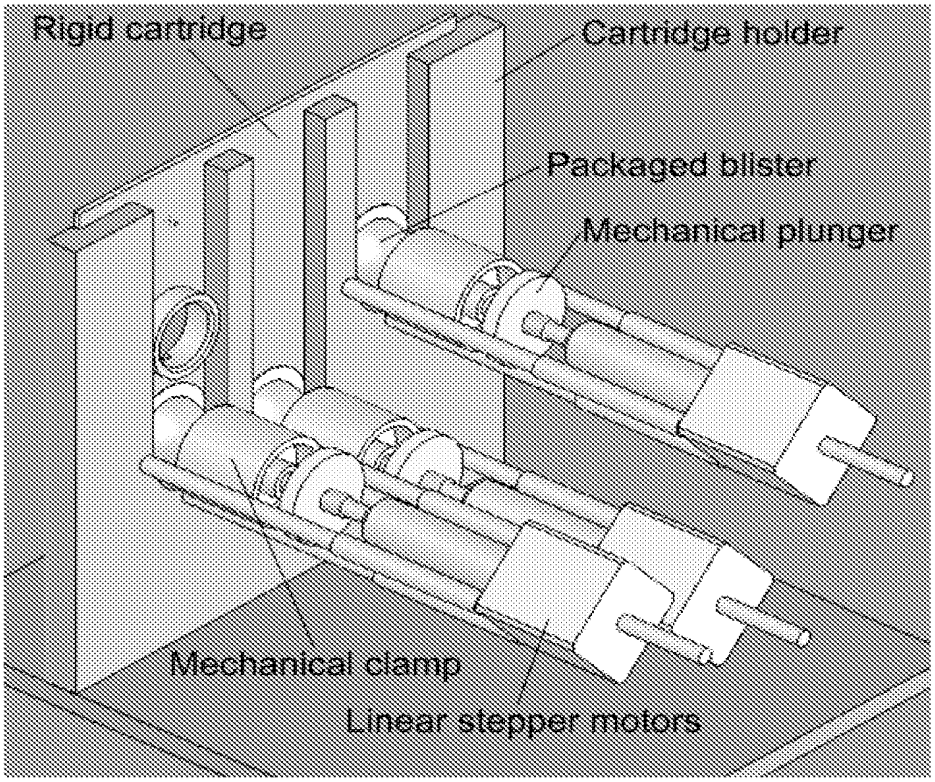


Figure 10



IDT Exhibit 1020.4354

Figure 11



IDT Exhibit 1020.4355

Figure 12

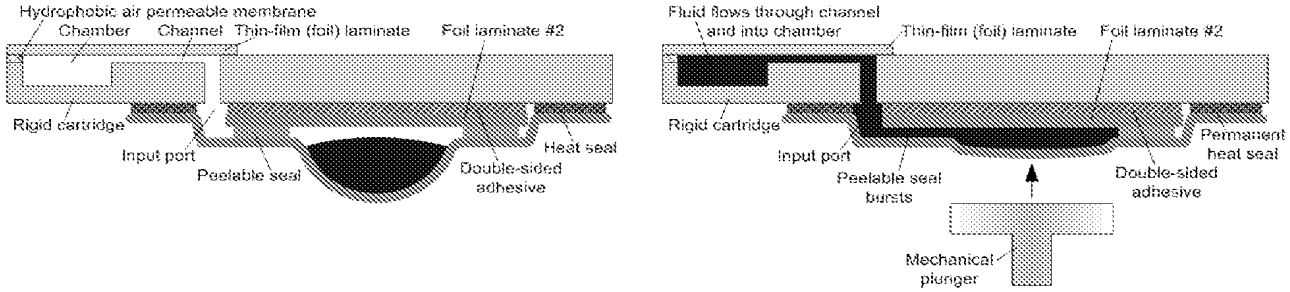
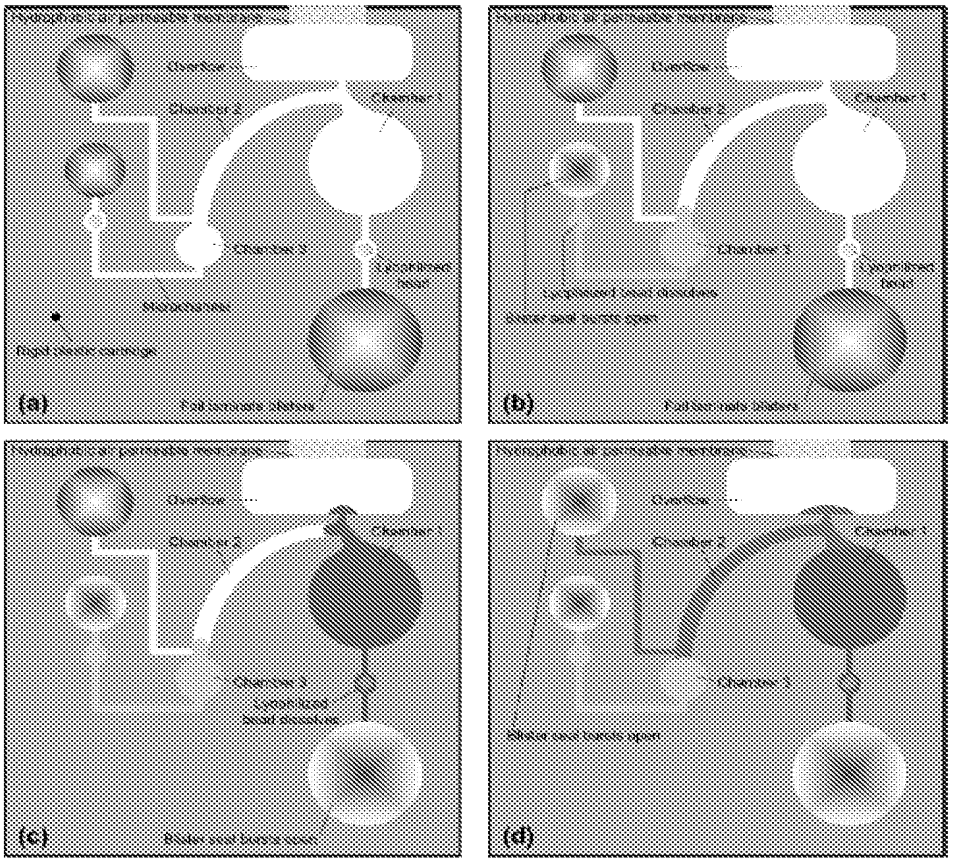


Figure 13



IDT Exhibit 1020.4356

Figure 14

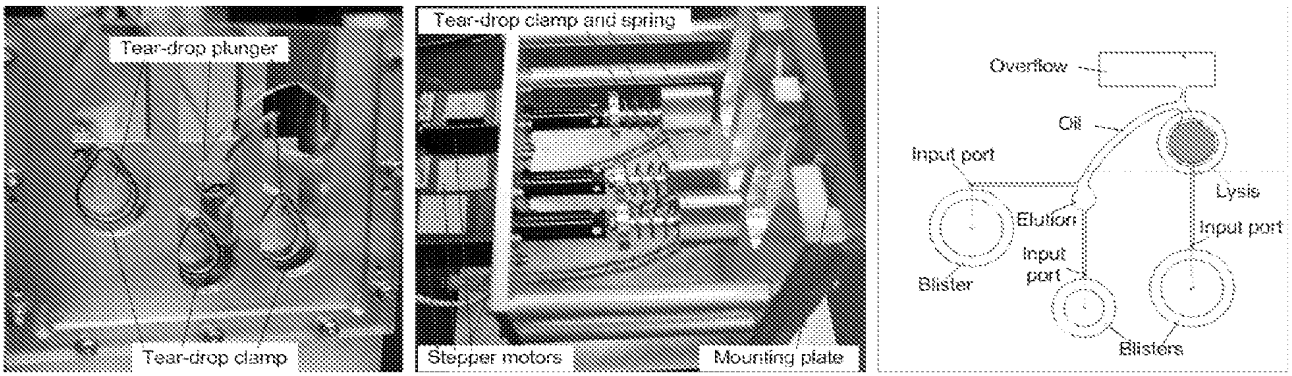
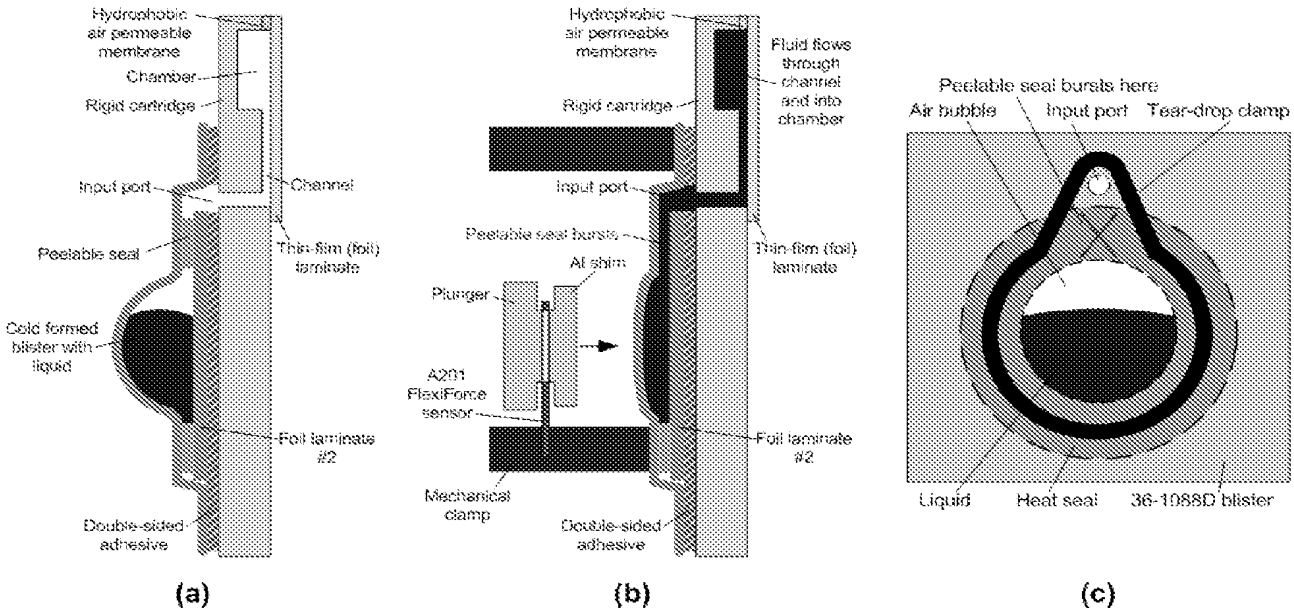


