



US 20250257387A1

(19) United States

(12) Patent Application Publication (10) Pub. No.: US 2025/0257387 A1  
RAYMOND et al. (43) Pub. Date: Aug. 14, 2025

(54) METHODS FOR QUANTITATIVE GENETIC ANALYSIS OF CELL FREE DNA

(71) Applicant: Resolution Bioscience, Inc., Kirkland, WA (US)

(72) Inventors: Christopher K. RAYMOND, Seattle, WA (US); Lee P. LIM, Kirkland, WA (US); Christopher D. ARMOUR, Kirkland, WA (US)

(21) Appl. No.: 19/025,287

(22) Filed: Jan. 16, 2025

**Related U.S. Application Data**

(63) Continuation of application No. 18/983,944, filed on Dec. 17, 2024, which is a continuation of application No. 15/727,887, filed on Oct. 9, 2017, now Pat. No. 12,203,127, which is a continuation of application No. 14/466,741, filed on Aug. 22, 2014, now abandoned.

**Publication Classification**

(51) Int. Cl.

CI2Q 1/6806	(2018.01)
CI2Q 1/6869	(2018.01)

*G16B 20/00* (2019.01)*G16B 20/10* (2019.01)*G16B 20/20* (2019.01)*G16B 25/00* (2019.01)*G16B 25/10* (2019.01)*G16B 30/00* (2019.01)*G16B 30/10* (2019.01)

(52) U.S. Cl.

CPC ..... *CI2Q 1/6806* (2013.01); *CI2Q 1/6869* (2013.01); *G16B 20/00* (2019.02); *G16B 20/10* (2019.02); *G16B 20/20* (2019.02); *G16B 25/10* (2019.02); *G16B 30/00* (2019.02); *G16B 30/10* (2019.02); *G16B 25/00* (2019.02)(57) **ABSTRACT**

The invention provides a method for genetic analysis in individuals that reveals both the genetic sequences and chromosomal copy number of targeted and specific genomic loci in a single assay. The present invention further provides methods for the sensitive and specific detection of target gene sequences and gene expression profiles.

Specification includes a Sequence Listing.