Figure 15



IDT Exhibit 1020.4357

Figure 16

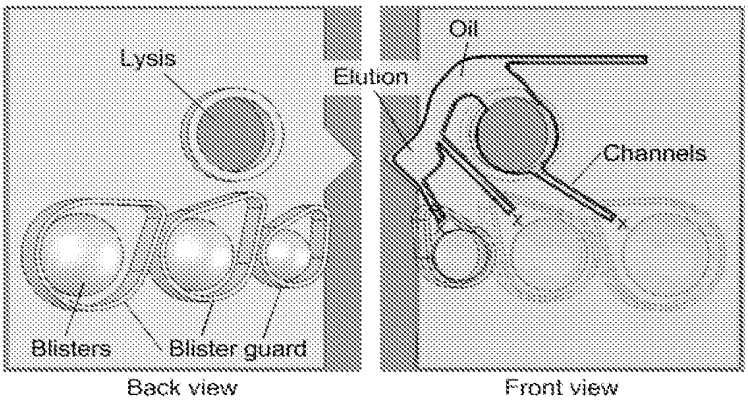
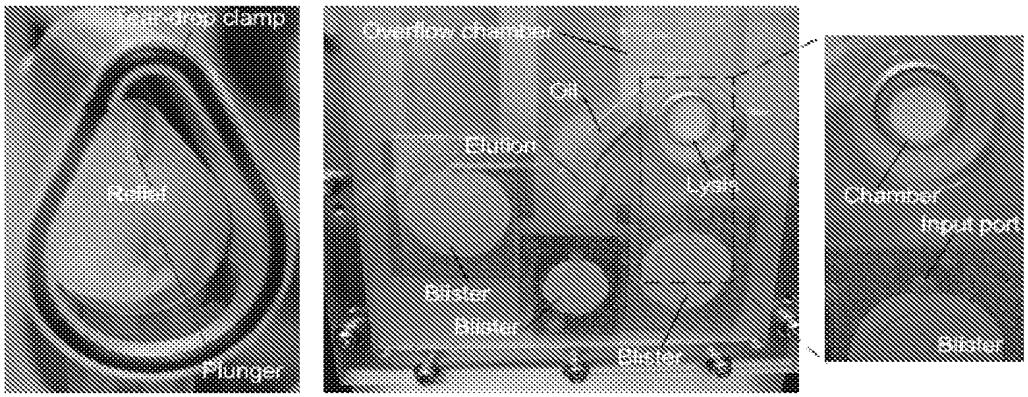


Figure 17



IDT Exhibit 1020.4358

Figure 18

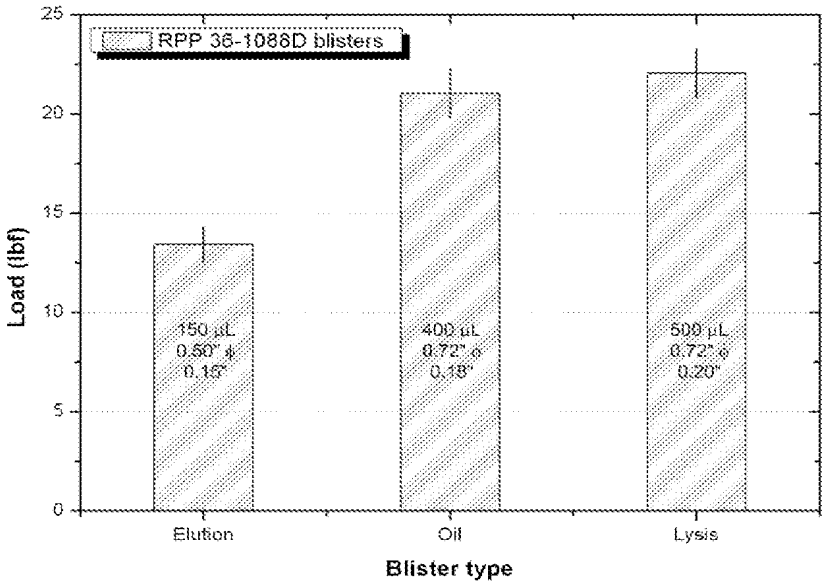
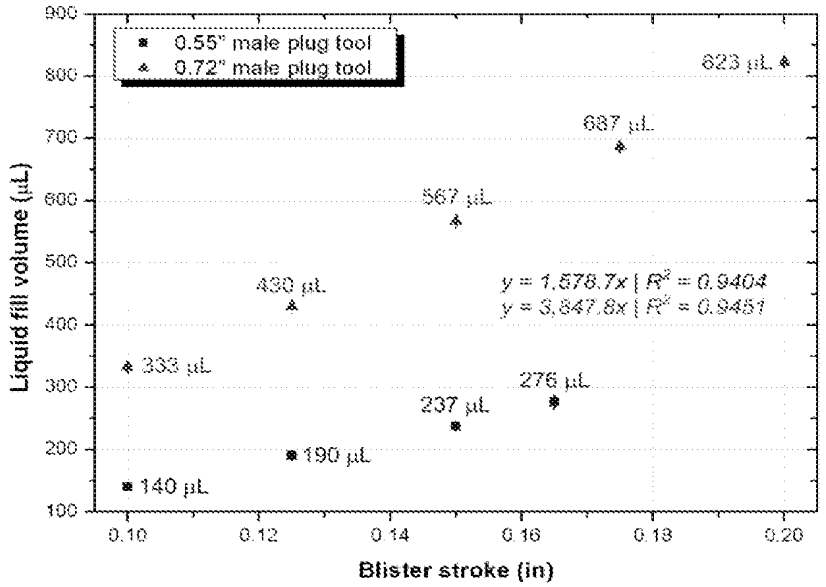


Figure 19



IDT Exhibit 1020.4359

Figure 20

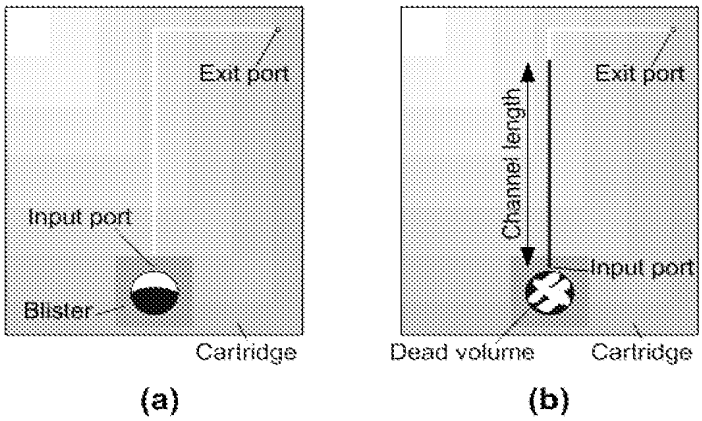


Figure 21

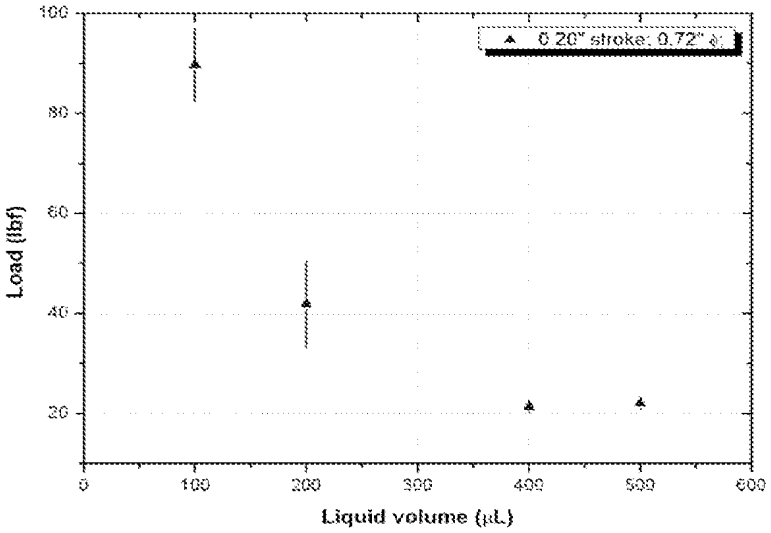


Figure 22

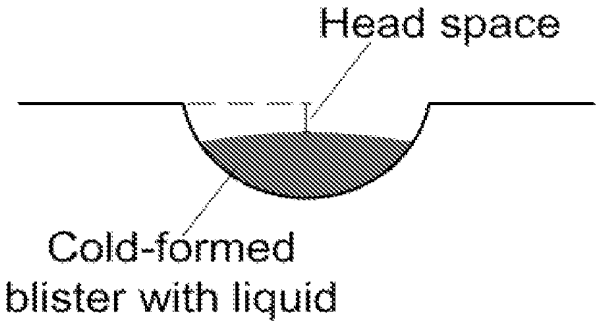
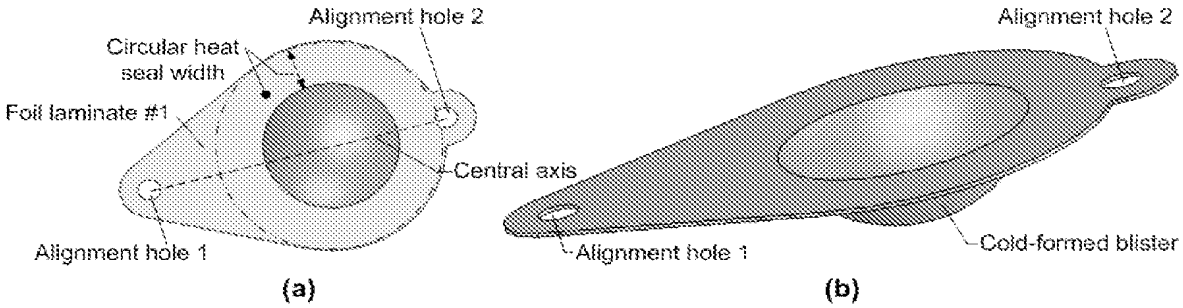


Figure 23



IDT Exhibit 1020.4361

Figure 24

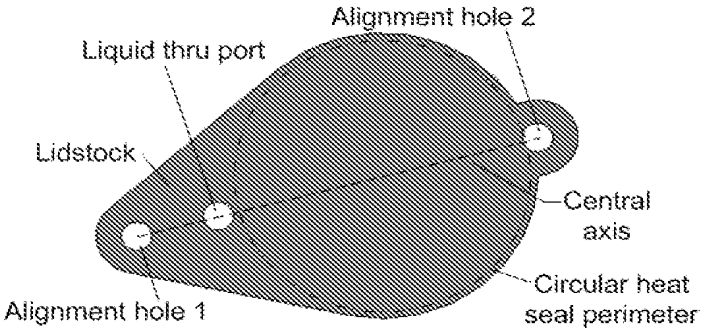
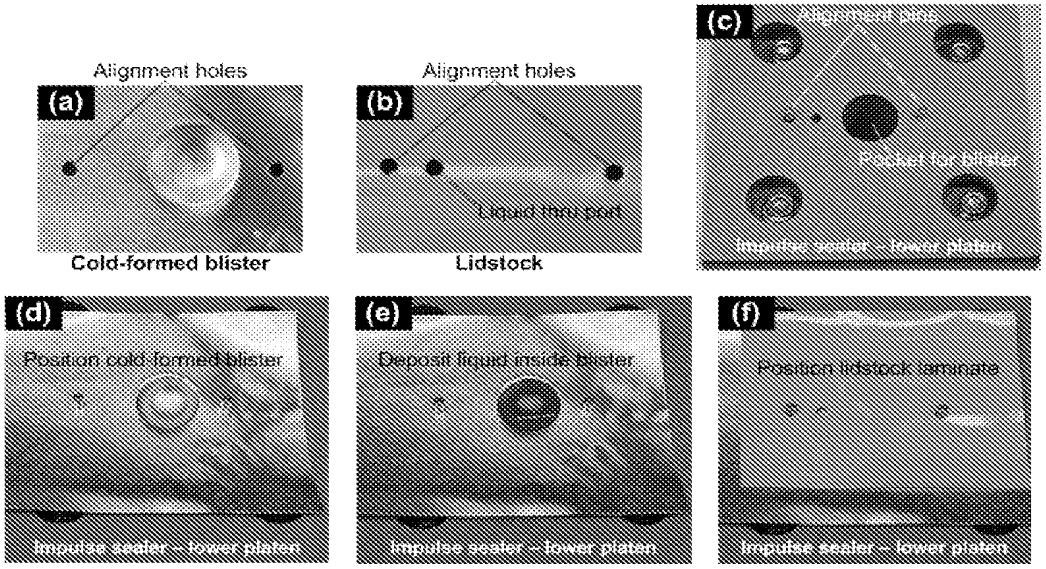