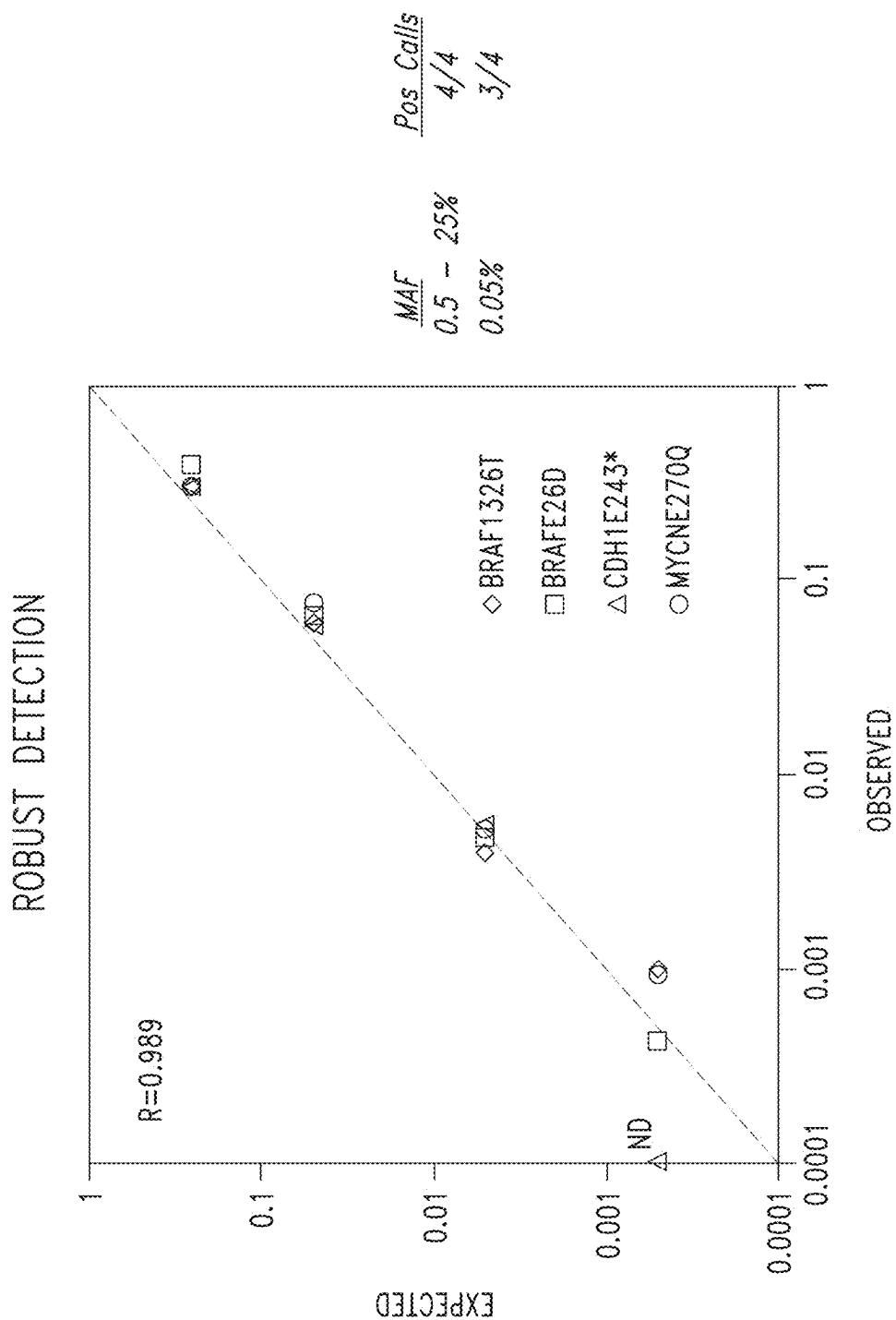


FIG. 1

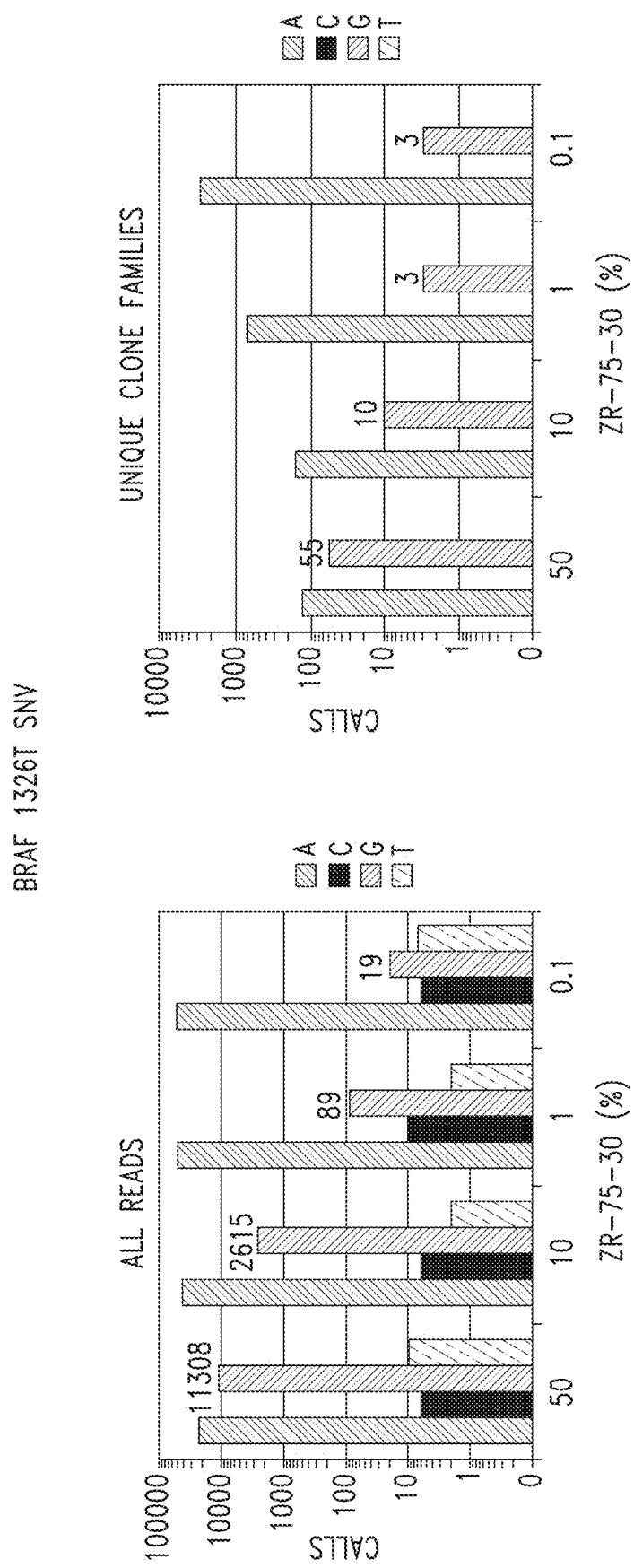
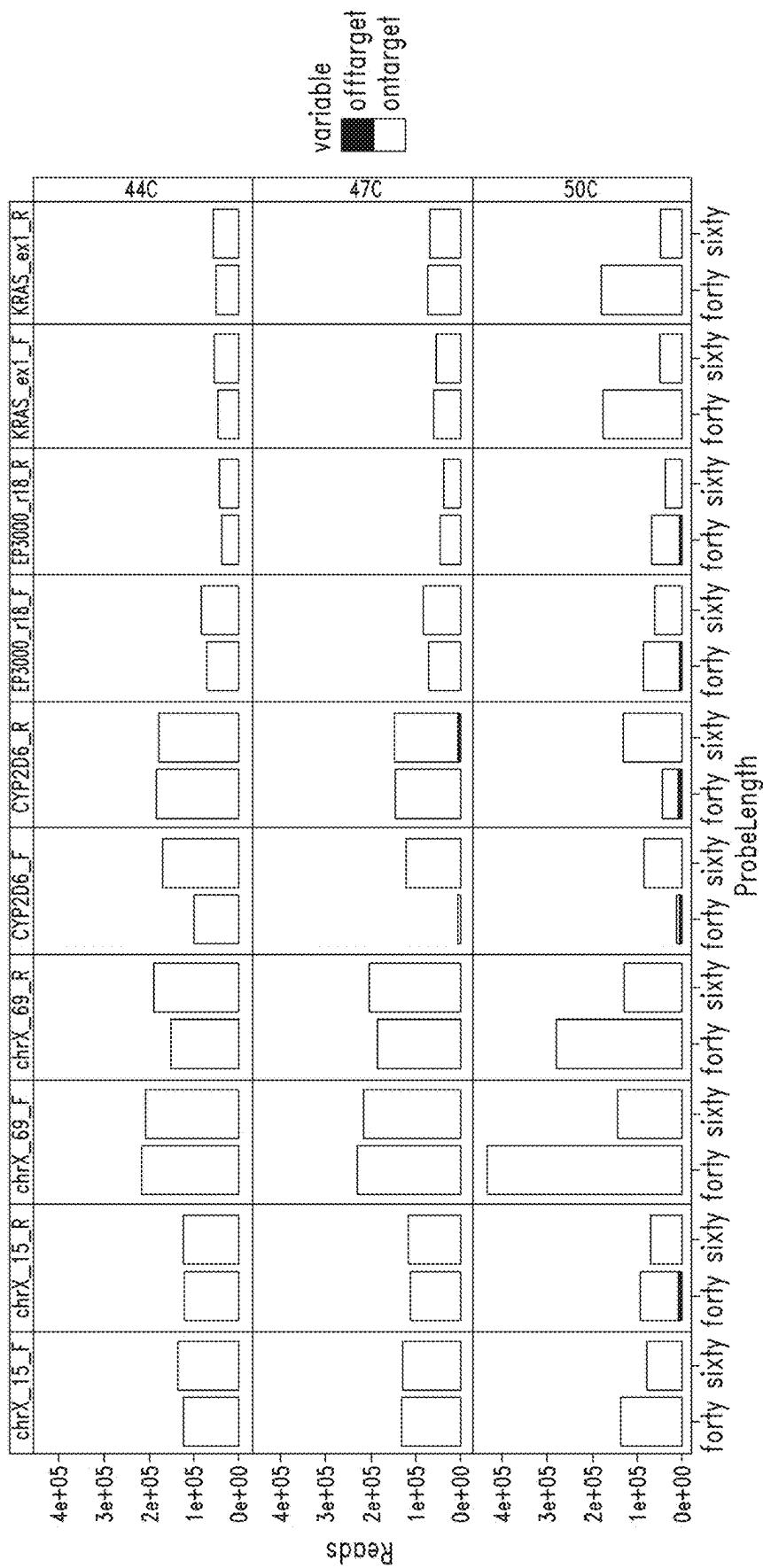
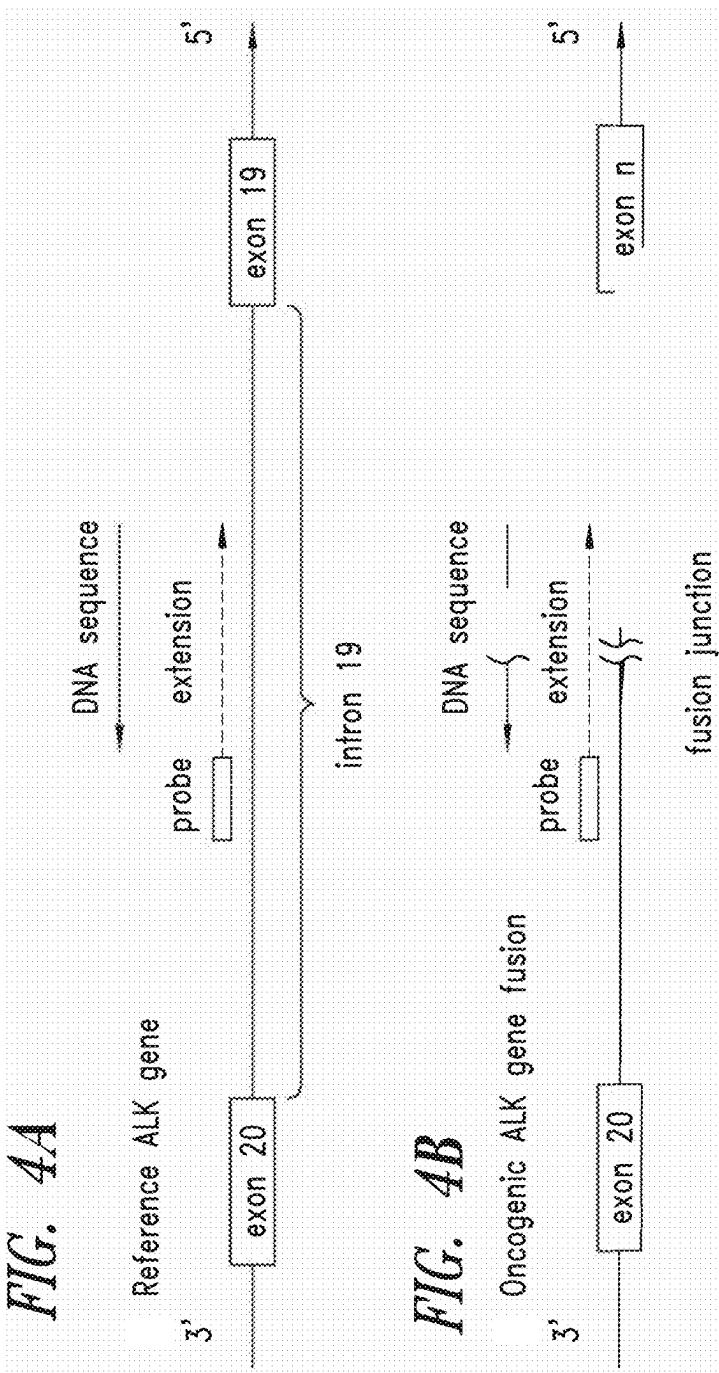