Figure 25



IDT Exhibit 1020.4362

Figure 26

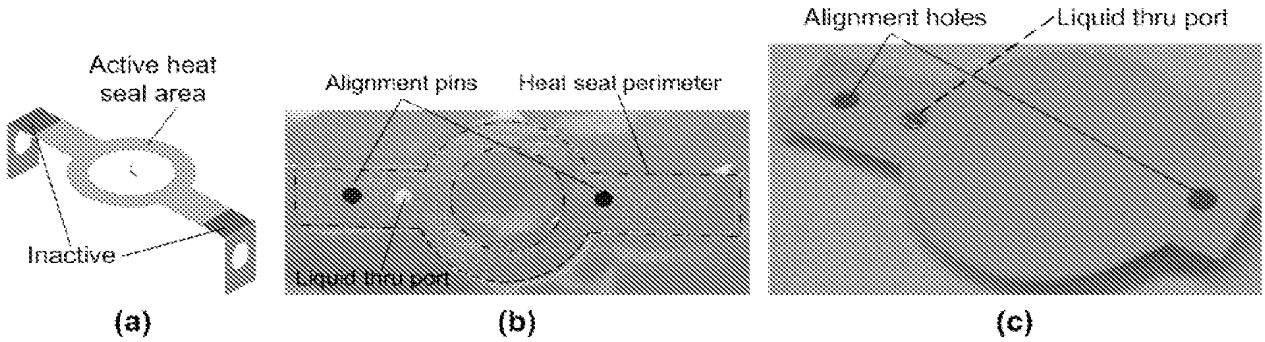
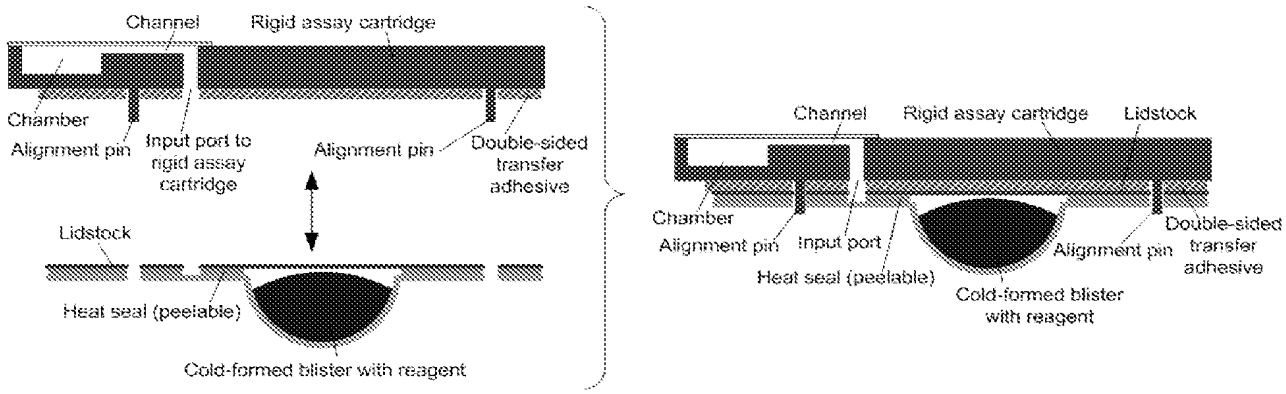
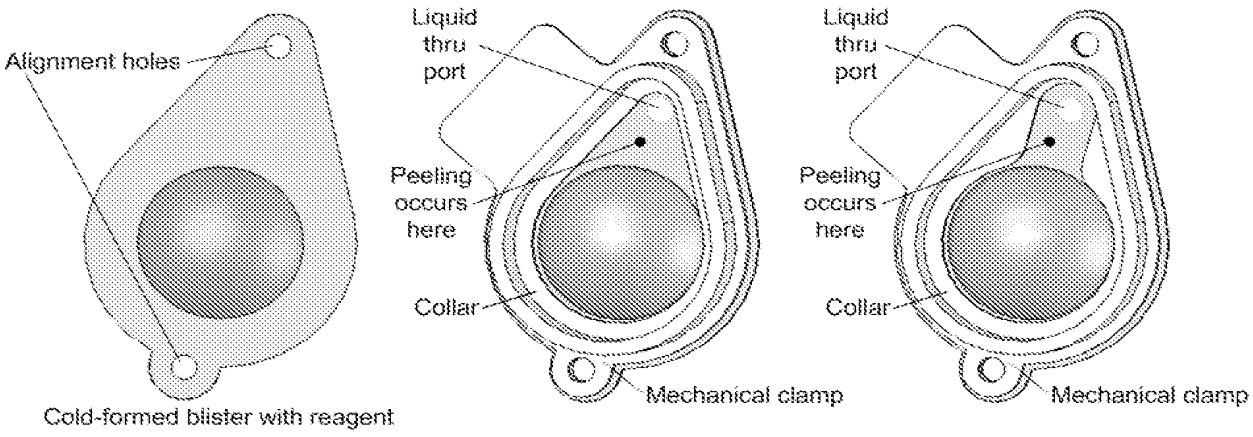


Figure 27



IDT Exhibit 1020.4363

Figure 28



IDT Exhibit 1020.4364

Espacenet

Bibliographic data: WO2013112923 (A1) — 2013-08-01

COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION

Inventor(s): AMORESE DOUG [US]; ARMOUR CHRIS [US]; KURN NURITH [US] ± (AMORESE, DOUG, ; ARMOUR, CHRIS, ; KURN, NURITH)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; AMORESE DOUG [US]; ARMOUR CHRIS [US]; KURN NURITH [US] ± (NUGEN TECHNOLOGIES, INC, ; AMORESE, DOUG, ; ARMOUR, CHRIS, ; KURN, NURITH)

Classification: **- international:** ***C40B50/06***
- cooperative: **C12N15/1068**; **C12N15/1072**; **C12N15/66**; C12Q2525/191 more

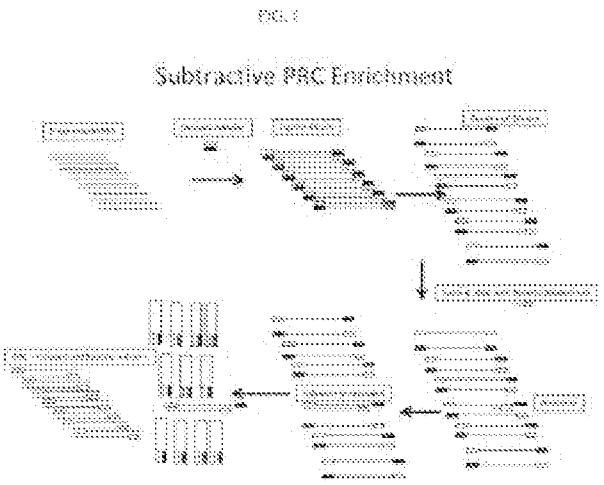
Application number: WO2013US23278 20130125 Global Dossier

Priority number (s): US201261591241P 20120126

Also published as: CA2862552 (A1) CN104093890 (A) CN104093890 (B) CN105861487 (A) EP2807292 (A1) more

Abstract of WO2013112923 (A1)

The present invention provides methods, compositions and kits for targeted nucleic acid sequence enrichment in a nucleic acid sample and for high efficiency nucleic acid library generation for next generation sequencing (NGS). Specifically, the methods, compositions and kits provided herein are useful for the production and capture of amplification-ready, target-specific and strand-specific regions of interest from nucleic acid samples containing complex DNA.



IDT Exhibit 1020.4365

IDT Exhibit 1020.4366

```
ERROR: ioerror
OFFENDING COMMAND: setfileposition

STACK:

-9
-filestream-
694
```

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2013 (01.08.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/112923 A1

(51) International Patent Classification:
C40B 50/06 (2006.01)

(21) International Application Number:
PCT/US2013/023278

(22) International Filing Date:
25 January 2013 (25.01.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/591,241 26 January 2012 (26.01.2012) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, San Carlos, CA 95050 (US).

(72) Inventors; and

(71) Applicants *(for US only)*: **AMORESE, Doug** [US/US]; 357 S. El Monte Ave., Los Altos, CA 94022 (US). **ARMOUR, Chris** [US/US]; 13531 104th Place, Ne., Kirkland, WA 98034 (US). **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) Agents: **ANDRES JR, Charles** et al.; Wilson Sonsini Goodrich & Rosati, 650 Page Mill Road, Palo Alto, CA 94304-1050 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

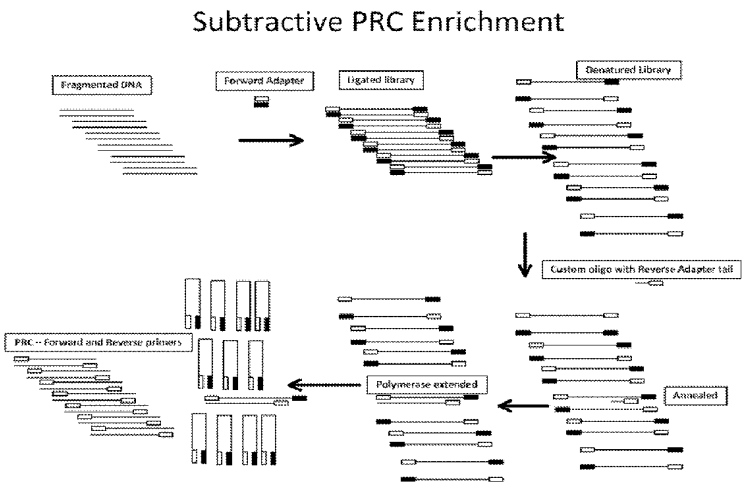
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

(54) **Title**: COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION

FIG. 1



(57) **Abstract**: The present invention provides methods, compositions and kits for targeted nucleic acid sequence enrichment in a nucleic acid sample and for high efficiency nucleic acid library generation for next generation sequencing (NGS). Specifically, the methods, compositions and kits provided herein are useful for the production and capture of amplification-ready, target-specific and strand-specific regions of interest from nucleic acid samples containing complex DNA.

WO 2013/112923 A1

COMPOSITIONS AND METHODS FOR TARGETED NUCLEIC ACID SEQUENCE ENRICHMENT AND HIGH EFFICIENCY LIBRARY GENERATION

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/591,241, filed January 26, 2012, which application is incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] With the rapid development of next generation sequencing (NGS) technologies and platforms, whole genome sequencing is becoming increasingly feasible. Researchers are driven to generate increasing amounts of data to achieve greater understanding of variance and biological trends, and to generate data from smaller sample sizes to avoid averaging across multiple cells within a tissue.

[0003] Although the cost of whole genome sequencing is decreasing and the throughput of the NGS platforms is increasing, it is nonetheless often more practical and cost-effective to select genomic regions of interest for sequencing and analysis. Target enrichment is a commonly employed strategy in genomic DNA sequencing in which genomic regions of interest are selectively captured from a DNA sample before sequencing. Focused target enrichment is an important tool especially in the fields of study where sequencing of a large number of samples is necessary (e.g. population-based studies of disease markers or SNPs), making whole genome sequencing cost-prohibitive. Similarly, improvements have been made that enable DNA libraries to be made from nucleic acid from fewer number of cells, but these are bound by the limitations of the efficiency of ligation reactions.

[0004] Several approaches to target enrichment have been developed which vary from one another in terms of sensitivity, specificity, reproducibility, uniformity, cost and ease of use. The target enrichment methods commonly employed today can be divided into three major categories, each with its distinct advantages and disadvantages: 1) PCR-based methods; 2) capture-by-hybridization, i.e. on-array or in-solution hybrid capture; and 3) capture-by-circularization, i.e. molecular inversion probe-based methods.