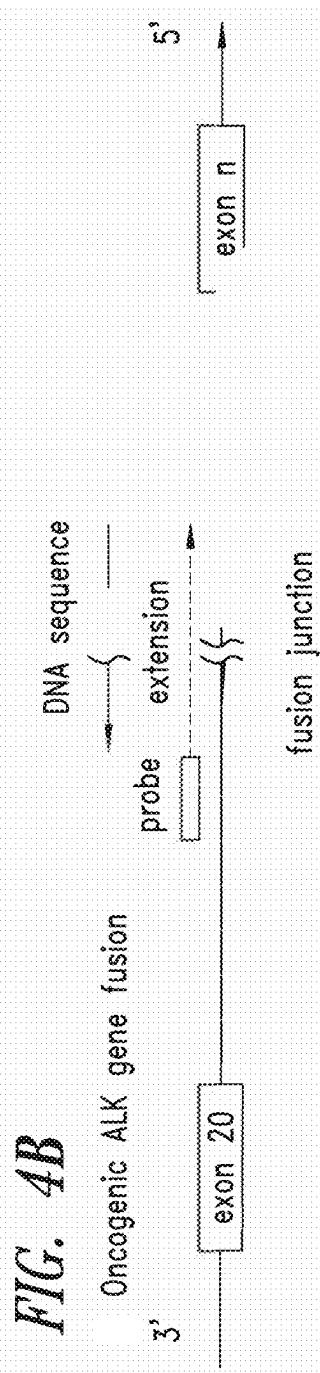
FIG. 2



*FIG. 4A*



*FIG. 4B*



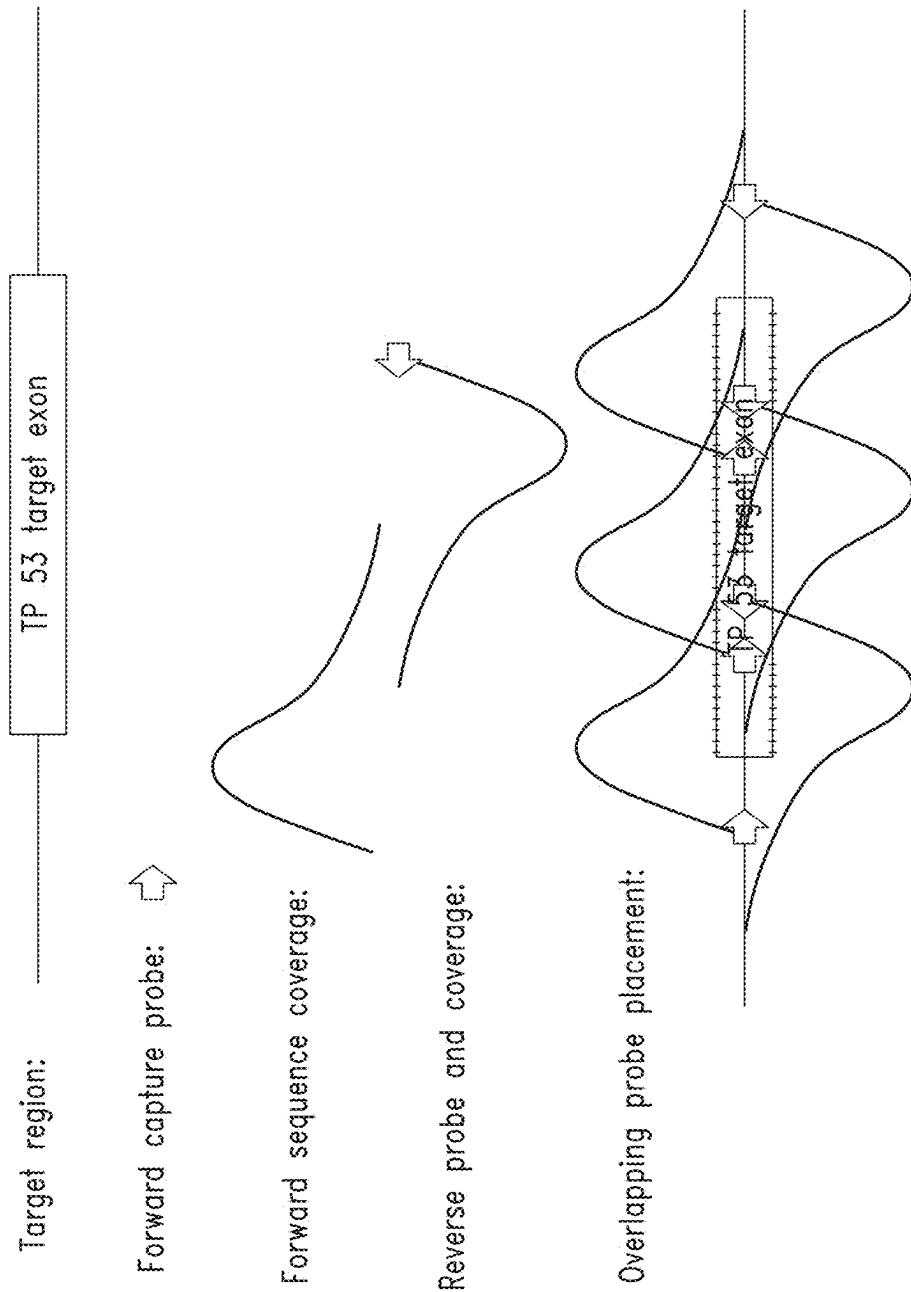
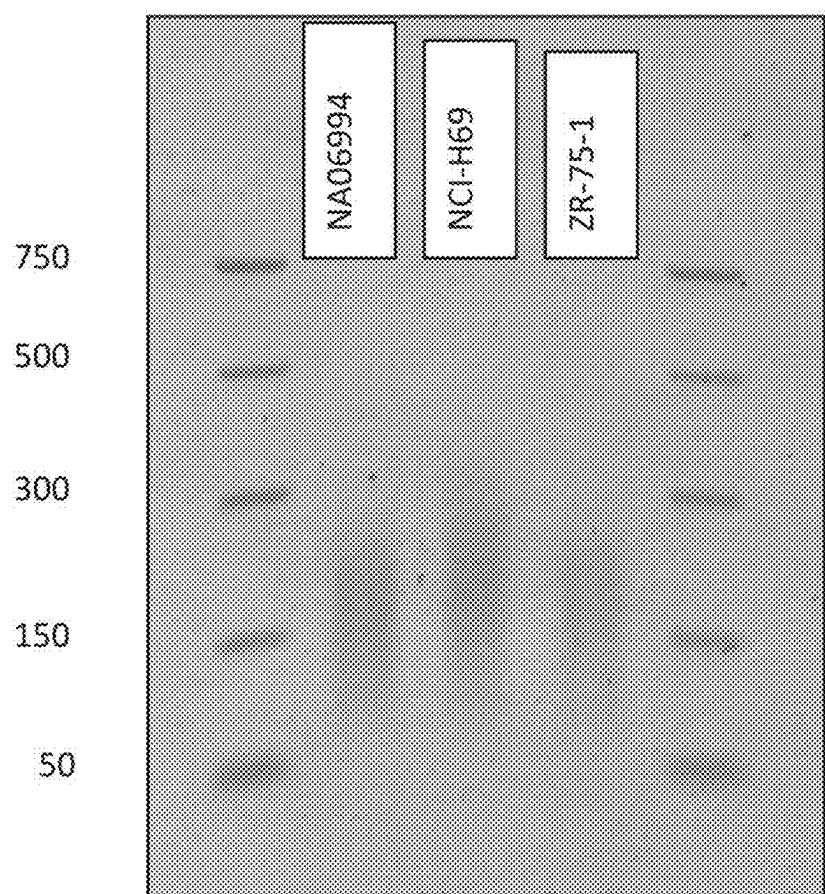


FIG. 5



*FIG. 6*

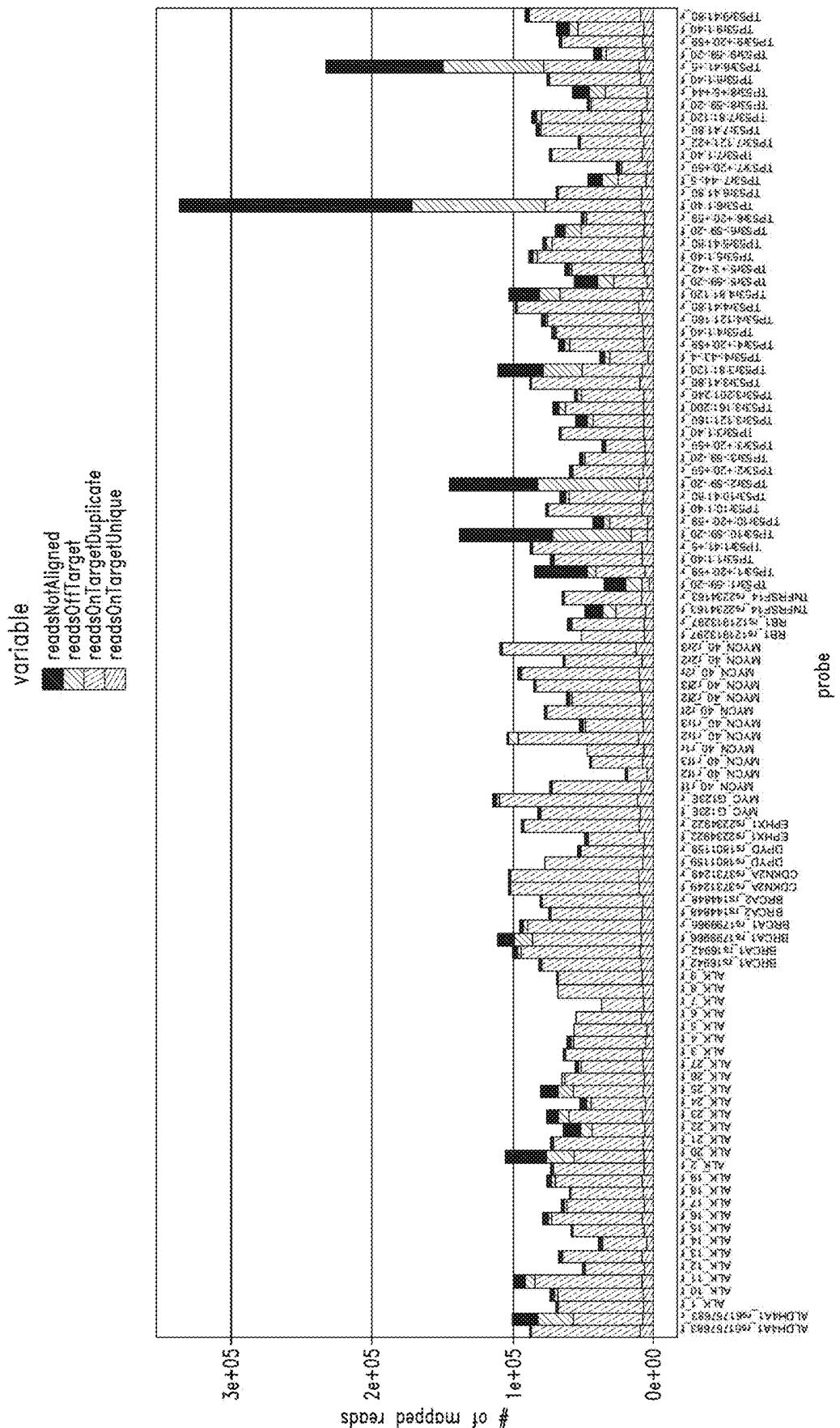


FIG. 7