[0005] The PCR-based methods employ highly parallel PCR amplification, where each target sequence in the sample has a corresponding pair of unique, sequence-specific primers. The simultaneous use of numerous primer pairs makes multiplex PCR

IDT Exhibit 1020.4370

impractical due to high level of non-specific amplification and primer-primer interactions. Recently developed microdroplet PCR technology (Tewhey *et al.*, 2009) in which each amplification reaction is physically separated into an individual droplet removes the constraints of multiplex PCR relating to non-specific amplification and primer-primer interactions. However, microdroplet PCR and other improved PCR-based methods require special instrumentation or platforms, are limited in their throughput, and, as with conventional multiplex PCR, require a large number of individual primer pairs when enriching for a multitude of regions on interest, thus making target enrichment costly.

[0006] Hybrid capture methods are based on the selective hybridization of the target genomic regions to user-designed oligonucleotides. The hybridization can be to oligonucleotides immobilized on high or low density microarrays (on-array capture), or solution-phase hybridization to oligonucleotides modified with a ligand (e.g. biotin) which can subsequently be immobilized to a solid surface, such as a bead (in-solution capture). The hybrid capture methods require complex pools of costly long oligonucleotides and long periods (typically 48 hours) of hybridization for efficient capture. For on-array hybrid capture, expensive instrumentation and hardware is also required. Because of the relatively low efficiency of the hybridization reaction, large quantities of input DNA are needed.

[0007] The molecular inversion probe (MIP) based method relies on construction of numerous single-stranded linear oligonucleotide probes, consisting of a common linker flanked by target-specific sequences. Upon annealing to a target sequence, the probe gap region is filled via polymerization and ligation, resulting in a circularized probe. The circularized probes are then released and amplified using primers directed at the common linker region. One of the main disadvantages of the MIP-based target enrichment is its relatively low capture uniformity, meaning there is large variability in sequence coverage across the target regions. As with PCR and hybrid capture, the MIP-based method requires a large number of target-specific oligonucleotides, which can be costly.

[0008] There is a need for improved methods for selective target enrichment that allow for low-cost, high throughput capture of genomic regions of interest without specialized instrumentation. Additionally, there is also a need for high efficiency nucleic acid library generation. The methods of the invention described herein fulfills these needs.

IDT Exhibit 1020.4371

SUMMARY OF THE INVENTION

[0009] In one aspect, disclosed herein are methods for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising: (a) fragmenting the nucleic acids, thereby generating nucleic acid fragments; (b) appending a first adaptor to a 5' end of each nucleic acid fragment; (c) annealing one or more oligonucleotides to the nucleic acid fragments, whereby each of the one or more oligonucleotides comprise a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid fragments, and a 5' portion comprising a second adapter sequence; (d) extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products with the first adaptor at a first end and the second adaptor sequence at a second end; and (e) amplifying the one or more oligonucleotide extension product using a first primer that is complementary to the first adaptor and a second primer that is complementary to the second adaptor sequence to enrich for nucleic acid fragments containing the first adaptor and the second adaptor sequence at each end. In one embodiment, the method further comprises an additional step of sequencing the one or more oligonucleotide extension product following amplification. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA, RNA, or cDNA. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA. In one embodiment, the target nucleic acid sequences of interest comprise cDNA. In one embodiment, the method further comprises denaturing the nucleic acid fragments prior to step c, thereby generating single-stranded nucleic acid fragments with the first adaptor sequence at the 5'end. In one embodiment, the first adaptor can be common to each nucleic acid fragment. In one embodiment, the second adaptor sequence can be common to the one or more oligonucleotides. In one embodiment, the first adaptor and the second adaptor sequence can be distinct from each other. In one embodiment, the first adaptor and/or the second adaptor sequence further comprise barcode sequence. In one embodiment, step b can be performed by ligation. In one embodiment, the method further comprises an additional step of performing gap repair following ligation of the first adapter to create nucleic acid fragments with complementary termini. In one embodiment, a composition comprising enriched target nucleic acid sequences of interest can be generated by the methods disclosed herein. In one embodiment, the polymerase can be a DNA polymerase.

IDT Exhibit 1020.4372

[0010] In another aspect, disclosed herein are methods for enriching for target nucleic acid sequences of interest in a sample comprising nucleic acids, the method comprising: (a) fragmenting the nucleic acids, thereby generating nucleic acid fragments; (b) appending a first adaptor to the nucleic acid fragments wherein the first adaptor comprises a partial duplex with a short strand and a long strand wherein a 3' end of the short strand of the partial duplex adaptor comprises a blocking group, and a 5' end of the long strand of the partial duplex adaptor comprises a restriction and/or cleavage site for a nucleic acid modifying enzyme; (c) denaturing the nucleic acid fragments, thereby creating single-stranded nucleic acid fragments; (d) annealing one or more oligonucleotides to the single stranded nucleic acid fragments, whereby each of the one or more oligonucleotides comprise sequence that is complementary to a target nucleic acid sequence of interest present in one or more of the single-stranded nucleic acid fragments, extending the one or more oligonucleotides with a polymerase to produce one or more double stranded nucleic acid complexes comprising the target nucleic acid sequences and their complements, a first end with a double stranded restriction and/or cleavage site for the nucleic acid modifying enzyme , and a second end with a 3' overhang comprising the short strand of the first adaptor; (e) cleaving the double stranded restriction and/or cleavage site with the nucleic acid modifying enzyme, thereby generating a cleavage site; (f) ligating a second adaptor to the cleavage site, wherein the second adaptor comprises a duplex with two strands; (g) denaturing the one or more double-stranded nucleic acid complexes, thereby generating one or more single stranded nucleic acid fragments comprising the target nucleic acid sequences, a strand from the second adapter at the first end, and the short strand of the first adaptor at the second end; and (h) amplifying the one or more single stranded nucleic acid fragments comprising the one or more target nucleic acid sequences with a first primer comprising sequence complementary to the strand from the second adapter and a second primer comprising sequence complementary to the short strand of the first adaptor, thereby enriching for the one or more target nucleic acid sequences. In one embodiment, the method further comprises an additional step of sequencing the one or more single stranded nucleic acid fragments from step h following amplification. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA, RNA, or cDNA. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA. In one embodiment, the target nucleic acid sequences of interest comprise cDNA. In one embodiment, the first adaptor and the second adaptor can

IDT Exhibit 1020.4373

be common to each of the nucleic acid fragments. In one embodiment, the first adaptor and the second adaptor can be distinct from each other. In one embodiment, the first adaptor and/or the second adaptor further comprise barcode sequence. In one embodiment, the double stranded restriction and/or cleavage site for the nucleic acid modifying enzyme from step e comprises the 5' end of the long strand of the partial duplex of the first adaptor and sequence complementary to the 5' end of the long strand of the partial duplex of the first adaptor generated from extension of the one or more oligonucleotides. In one embodiment, the nucleic acid modifying enzyme comprises a restriction enzyme. In one embodiment, step b can be performed by ligation. In one embodiment, a composition comprising enriched target nucleic sequences of interest can be generated by the methods disclosed herein. In one embodiment, the polymerase can be a DNA polymerase.

[0011] In yet another aspect, disclosed herein are methods for generating a library of nucleic acid sequences, the method comprising: (a) fragmenting a sample comprising nucleic acids, thereby generating nucleic acid fragments; (b) appending a first adapter to each of the nucleic acid fragments; (c) denaturing the nucleic acid fragments, thereby generating a library of single-stranded nucleic acid fragments; (d) annealing one or more oligonucleotides to the single-stranded nucleic acid fragments wherein each of the one or more oligonucleotides comprises a 3' portion complementary to sequence in one or more of the single-stranded nucleic acid fragments and a 5' portion comprising a second adaptor sequence; (e) extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products comprising the first adaptor at a first end and the second adapter sequence at a second end; and (f) amplifying the one or more oligonucleotide extension products with a set of primers specific to the first adaptor and the second adaptor sequence to generate a library of nucleic acid fragments comprising the first adaptor and second adaptor sequence at each end. In one embodiment, the method further comprises an additional step of performing a gap repair reaction following ligation of the forward adapter to create nucleic acid fragments with complementary termini. In one embodiment, the method further comprises an additional step of sequencing the amplified one or more oligonucleotide extension product from step f. In one embodiment, the nucleic acid sequence comprises genomic DNA. In one embodiment, the nucleic acid sequence comprises cDNA. In one embodiment, step c can be omitted wherein the nucleic acid fragments are double-stranded. In one embodiment,

IDT Exhibit 1020.4374

the 3' portion of the one or more oligonucleotides of step d comprises random sequence. In one embodiment, step b can be performed by ligation. In one embodiment, the first adaptor and the second adaptor sequence can be common to each nucleic acid fragment. In one embodiment, the first adaptor and the second adaptor sequence can be distinct from each other. In one embodiment, the first adaptor and/or the second adaptor sequence further comprise barcode sequence. In one embodiment, the methods disclosed herein can generate a composition comprising a library of nucleic sequences. In one embodiment, the polymerase can be a DNA polymerase.

[0012] In further aspect, disclosed herein are methods for enriching for target nucleic acid sequences of interest from a library comprising nucleic acid inserts with a first adaptor on a first end and a second adaptor on a second end, the method comprising: (a) denaturing the nucleic acid inserts, thereby generating a library of single stranded nucleic acid inserts; (b) annealing one or more oligonucleotides to the single stranded nucleic acid inserts, wherein each of the one or more oligonucleotides comprises a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid inserts, and a 5' portion comprising a third adaptor sequence; (c) extending the one or more oligonucleotides with a polymerase thereby generating one or more oligonucleotide extension products with the first adaptor at the first end and the third adaptor sequence at the second end; and (d) amplifying the one or more oligonucleotide extension products using a first primer that is complementary to the first adaptor and a second primer that is complementary to the third adaptor sequence to enrich for nucleic acid fragments containing the first adaptor and the third adaptor sequence at each end. In one embodiment, the method further comprises an additional step of sequencing the amplified one or more oligonucleotide extension products from step d. In one embodiment, the target nucleic acid sequences of interest comprise genomic DNA. In one embodiment, the target nucleic acid sequences of interest comprise cDNA. In one embodiment, step a can be omitted wherein the nucleic acid fragments can be double-stranded. In one embodiment, the first adaptor and the second adaptor can be common to each nucleic acid fragment. In one embodiment, the third adaptor sequence can be common to the one or more oligonucleotides. In one embodiment, the first adaptor and the second adaptor can be distinct from each other. In one embodiment, the first adaptor and the second adaptor can be the same. In one embodiment, the first adaptor, the second adaptor and/or the third adaptor sequence further comprise barcode sequence. In one

IDT Exhibit 1020.4375

embodiment, the methods disclosed herein can generate a composition comprising enriched target nucleic acid sequences of interest.

INCORPORATION BY REFERENCE

[0013] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0015] FIG. 1 depicts selective target enrichment using ligation of a single forward adaptor at the ends of the DNA fragments in the DNA library. The sequence-specific oligonucleotide that anneals to the target region of interest contains a common reverse adaptor sequence at its 5'end, and following sequence-specific oligonucleotide extension, PCR is performed using a set of primers specific to the forward and reverse adaptors.

[0016] FIG. 2 depicts an alternative ligation protocol wherein the DNA-fragment-adaptor complex is denatured following ligation without nick repair and adaptor fill-in, generating ligation products where non-complementary ends exist on each insert.

[0017] FIG. 3 depicts selective target enrichment using ligation of partial duplex adaptors. Cleavage of the 5' end of the long strand of the partial duplex adaptor (and the corresponding complementary sequence of the extended sequence-specific oligonucleotide) by a nucleic acid modifying enzyme specific for double-stranded DNA allows for ligation of a new adaptor pair , and consequently, amplification with primers corresponding to the new adaptors.

[0018] FIG. 4 depicts high efficiency NGS library generation using random priming. The oligonucleotide that anneals to the DNA fragment contains a common reverse adaptor sequence at its 5' end, and following primer extension, PCR is performed using a set of primers specific to the forward and reverse adaptors.

IDT Exhibit 1020.4376

DETAILED DESCRIPTION OF THE INVENTION

General

[0019] The methods of the invention can be used for the selective enrichment of a plurality of defined target sequences from complex DNA with a set of common primers and adaptors, thus circumventing the need for multiplex PCR and multiple primer pairs. A multiplicity of target regions of interest are envisioned: for example, the regions of interest can represent all known coding regions, the entire exome, selected regions of coding genomic regions representing selected pathways, selected genomic regions known to comprise genomic variation related to altered phenotype, entire or selected regions of a specific chromosome, and the like. In another aspect, the methods of the invention can be used for high efficiency nucleic acid library production as well.

[0020] Altogether, the methods of the present invention create a simple, low cost, high throughput system for target enrichment and library preparation.

[0021] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the invention. On the contrary, the invention is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the invention.

[0022] In one embodiment, the present invention provides methods and compositions for the enrichment of specific target sequences of interest from a sample comprising nucleic acids. The methods described herein enrich target sequences using conventional duplex adaptors and/or partial duplex adaptors, sequence specific oligonucleotides, restriction enzymes and ligation. The methods further enable enrichment of target sequences from specific strands of template nucleic acids which can be further amplified using a variety of amplification methods. In another embodiment, the present invention provides methods for high efficiency generation of libraries comprising specific nucleic acid sequences of interest.

[0023] In one embodiment, the present invention provides methods and compositions for the enrichment of target nucleic acid sequences from a sample comprising nucleic acids. In one aspect, the method comprises fragmenting nucleic acids in an input sample to generate nucleic acid fragments. The nucleic acids can be DNA, or RNA. The nucleic acids can be single or double stranded. The DNA can be genomic DNA or cDNA or any

IDT Exhibit 1020.4377

combination thereof. In one embodiment, the nucleic acids in an input sample are double stranded DNA. In one embodiment, fragmentation of the nucleic acids can be achieved through methods known in the art. Fragmentation can be through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include nebulization, sonication, and/or hydrodynamic shearing. In some embodiments, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acids in the input sample to acoustic sonication. In some embodiments, the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of nucleic acid or polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g, from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the nucleic acids in the input sample with one or more restriction endonucleases. Fragmentation can produce fragments having 5' overhangs, 3' overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[0024] In some embodiments, the nucleic acids in the input sample can be fragmented into a population of fragmented nucleic acid molecules or polynucleotides of one or more specific size range(s). In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 10,000 nucleotide, such as less than 5,000 nucleotides, less than 2,500 nucleotides, less than 2,500 nucleotides, less than 1,000 nucleotides, less than 500

nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides.

[0025] In one embodiment, fragmentation of the nucleic acids can be followed by end repair of the nucleic acid fragments. End repair can include the generation of blunt ends, non-blunt ends (i.e sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the nucleic acid fragments, by a polymerase lacking 3′-exonuclease activity. End repair can be performed using any number of enzymes and/or methods known in the art including, but not limited to, commercially available kits such as the Encore ™ Ultra Low Input NGS Library System I. In a preferred embodiment, end repair can be performed on double stranded DNA fragments to produce blunt ends wherein the double stranded DNA fragments contain 5' phosphates and 3' hydroxyls. In some embodiments, the double-stranded DNA fragments can be blunt-end polished (or "end repaired") to produce DNA fragments having blunt ends, prior to being joined to adapters. Generation of the blunt ends on the double stranded fragments can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded DNA fragments can be blunt ended by the use of a single stranded specific DNA endonuclease, for example, but not limited to, mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, or any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded fragments generated by fragmenting the sample comprising nucleic acids. In still other cases, the nucleic acid fragments can be made blunt ended by filling in the overhanging single stranded ends of the double stranded fragments. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions of the double stranded fragments. Alternatively, the double

IDT Exhibit 1020.4379

stranded DNA fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0026] In some embodiments, the 5' and/or 3' end nucleotide sequences of fragmented nucleic acids are not modified or end-repaired prior to ligation with the adapter oligonucleotides of the present invention. For example, fragmentation by a restriction endonuclease can be used to leave a predictable overhang, followed by ligation with one or more adapter oligonucleotides comprising an overhang complementary to the predictable overhang on a nucleic acid fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of blunt-ended nucleic acid fragments to adapter oligonucleotides comprising a blunt end. In some embodiments, end repair can be followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. Nucleic acid fragments having an overhang can be joined to one or more adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3' ends of end repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3' end. In some embodiments, adapter oligonucleotides can be joined to blunt end double-stranded nucleic acid fragments which have been modified by extension of the 3' end with one or more nucleotides followed by 5' phosphorylation. In some cases, extension of the 3' end can be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, nucleic acid fragments having blunt ends can be joined to one or more adapters comprising a blunt end. Phosphorylation of 5' ends of nucleic acid fragments can be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented nucleic acid molecules may optionally be treated to dephosphorylate 5' ends or 3' ends, for example, by using enzymes known in the art, such as phosphatases.

[0027] The methods described herein for enriching for target nucleic acid sequences further comprise appending a first adaptor to the nucleic acid fragments generated by the methods described herein. In one embodiment, the first adaptor can be a forward adaptor.

IDT Exhibit 1020.4380

Appending the first adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction. In one embodiment, appendage of a first adaptor to the nucleic acid fragments comprises ligation. In one embodiment, ligation of the first adaptor to the nucleic acid fragments can be following end repair of the nucleic acid fragments. In another embodiment, the ligation of the first adaptor to the nucleic acid fragments can be following generation of the nucleic acid fragments without end repair of the nucleic acid fragments. The first adaptor can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors in which the adaptor comprises two complementary strands. In a preferred embodiment, the first adaptor can be a double stranded DNA adaptor. In one embodiment, the first adaptor can be an oligonucleotide of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the first adaptor(s) is appended or attached. In one embodiment, the first adaptor can be a conventional duplex adaptor, wherein the first adaptor comprises sequence well known in the art. In a preferred embodiment, the first adaptor can be appended to the nucleic acid fragments generated by the methods described herein in multiple orientations. In a preferred embodiment, the methods described herein can involve the use of a first duplex adaptor comprising double stranded DNA of known sequence that is blunt ended and can bind to the double stranded nucleic acid fragments generated by the methods described herein in one of two orientations. In one embodiment, the first adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same first adaptor. In other words, each of the nucleic acid fragments comprises a common first adaptor. In another embodiment, a first adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein such that each nucleic acid fragment in the library of nucleic acid fragments comprises the first adaptor ligated to one or both ends.

[0028] In one embodiment, the first adaptor can be ligated or appended to the 5' and/or 3' ends of the nucleic acid fragments generated by the methods described herein. The first adaptor can comprise two strands wherein each strand comprises a free 3' hydroxyl group but neither strand comprises a free 5' phosphate. In one embodiment, the free 3' hydroxyl group on each strand of the first adaptor can be ligated to a free 5' phosphate present on either end of the nucleic acid fragments of the present invention. In this

IDT Exhibit 1020.4381

embodiment, the first adaptor comprises a ligation strand and a non-ligation strand whereby the ligation strand can be ligated to the 5'phosphate on either end of the nucleic acid fragment while a nick or gap can be present between the non-ligation strand of the first adaptor and the 3' hydroxyl on either end of the nucleic acid fragment. In one embodiment, the nick or gap can be filled in by performing a gap repair reaction. In one embodiment, the gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In one embodiment, the gap repair can be performed using a DNA dependent DNA polymerase with weak or no strand displacement activity. In one embodiment, the ligation strand of the first adaptor can serve as the template for the gap repair or fill-in reaction. In this embodiment, the gap repair or fill-in reaction comprises an extension reaction wherein the ligation strand of the first adaptor serves as a template and leads to the generation of nucleic acid fragments with complementary termini or ends as depicted, for example, in FIG. 1. In one embodiment, the gap repair can be performed using *Taq* DNA polymerase. In one embodiment, the ligation of the first adaptor to the nucleic acid fragments generated by the methods described herein may not be followed gap repair as depicted, for example, in FIG. 2. In this embodiment, the nucleic acid fragments comprise first adaptor sequence ligated only at the 5' end of each strand.

[0029] Ligation and, optionally gap repair, of the first adaptor to the nucleic acid fragments generates a first adaptor-nucleic acid fragment complex. In one embodiment, the first adaptor-nucleic acid fragment complex can be denatured. Denaturation can be achieved using any of the methods known in the art including, but not limited to, physical, thermal, and/ or chemical denaturation. In one embodiment, denaturation can be achieved using thermal or heat denaturation. In one embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising first adaptor sequence at only the 5'end of the nucleic acid fragments as depicted, for example, in FIG. 2. In another embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising first adaptor sequence at both the 5'end and 3'end of the nucleic acid fragments as depicted, for example, in FIG.1.

[0030] In one embodiment, the nucleic acid fragments comprising first adaptor sequence appended to either the 5' end or both the 5' and 3' end can be denatured to generate single stranded nucleic acid fragments comprising first adaptor sequence appended to either the

IDT Exhibit 1020.4382

5' end or both the 5' and 3' end. In one embodiment, the methods of the present invention described herein can be used to generate a plurality of single stranded nucleic acid fragments comprising first adaptor sequence appended to either the 5' end or both the 5' and 3' end. In one embodiment, an oligonucleotide comprising at a first end sequence complementary to a target sequence of interest present in a single stranded nucleic acid fragment and at a second end sequence from a second adaptor , wherein the second adaptor sequence is not complementary to the target nucleic acid can be annealed to the single stranded nucleic acid fragments. In one embodiment, the second adaptor sequence can be sequence from a reverse adaptor. In one embodiment, the target nucleic acid sequence of interest can be present in one or more of the single stranded nucleic acid fragments. In one embodiment, different or distinct target nucleic acid sequences of interest can be present in one or more of the single stranded nucleic acid fragments. In one embodiment, one or more oligonucleotides can comprise sequence complementary to the same sequence of interest present in one or more single stranded nucleic acid fragments. In this embodiment, the one or more oligonucleotides can comprise sequence that is complementary to different parts or regions of the same sequence of interest. In one embodiment, the different regions can be adjacent to each other. In one embodiment, the different regions can be non-adjacent to each other. In a preferred embodiment, the one or more oligonucleotides that comprise sequence complementary to the same target nucleic acid sequence of interest further comprise the same second adaptor sequence. In another embodiment, one or more oligonucleotides can comprise sequence complementary to different or distinct sequences of interest which can be present in one or more single stranded nucleic acid fragments. In a preferred embodiment, the one or more oligonucleotides that comprise sequence complementary to different or distinct target nucleic acid sequences of interest further comprise the same second adaptor sequence. In one embodiment, the sequence complementary to the target sequence of interest can be at the 3'end of the oligonucleotide and the second adaptor sequence can be at the 5'end of the oligonucleotide. In a preferred embodiment, the second adaptor sequence is non-complementary to the target nucleic acid sequence of interest. In this manner, the second adaptor sequence serves as a tail. The second adaptor sequence can be a conventional adaptor sequence. In a preferred embodiment, the second adaptor sequence can be a conventional adaptor sequence that is different than or distinct from the sequence of the first adaptor appended to the single stranded nucleic acid fragment as

IDT Exhibit 1020.4383

described above. In one embodiment, the second adaptor sequence can be of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the second adaptor sequence is appended or attached. In a separate embodiment, the oligonucleotide can be annealed to the nucleic acid fragments comprising the first adaptor sequence appended to either the 5' end or both the 5' and 3' end without prior denaturation. In this embodiment, annealing of the oligonucleotide can be via formation of a triple helix or triplex between the oligonucleotide and a double stranded nucleic acid fragment comprising the first adaptor sequence appended to either the 5' end or both the 5' and 3' ends of the double stranded nucleic acid fragment. In this embodiment, the double stranded nucleic acid fragment comprises a sequence of interest and can be present amongst a plurality of double stranded nucleic acid fragments comprising first adaptor sequence appended to either the 5' end or both the 5' and 3' end. Further to this embodiment, the oligonucleotide comprises sequence complementary to the sequence of interest in the double stranded nucleic acid fragment. Overall, the use of the oligonucleotide comprising sequence complementary to a target sequence of interest present in a nucleic acid fragment amongst one or more or a plurality of nucleic acid fragments allows for selective binding and subsequent enrichment of said nucleic fragment using the methods described herein.