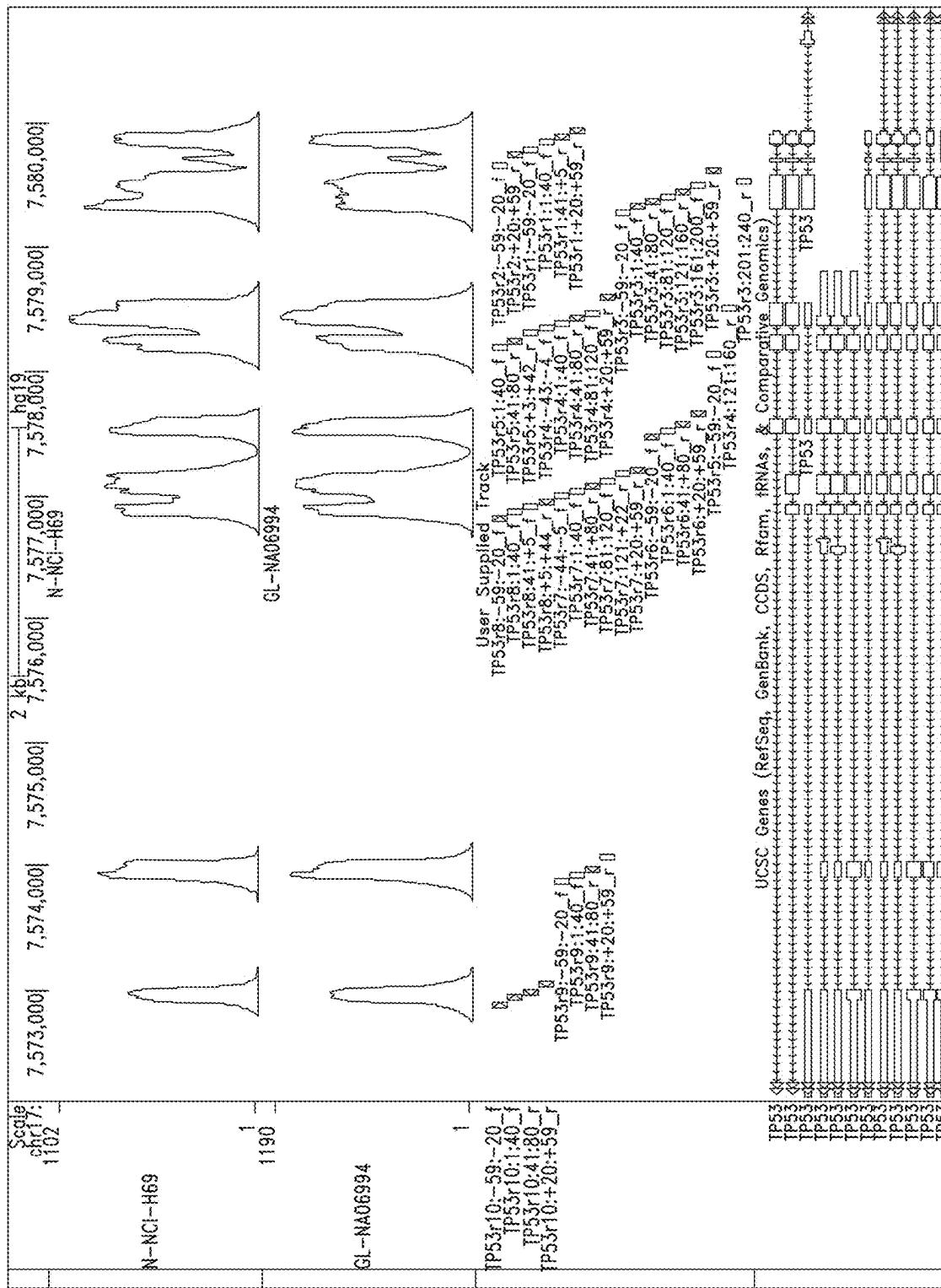


FIG. 8

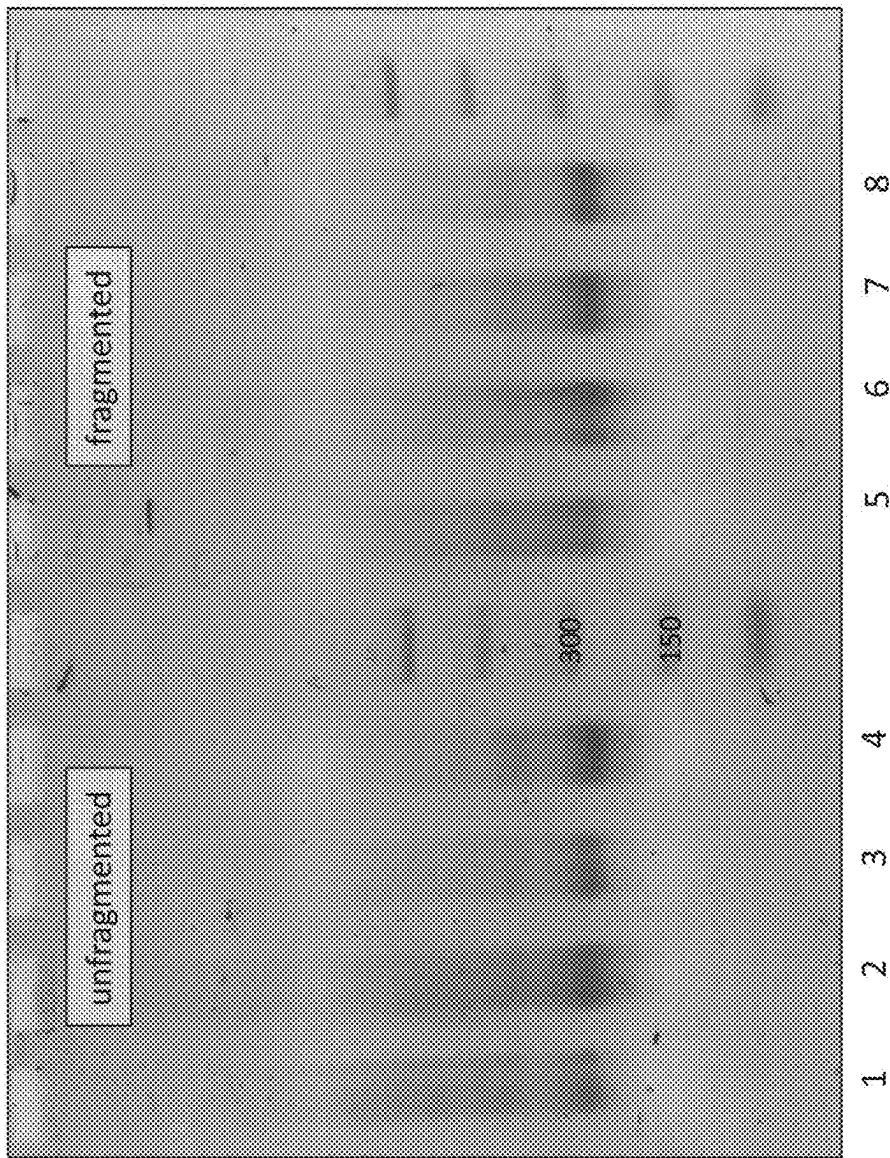


FIG. 9A