[0031] Following annealing of the oligonucleotide as described above, a polymerase can be used to extend the oligonucleotide. In one embodiment, the polymerase can be a DNA dependent DNA polymerase. In one embodiment, the DNA dependent DNA polymerase can be any of the DNA dependent DNA polymerases as described herein and extension of the oligonucleotide can be by any of the methods known in the art. In one embodiment, an oligonucleotide comprising the second adaptor sequence, wherein the second adaptor sequence is not complementary to the target nucleic acid, and sequence complementary to a target sequence of interest present in a nucleic acid fragment comprising a first adaptor appended to one and/or both ends can be annealed to the nucleic acid fragment and extended with a polymerase to generate an oligonucleotide extension product comprising the first adaptor sequence at a first end and the second adaptor sequence at a second end. In one embodiment, the nucleic acid fragment can be present amongst a plurality of nucleic acid fragments comprising first adaptor appended to one and/or both ends. In this

IDT Exhibit 1020.4384

embodiment, the oligonucleotide extension product can only be generated for a nucleic acid fragment that contains the target sequence of interest.

[0032] In one embodiment, the oligonucleotide extension product generated by the methods described herein can be subjected to an amplification reaction. In one embodiment, the amplification reaction can be exponential, and may be carried out at various temperature cycles or isothermal. In one embodiment, the amplification can be polymerase chain reaction. In one embodiment, the amplification reaction can be isothermal. In one embodiment, the oligonucleotide extension product comprises first adaptor sequence on one end and second adaptor sequence on the other end as generated by the methods described herein. In a preferred embodiment, the oligonucleotide extension product can be separated from the template nucleic acid fragment in order to generate a single stranded oligonucleotide extension product with first adaptor sequence on the 5' end and second adaptor sequence on the 3' end. The single stranded oligonucleotide extension product can then be amplified using a first primer comprising sequence complementary to the first adaptor and a second primer comprising sequence complementary to the second adaptor sequence. In this manner only oligonucleotide extension products comprising both the first and the second adaptor sequence will be amplified and thus enriched. In one embodiment, the first adaptor and/or the second adaptor sequence can comprise an identifier sequence. In one embodiment, the identifier sequence can be barcode sequence. In one embodiment, the barcode sequence can be the same or different for the first adaptor and the second adaptor sequence. In one embodiment, the first adaptor and/or the second adaptor sequence can comprise sequence that can be used for downstream applications such as, for example, but not limited to, sequencing. In one embodiment, the first adaptor and/or the second adaptor sequence can comprise flow cell sequences which can be used for sequencing with the sequencing method developed by Illumina and described herein.

[0033] In an alternate embodiment, the methods of the present invention can be used to generate a library of nucleic acid fragments or inserts wherein each nucleic acid fragment comprises an adaptor at one or both ends. In one embodiment, the adaptors can be present at both ends and can be distinct from each other. In one embodiment, the adaptors can be present at both ends and can comprise the same adaptor sequence. The generation of the library comprising nucleic acid inserts with distinct adaptors at both ends can involve the methods for generating oligonucleotide extension products comprising first

IDT Exhibit 1020.4385

adaptor sequence on one end and second adaptor sequence on the other end as described above with the exception that the oligonucleotide that binds to the nucleic acid fragments and can be extended comprises random sequence. In this embodiment, the oligonucleotide comprises random sequence at the 3' portion that is hybridizable to one or more nucleic acid fragments and further comprises second adaptor sequence at the 5'-portion. Extension of the oligonucleotide along the nucleic acid fragment and the corresponding first adaptor generates a product, or products, comprising the second adaptor at one end and a sequence complementary to the first adaptor at the other end, as illustrated in Figure 4. In one embodiment, the random sequence present in the oligonucleotide can bind to and be extended on one or more nucleic acid inserts. In one embodiment, one or more oligonucleotides comprising a 3' portion comprising random sequence and a 5' portion comprising second adaptor sequence can be annealed to a library comprising nucleic acid inserts comprising a first adaptor sequence on one or both ends of each of the nucleic acid inserts. In one embodiment, the first adaptor can be the same or common to each of the nucleic acid inserts. In one embodiment, the second adaptor sequence can be the same or common to each of the one or more oligonucleotides. In one embodiment, the methods described above can be used to generate a library of nucleic acid inserts comprising wherein each of the nucleic acid inserts comprises a common first adaptor on one end and a common second adaptor sequence on a second end. In one embodiment, the first adaptor and the second adaptor sequence can be distinct from each other. In one embodiment, the first adaptor and the second adaptor sequence can comprise the same adaptor sequence. Overall, the methods of the present invention as described above can be used for the high efficiency generation of a library of nucleic acid sequences.

[0034] In yet another alternate embodiment to the methods of the invention as described above, the first adaptor can be a double stranded DNA adaptor comprising a partial duplex, wherein the two strands of the adaptor can be different lengths with a complementary region and an overhanging region at the 5' end. In this embodiment, the 5' end of the long strand of the partial duplex adaptor can comprise a unique site for a nucleic acid modifying enzyme, such as a restriction enzyme, that is absent from the short strand of the duplex adaptor. In a further embodiment, the 3' end of the short strand adaptor can be modified by a replacement of the 3' OH-group by a blocking group, for example, a dideoxynucleotide (ddCMP, ddAMP, ddTMP, or ddGMP) to prevent

IDT Exhibit 1020.4386

polymerase extension. In this embodiment, the first adaptor comprising the partial duplex can be ligated to nucleic acid fragments generated by the methods described herein. In one embodiment, ligation of the partial duplex first adaptor can be followed by a gap repair reaction as described above. In this embodiment, ligation of the partial duplex first adaptor is not followed by a gap repair reaction. In a preferred embodiment, the partial duplex first adaptor comprises a free 5' phosphate on the short strand and a free 3'hydroxyl on the long strand. In this embodiment, ligation of the partial duplex adaptor generates double stranded nucleic acid fragments wherein both ends of the double stranded nucleic acid fragment comprise the long strand and short strand of the partial duplex first adaptor. A double stranded partial duplex first adaptor-nucleic acid fragment complex can be generated by ligation. In one embodiment, the double stranded partial duplex first adaptor-nucleic acid fragment complex can be denatured to generate a single stranded nucleic acid fragment comprising the long strand of the first adaptor on a first end and the short strand of the first adaptor on a second end. In this embodiment, the first end is the 5' end and the second end is the 3' end. In one embodiment, the first adaptor can be appended to one or more nucleic acid fragments as generated by the methods described herein such that each of the nucleic acid fragments comprises the same first adaptor or, in other words, the first adaptor can be common to each of the nucleic acid inserts. An oligonucleotide or primer comprising sequence complementary to a sequence of interest in the single stranded nucleic acid fragment can be annealed to the single stranded nucleic acid fragment and extended using a polymerase. In one embodiment, the polymerase can be a DNA dependent DNA polymerase. In one embodiment, the DNA dependent DNA polymerase can be any of the DNA dependent DNA polymerases as described herein and extension of the oligonucleotide can be by any of the methods known in the art. Extension of the primer annealed to the single stranded nucleic acid fragment generates an oligonucleotide extension product comprising sequence complementary to the long strand of the first adaptor on one end. In one embodiment, the oligonucleotide extension product remains hybridized to the single stranded nucleic acid fragment such that the restriction and/or cleavage site specific for a nucleic acid modifying enzyme is made double stranded. The double stranded site can then be cleaved by the nucleic acid modifying enzyme specific for the double stranded restriction site. In one embodiment, the nucleic acid modifying enzyme can be a restriction enzyme. In one embodiment, the restriction enzyme can be specific for a double stranded restriction site.

IDT Exhibit 1020.4387

In one embodiment, cleavage of the restriction site can generate a blunt end or non-blunt end. In one embodiment, end repair by any of the methods described herein can be performed on the end of the nucleic acid fragment following cleavage. Cleavage of the restriction and/or cleavage site generates a site to which a second adaptor can be ligated. Ligation of the second adaptor can be through any of the methods for ligation as described herein. In one embodiment, ligation generates a double stranded nucleic acid fragment comprising the second adaptor on a first end and a partial duplex on a second end, wherein the partial duplex comprises a 3' overhang comprising the sequence of the short strand of the first adaptor. The double stranded nucleic acid fragment can then be denatured using any of the methods for denaturation disclosed herein to generate a single stranded nucleic acid fragment comprising the second adaptor sequence on the first end and the sequence of the short strand of the first adaptor on the second end. In one embodiment, the first end and second end comprise the 5' end and 3' end, respectively. In one embodiment, the second adaptor can be appended to one or more nucleic acid fragments following cleavage of the double stranded restriction site such that each of the nucleic acid fragments comprises the same second adaptor or, in other words, the second adaptor can be common to each of the nucleic acid inserts. The single stranded nucleic acid fragment can then be amplified using a first primer specific for the second primer and a second primer specific for sequence present in the short strand of the first adaptor. In one embodiment, the amplification reaction can be exponential, and may be carried out at various temperature cycles or isothermal. In one embodiment, the amplification can be polymerase chain reaction. In one embodiment, the amplification reaction can be isothermal. Overall, only a fragment comprising the second adaptor and the short strand of the first adaptor will be amplified or enriched. In so far as the method provides for enrichment of targeted fragments of the library, and not enrichment of oligonucleotide extension products generated by the extension of the oligonucleotide comprising sequence complementary to a target sequence of interest, there is no distortion of the original DNA library, and the enrichment is independent of the insert length. Because the 3' end of the short strand of the partial duplex adaptor is 3' blocked, the method enables directional or asymmetric ligation. In one embodiment, the oligonucleotide that comprises sequence complementary to a sequence of interest in a nucleic acid fragment further comprises reverse adaptor sequence. In this embodiment, the sequence complementary to a sequence of interest in the nucleic acid fragment can be present in a 3' portion of the

IDT Exhibit 1020.4388

oligonucleotide and the reverse adaptor sequence can be present at a 5' portion. Further to this embodiment, the reverse adaptor sequence can be a common or conventional adaptor sequence and can be different or distinct from the first and/or second adaptors. Further still to this embodiment, the methods described above can lead to the generation of a single stranded nucleic acid fragment comprising the second adaptor at one end and the reverse adaptor sequence at the other end. Subsequent to this embodiment, the single stranded nucleic acid fragment can be enriched through amplification using a first primer specific to the second adaptor and a second primer specific to the third adaptor sequence.

[0035] The methods of the inventions are further applicable to any enrichment of target nucleic acid sequences of interest from libraries comprising fragments of nucleic acid of a sample appended with adaptor sequence at one or both ends, wherein the libraries are generated using ligation of the adaptor or adaptor sequences to one or both ends as described herein or by ligation independent methods, such as for example Nextera, a transposome driven method. In one embodiment, the nucleic acid can be DNA such as genomic DNA or cDNA. In one embodiment, the nucleic acid can be double stranded. Enrichment of nucleic acid sequences of interest can be achieved using the methods described herein for target enrichment. In one embodiment, the method for enriching for target nucleic acid sequences of interest from a library comprising nucleic inserts with adaptors appended to one or both ends comprises denaturing the nucleic acid inserts to generate a library of single stranded nucleic acid inserts. In one embodiment, each of the nucleic acid inserts can comprise a first adaptor sequence on one end and a second adaptor sequence on an opposite end. In one embodiment, the first adaptor and the second adaptor can be distinct from each other. In one embodiment, the first adaptor and the second adaptor can comprise the same adaptor sequence. In one embodiment, each of the nucleic acid inserts can comprise a first adaptor sequence on one end and a second adaptor sequence on an opposite end such that denaturation generates a library of single stranded nucleic acid inserts comprising the first adaptor sequence on one end and the second adaptor sequence on an opposite end. Denaturation can be achieved using any of the methods described herein. Further to the embodiments described above, one or more oligonucleotides can be annealed to the single stranded nucleic acid inserts. In one embodiment, each of the one or more oligonucleotides comprises a 3' portion that is complementary to a target nucleic acid sequence of interest present in one or more of the nucleic acid inserts, and a 5' portion comprising a third adaptor sequence. In one

IDT Exhibit 1020.4389

embodiment, the third adaptor sequence is distinct from either or both of the first adaptor and the second adaptor. The one or more oligonucleotides can be extended with a polymerase (i.e. a DNA polymerase) thereby generating one or more oligonucleotide extension products with the first or second adaptor at a first end and the third adaptor sequence at a second end. In one embodiment, the first end comprises the 5' end and the second end comprises the 3' end. The one or more oligonucleotide extension products can be amplified using a first primer that can be complementary to the first or second adaptor and a second primer that can be complementary to the third adaptor sequence to enrich for nucleic acid fragments comprising the first or second adaptor and the third adaptor sequence at each end. In one embodiment, the first and second adaptors can be common to each of the nucleic acid inserts in the library. In one embodiment, the third adaptor sequence can be common to each of the one or more oligonucleotides. Overall, the target enrichment methods as described above can be used to generate a composition comprising a library of nucleic acid inserts enriched for any target sequence of interest from a non-enriched library comprised of nucleic acid inserts with an adaptor ligated to one or both ends.

[0036] A schematic of a preferred embodiment of the methods described herein for enriching for target sequences of interest is illustrated in FIG. 1. Overall, FIG. 1 depicts a method for isolating or enriching for a nucleic acid fragment or insert comprising a target nucleic acid sequence from a library or plurality of nucleic acid fragments. The method in FIG. 1 involves generation of a ligated library of nucleic acid fragments or inserts wherein each fragment or insert of the ligated library comprises a common forward adaptor and a fragment or insert specific reverse adaptor distinct from the forward adaptor such that subtractive PCR using a primer directed against the common forward adaptor and a primer directed against the reverse adaptor enriches for a nucleic acid fragment or insert comprising a target nucleic acid sequence. The input for the method depicted in FIG.1 is fragmented DNA. The fragmented DNA is double stranded and comprises a plurality or library of DNA fragments. In one embodiment, the DNA fragments can be derived from complex DNA, such as double-stranded DNA, genomic DNA or mixed DNA from more than one organism. In one embodiment, the DNA fragments can be derived from RNA that has been converted to cDNA through a first strand synthesis reaction using any of the methods well known in the art for generating cDNA from an RNA template which can include, but is not limited to, combining the RNA with a primer

IDT Exhibit 1020.4390

(i.e. random primer), and reverse transcribing the RNA template with an RNA-dependent DNA polymerase. In one embodiment, the DNA fragments can be derived from RNA that has been converted to double stranded cDNA through a first and second strand synthesis reaction using any of the methods well known in the art. Fragmentation of the DNA to produce the DNA fragments can be achieved through any of the methods described herein for fragmenting nucleic acids which can include, but are not limited to, physical (i.e. sonication), and/or chemical (i.e. restriction enzyme treatment) fragmentation reactions.

[0037] As depicted in FIG. 1, a single forward adaptor is ligated to the DNA fragments. In one embodiment, the single forward adaptor can comprise known sequence. In one embodiment, the single forward adaptor can be a common adaptor. In one embodiment, the DNA fragments can be subjected to an end repair reaction as described herein to produce blunt ends. In this embodiment, the single forward adaptor can also comprise blunt ends and ligation between the single forward adaptor and the DNA fragments can be through blunt end ligation as described herein. Ligation can be facilitated through the use of enzymes (i.e. T4 DNA ligase) and methods known in the art, including, but not limited to, commercially available kits such as the Encore ^{TM} Ultra Low Input NGS Library System. In FIG. 1, the forward adaptor can contain a strand (the ligation strand) that ligates with the free 5'phosphate on a 5' end of the DNA fragments and a strand that does not ligate (non-ligation strand) to a 3' end of the DNA fragments. In one embodiment, the ligation reaction can lead to the generation of a nick or gap between the non-ligation strand of the single forward adaptor and the 3' end of the DNA fragments. In this embodiment, the nick or gap can be repaired or filled in through a gap repair or fill-in reaction wherein the 3' end of the DNA fragments can be extended with a polymerase (preferably with a DNA dependent DNA polymerase such as *Taq* DNA polymerase) wherein the ligation strand of the forward adaptor can serve as template. In this embodiment, the gap repair generates DNA fragments with complementary ends. As depicted in FIG. 1, the DNA fragments with complementary ends are denatured to generate a denatured library comprising single stranded DNA fragments with complementary ends. Denaturation can be achieved using any of the methods known in the art which can include, but are not limited to, heat denaturation, and/or chemical denaturation.

IDT Exhibit 1020.4391

[0038] As depicted in FIG. 1, a custom oligonucleotide with a reverse adaptor tail is annealed to the single stranded DNA fragments with complementary ends. In one embodiment, the custom oligonucleotide with a reverse adaptor tail can comprise a 3' portion comprising sequence complementary to a target sequence of interest in one of the single-stranded DNA fragments and a 5' portion comprising reverse adaptor sequence that is not complementary to the single-stranded DNA fragments in the denatured library. In one embodiment, the reverse adaptor sequence can be known sequence. In a preferred embodiment, the reverse adaptor sequence can be distinct from the single forward adaptor as described herein. In one embodiment, a plurality of custom oligonucleotides with a reverse adaptor tail can be added to the denatured library wherein the plurality of custom oligonucleotides with a reverse adaptor tail comprise a 3' portion comprising sequence complementary to a target sequence of interest in one or more of the single-stranded DNA fragments of the denatured library and a 5' portion comprising a reverse adaptor sequence that is not complementary to the single-stranded DNA fragments in the denatured library. In one embodiment, the reverse adaptor tail comprises the same reverse adaptor sequence in each of the plurality of custom oligonucleotides with a reverse adaptor tail, and wherein the reverse adaptor sequence is distinct from the forward adaptor sequence. In another embodiment, the reverse adaptor tail comprises a different reverse adaptor sequence for each of the plurality of custom oligonucleotides with a reverse adaptor tail, and wherein each of the different reverse adaptor sequences is distinct from the forward adaptor sequence. In one embodiment, the 3' portion of the custom oligonucleotide with a reverse adaptor tail can be a specific sequence, wherein the custom oligonucleotide comprises a sequence complementary to the target sequence of interest and provides a means for targeted enrichment of sequence or sequences of interest using the methods of the invention. In another embodiment the 3' portion of the custom oligonucleotide with reverse adaptor tail can be a randomly generated sequence hybridizable to random sequences of the library of fragments with adaptor sequences on one or both ends, providing means for efficient, non-enriched library generation employing the methods of the invention.

[0039] Following annealing of the custom oligonucleotide with a reverse adaptor tail to a sequence of interest in a single-stranded DNA fragment of the denatured library, the custom oligonucleotide with a reverse adaptor tail is extended using any method known in the art, which can include but is not limited to, extension using a DNA dependent DNA

IDT Exhibit 1020.4392

polymerase using the single stranded DNA fragment of the denatured library as a template. Extension of the custom oligonucleotide with a reverse adaptor tail generates an oligonucleotide extension product with forward adaptor sequence at one end and reverse adaptor sequence at the other end. In this embodiment, the custom oligonucleotide with a reverse adaptor tail can only anneal to and be extended on DNA fragments in the denatured library comprising the target sequence of interest for which the custom oligonucleotide with a reverse adaptor tail is directed. As illustrated in FIG. 1, a subtractive polymerase chain reaction (PCR) procedure is subsequently performed using a first primer directed against the forward adaptor sequence and a second primer directed toward the reverse adaptor sequence such that only the oligonucleotide extension product with the forward adaptor sequence at one end and the reverse adaptor sequence at the other end can be amplified and thus enriched.

[0040] FIG. 2 depicts another embodiment of subtractive PCR enrichment method as described for FIG. 1, wherein ligation of a duplex forward adaptor (P1) to a double stranded nucleic acid fragment is not subjected to gap repair. The duplex forward adaptor (P1) comprises a strand (the ligation strand; P1) that ligates with the free 5'phosphate on a 5' end of the nucleic acid fragment and a strand that does not ligate (non-ligation strand; P1rc) to a 3' end of the nucleic acid fragment. In this embodiment, the ligation strand is ligated to the 5' end of both strands of a double stranded nucleic fragment whereas a gap or nick is generated between the non-ligation strand and the 3'end of both strands of the double stranded nucleic acid fragment. As depicted in FIG. 2, the ligation of the P1 adaptor to the nucleic acid fragment is followed by denaturation without a gap repair or fill-in reaction thereby generating a single stranded nucleic acid fragment with non-complementary ends. In this embodiment, the single stranded nucleic acid fragment with non-complementary ends comprises a P1 forward adaptor sequence at a 5' end and fragment specific sequence at a 3' end. In a further embodiment, the single stranded nucleic fragment with non-complementary ends can be further processed as described above and illustrated in FIG. 1 to generate a single stranded nucleic fragment with the P1 forward adaptor sequence at the 5' end and a distinct reverse adaptor sequence on the 3' end.

[0041] FIG. 4 illustrates another embodiment of the present invention for the high efficiency generation of libraries comprising nucleic acid fragments with distinct adaptors on each end. In this embodiment, the methods for generating a denatured library

IDT Exhibit 1020.4393

comprising single stranded DNA fragments with a single forward adaptor sequence on a 5' end and a reverse adaptor sequence on a 3' end is similar to the method described for FIG. 1. However, in FIG. 4, an oligonucleotide with a reverse adaptor tail is used whereby the oligonucleotide comprises a 3' portion comprising random sequence and a reverse adaptor tail wherein the reverse adaptor tail comprises reverse adaptor sequence that is distinct from the single forward adaptor sequence. As illustrated in FIG. 4, PCR can be carried out with a first primer directed against the single forward adaptor sequence and a second primer directed toward the reverse adaptor sequence wherein both the first and second primers further comprise flow cell sequences. In this manner, single stranded DNA fragments with a single forward adaptor sequence on a 5' end and a reverse adaptor sequence on a 3' end comprise flow cell sequences that can be used to adhere the amplified single stranded DNA fragments to flow cells for subsequent sequencing by the method commercialized by Illumina, as described U.S. Pat. Nos. 5,750,341; 6,306,597; and 5,969,119.