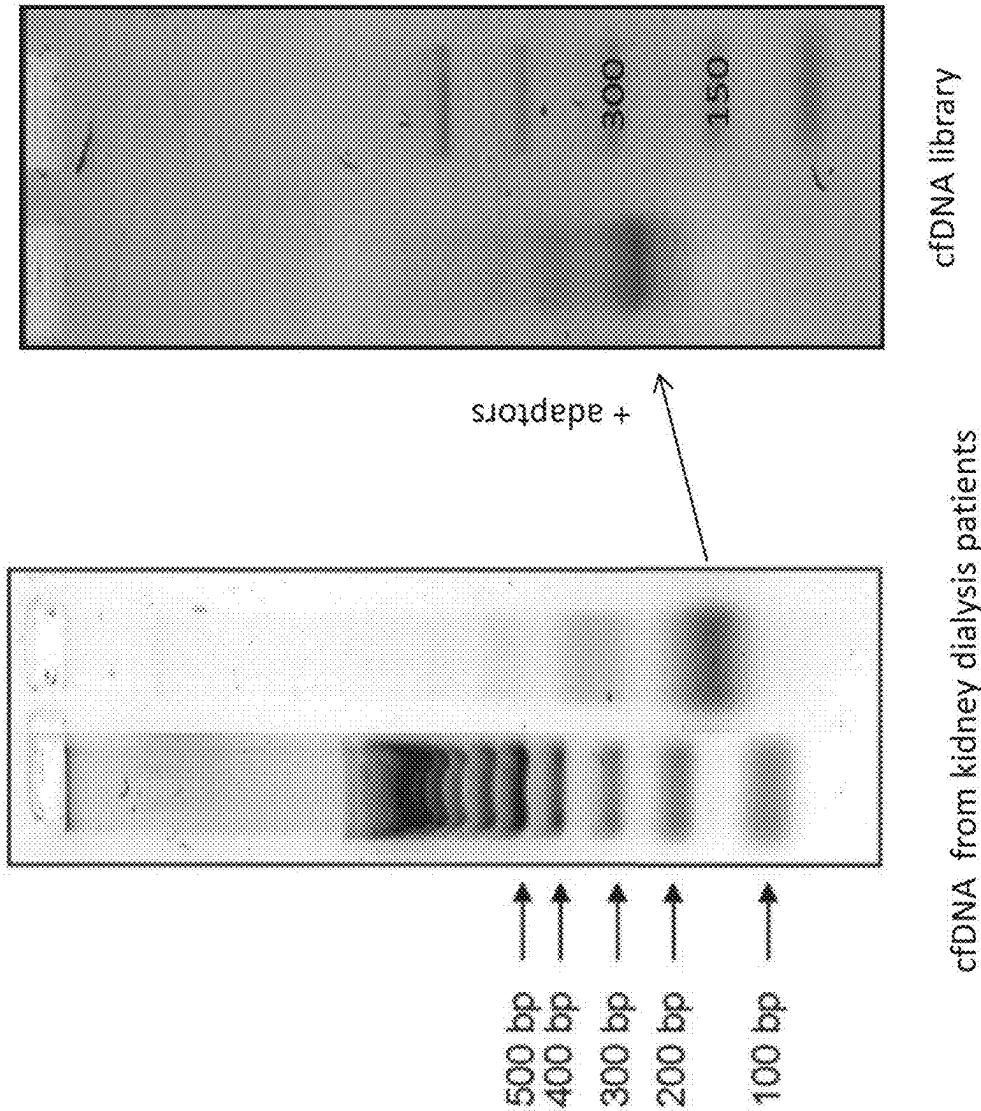
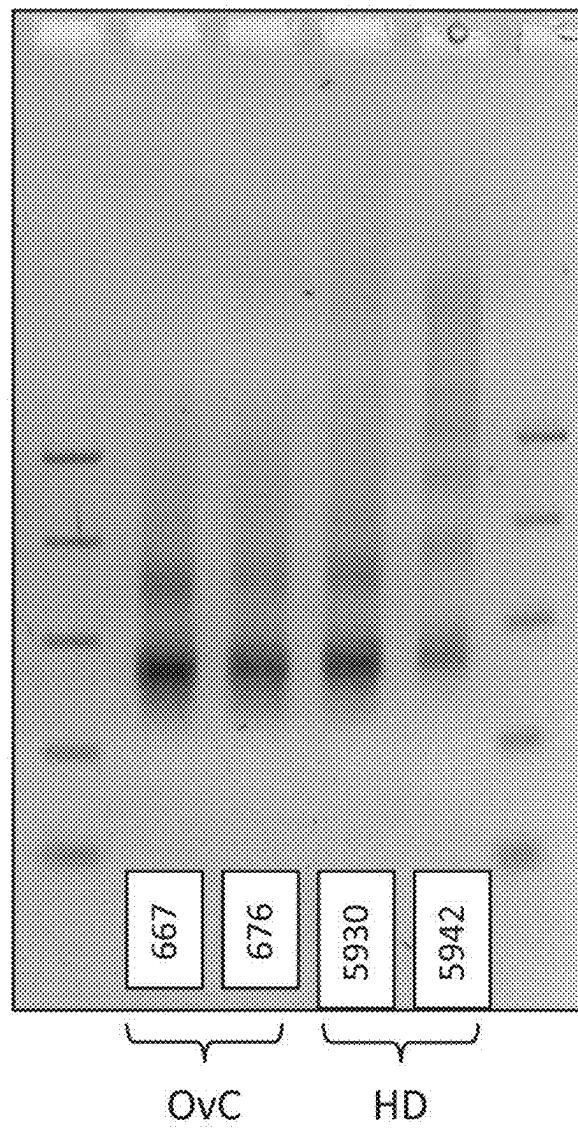