[0042] FIG. 3 illustrates a method for enrichment of a target nucleic acid sequence, or sequences of interest, contained in a double stranded nucleic acid insert from a complex library. In one embodiment, the complex library comprises nucleic acid inserts from a genomic DNA sample, In FIG. 3, the single forward adaptor comprises a partial duplex forward adaptor comprising a long strand, A, that forms a partial duplex with a short strand, B. Strand A of the partial duplex adaptor further comprises a restriction enzyme site while strand B does not contain the restriction enzyme site. Strand B further comprises a blocking group wherein the 3' end of strand B is modified by replacement of the 3' OH group with a blocking group that can prevent polymerase extension. In one embodiment, the partial duplex forward adaptor is ligated to the double stranded nucleic acid fragments such that a double stranded insert with the partial duplex forward adaptor appended to both ends is generated. In this embodiment, the 5' end of strand B of the partial duplex adaptor can contain a free 5' phosphate which can be ligated to a free 3' OH present on one or both strands of the double stranded insert. Subsequent denaturation generates a single stranded insert comprising sequence A on a 5'end and strand B on a 3' end. A primer, C, directed against a specific sequence of interest within the single stranded insert can be annealed to the specific sequence and extended with a DNA polymerase using the single stranded insert as a template. In one embodiment, primer C can be a sequence specific primer and is employed for enrichment of target, or targets, of

IDT Exhibit 1020.4394

interest according to the methods of the invention. In one embodiment, primer C can be a random primer. Extension of primer C with DNA polymerase generates an extended primer C product that comprises sequence complementary to sequence A at its' 3' end in a double stranded complex with the template insert strand such that a double stranded restriction site has been generated between sequence A and its' complement. In one embodiment, the double stranded restriction enzyme recognition site can be cleaved by a restriction enzyme specific for the double stranded restriction site thereby generating a truncated, or cleaved, adaptor sequence at one end. A second forward adaptor comprising a common, or conventional, duplex adaptor D is then ligated to the cleavage site using any of the ligation methods described herein, thereby generating a double stranded complex comprising the second forward adaptor D at one end and a 3' overhang comprising strand B on the opposite end. The double stranded complex comprising the second forward adaptor D at one end and a 3' overhang comprising strand B on the opposite end is denatured and amplified using a first primer directed against the second forward adaptor D and a second primer directed against strand B. In this manner, the methods depicted in FIG. 3 can be used to enrich for specific sequences of interest from a complex library since the methods are designed such that the second forward adaptor D can only bind to the double stranded cleavage site generated by restriction enzyme digestion of the double stranded restriction site created between sequence A and its complement following extension of primer C. As described, primer C can be directed against a target sequence of interest present in a single, or multiple insert, or inserts, amongst a plurality of inserts. Moreover, the method can be made strand specific by designing primer C to bind to target sequence of interest present on one strand or the other of an insert in an amongst a plurality of inserts.

[0043] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, *DNA Replication*, Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry*, Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human *Molecular Genetics*, Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford, 1984); and the like.

IDT Exhibit 1020.4395

Input Nucleic Acid

[0044] The input can be a nucleic acid. In one embodiment, the input can be DNA. In one embodiment, the input nucleic acid can be complex DNA, such as double-stranded DNA, genomic DNA or mixed DNA from more than one organism. In one embodiment, the input can be RNA. In one embodiment, the RNA can be obtained and purified using standard techniques in the art and include RNAs in purified or unpurified form, which include, but are not limited to, mRNAs, tRNAs, snRNAs, rRNAs, retroviruses, small non-coding RNAs, microRNAs, polysomal RNAs, pre-mRNAs, intronic RNA, viral RNA, cell free RNA and fragments thereof. The non-coding RNA, or ncRNA may include snoRNAs, microRNAs, siRNAs, piRNAs and long nc RNAs. In one embodiment, the input nucleic acid can be cDNA. The cDNA can be generated from RNA, e.g., mRNA. The cDNA can be single or double stranded. The input DNA can be of a specific species, for example, human, rat, mouse, other animals, specific plants, bacteria, algae, viruses, and the like. The input complex also can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The input DNA can be cDNA made from a mixture of genomes of different species. Alternatively, the input nucleic acid can be from a synthetic source. The input DNA can be mitochondrial DNA. The input DNA can be cell-free DNA. The cell-free DNA can be obtained from, e.g., a serum or plasma sample. The input DNA can comprise one or more chromosomes. For example, if the input DNA is from a human, the DNA can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. The DNA can be from a linear or circular genome. The DNA can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The input DNA can be from more than one individual or organism. The input DNA can be double stranded or single stranded. The input DNA can be part of chromatin. The input DNA can be associated with histones.

[0045] In some embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to single-stranded nucleic acid targets. In one embodiment, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to single-stranded DNA targets. In the case where the input nucleic acid sample comprises genomic DNA or other double-stranded DNA, the input nucleic acid sample can be first denatured to render the target single stranded and enable hybridization of the oligonucleotides to the desired sequence regions of interest. In these

IDT Exhibit 1020.4396

embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest. In some embodiments, the other double-stranded DNA can be double-stranded cDNA generated by first and second strand synthesis of one or more target RNAs.

[0046] In other embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to double-stranded nucleic acid targets, without denaturation of the double stranded nucleic acids. In other embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to hybridize to a double-stranded DNA target, without denaturation of the dsDNA. In these embodiments, the oligonucleotides targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the oligonucleotides to the double-stranded DNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out in situ, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

[0047] A "target nucleic acid sequence" or "target sequence" as used herein, is a polynucleotide sequence of interest, for which enrichment is desired. The target sequence may be known or not known, in terms of its actual sequence. Generally, a "template", as used herein, is a polynucleotide that contains the target nucleic acid sequence. The terms "target sequence," "target nucleic acid sequence," "target nucleotide sequence," "regions of interest," or "sequence of interest" and, variations thereof, are used interchangeably.

Oligonucleotides of the invention

[0048] As used within the invention, the term "oligonucleotide" refers to a polynucleotide chain, typically less than 200 residues long, most typically between 15 and

IDT Exhibit 1020.4397

100 nucleotides long, but also intended to encompass longer polynucleotide chains. Oligonucleotides may be single-or double-stranded. As used in this invention, the term "oligonucleotide" may be used interchangeably with the terms "primer" and "adaptor".

[0049] As used herein, the terms "hybridization"/ "hybridizing" and "annealing" are used interchangeably and refer to the pairing of complementary nucleic acids.

[0050] The term "primer", as used herein, can refer to a nucleotide sequence, generally with a free 3' hydroxyl group, that is capable of hybridizing with a template (such as one or more target polynucleotides, one or more target DNAs, one or more target RNAs or a primer extension product) and is also capable of promoting polymerization of a polynucleotide complementary to the template. A primer can be, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase [i.e. RNase H] cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer. A primer may contain a non-hybridizing sequence that constitutes a tail of the primer. A primer may still be hybridizing to a target even though its sequences are not fully complementary to the target.

[0051] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as in PCR, SPIA or cDNA synthesis, for example. The oligonucleotide primer can be a synthetic polynucleotide that is single stranded, containing a sequence at its 3′-end that is capable of hybridizing with a sequence of the target polynucleotide. Normally, the 3′ region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a sequence or primer binding site.

[0052] "Complementary", as used herein, can refer to complementarity to all or only to a portion of a sequence. The number of nucleotides in the hybridizable sequence of a specific oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer will prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the hybridizing portion of the oligonucleotide primer will be at least as great as the defined sequence on the target polynucleotide that the oligonucleotide primer hybridizes to, namely, at least 5, at least 6,

IDT Exhibit 1020.4398

at least 7, at least 8, at least 9, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least about 20, and generally from about 6 to about 10 or 6 to about 12 or 12 to about 200 nucleotides, usually about 20 to about 50 nucleotides. In general, the target polynucleotide is larger than the oligonucleotide primer or primers as described previously.

[0053] In some cases, the identity of the investigated target polynucleotide sequence is known, and hybridizable sequence specific oligonucleotides or primers can be synthesized precisely according to the antisense sequence of the aforesaid target polynucleotide sequence. In some embodiments, multiple sequence-specific oligonucleotides or primers are employed to hybridize to a multiplicity of genomic regions of interest, allowing for selective enrichment of the regions of interest. In so far as the genomic regions may be very long, multiple oligonucleotides can be designed to hybridize to different sequence regions within the genomic regions of interest. In other embodiments, when the target polynucleotide sequence is unknown, the hybridizable sequence of an oligonucleotide or primer is a random sequence. Oligonucleotides or primers comprising random sequences may be referred to as "random primers", or "random oligonucleoties," as described herein. In one embodiment, an oligonucleotide or primer of the present invention hybridizable to a target sequence may comprise a mixture of primers or oilognucleotides designed to hybridize to a plurality (e.g. 2, 3, 4, about 6, 8, 10, 20, 40, 80, 100, 125, 150, 200, 250, 300, 400, 500, 600, 800, 1000, 1500, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 10,000, 20,000, 25,000 or more) of target sequences. In some cases, the plurality of target sequences may comprise a group of related sequences, random sequences, a whole transcriptome or fraction (e.g. substantial fraction) thereof, or any group of sequences such as mRNA. In some embodiments, the primers can be directed to known sequences present in the adaptors used in the invention as described herein. In this embodiment, the primers can comprise groups of primers comprising one or more primers in each group, wherein each group of primers can be directed against distinct adaptors.

[0054] Tailed primers or oligonucleotides can be employed in certain embodiments of the invention. In general, a tailed primer comprises a 3' portion that is hybridizable to one or more target polynucleotides, and a 5' portion that is not hybridizable to the one or more target polynucleotides. In general, the non-hybridizable 5' portion does not hybridize to the one or more target polynucleotides under conditions in which the hybridizable 3'

IDT Exhibit 1020.4399

portion of the tailed primer hybridizes to the one or more target polynucleotides. In some embodiments, the non-hybridizable 5' portion comprises an adaptor sequence. In some embodiments, the non-hybridizable 5' portion comprises a common or conventional adaptor sequence. In some embodiments, the non-hybridizable 5' portion comprises a common or conventional adaptor sequence that is distinct or different from the sequence of other adaptors used in the present invention. In some embodiments, the non-hybridizable 5' portion comprises a promoter-specific sequence. Generally, a promoter-specific sequence comprises a single-stranded DNA sequence region which, in double-stranded form is capable of mediating RNA transcription. Examples of promoter-specific sequences are known in the art, and include, without limitation, T7, T3, or SP6 RNA polymerase promoter sequences. When the tailed primer is extended with a DNA polymerase, a primer extension product with a 5' portion comprising a defined sequence can be created. This primer extension product can then have a second primer anneal to it, which can be extended with a DNA polymerase to create a double stranded product comprising a defined sequence at one end. In some embodiments, where the non-hybridizable 5' portion of one or more tailed primers comprises a promoter-specific sequence, creation of a double-stranded product comprising a defined sequence at one end generates a double-stranded promoter sequence that is capable of mediating RNA transcription. In some embodiments, a double-stranded promoter sequence can be generated by hybridizing to the promoter-specific sequence an oligonucleotide comprising a sequence complementary to the promoter-specific sequence. In some embodiments, formation of a double-stranded promoter can be followed by the generation of single-stranded RNA by RNA transcription of sequence downstream of the double-stranded promoter, generally in a reaction mixture comprising all necessary components, including but not limited to ribonucleoside triphosphates (rNTPs) and a DNA-dependent RNA polymerase. Tailed primers can comprise DNA, RNA, or both DNA and RNA. In some embodiments, the tailed primer consists of DNA.

[0055] Composite primers can be employed in certain embodiments of the invention. Composite primers are primers that are composed of RNA and DNA portions. In some aspects, the composite primer can be a tailed composite primer comprising, for example, a 3'-DNA portion and a 5'-RNA portion. In the tailed composite primer, a 3'-portion, all or a portion of which comprises DNA, is complementary to a polynucleotide; and a 5'-portion, all or a portion of which comprises RNA, is not complementary to the

IDT Exhibit 1020.4400