FIG. 9B



*FIG. 9C*

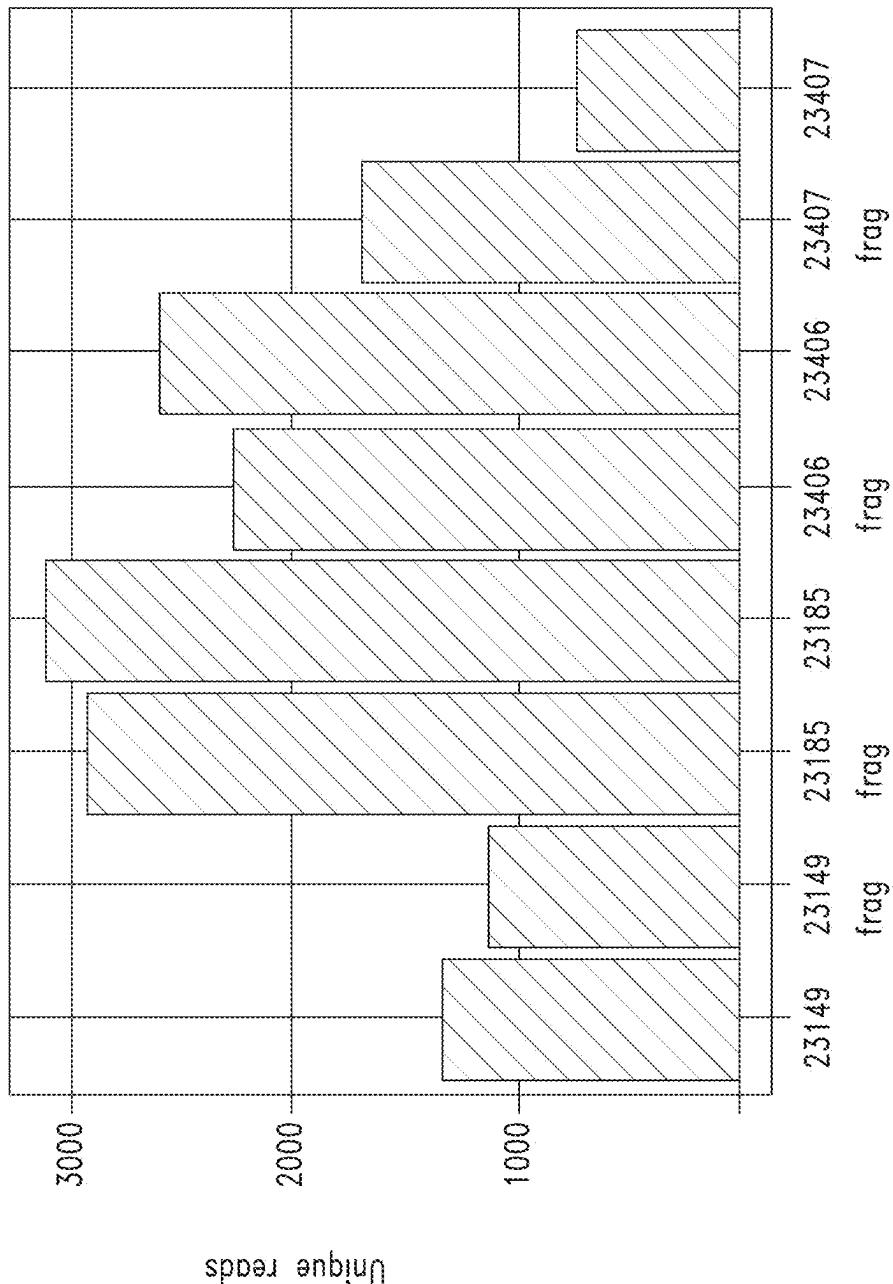


FIG. 10

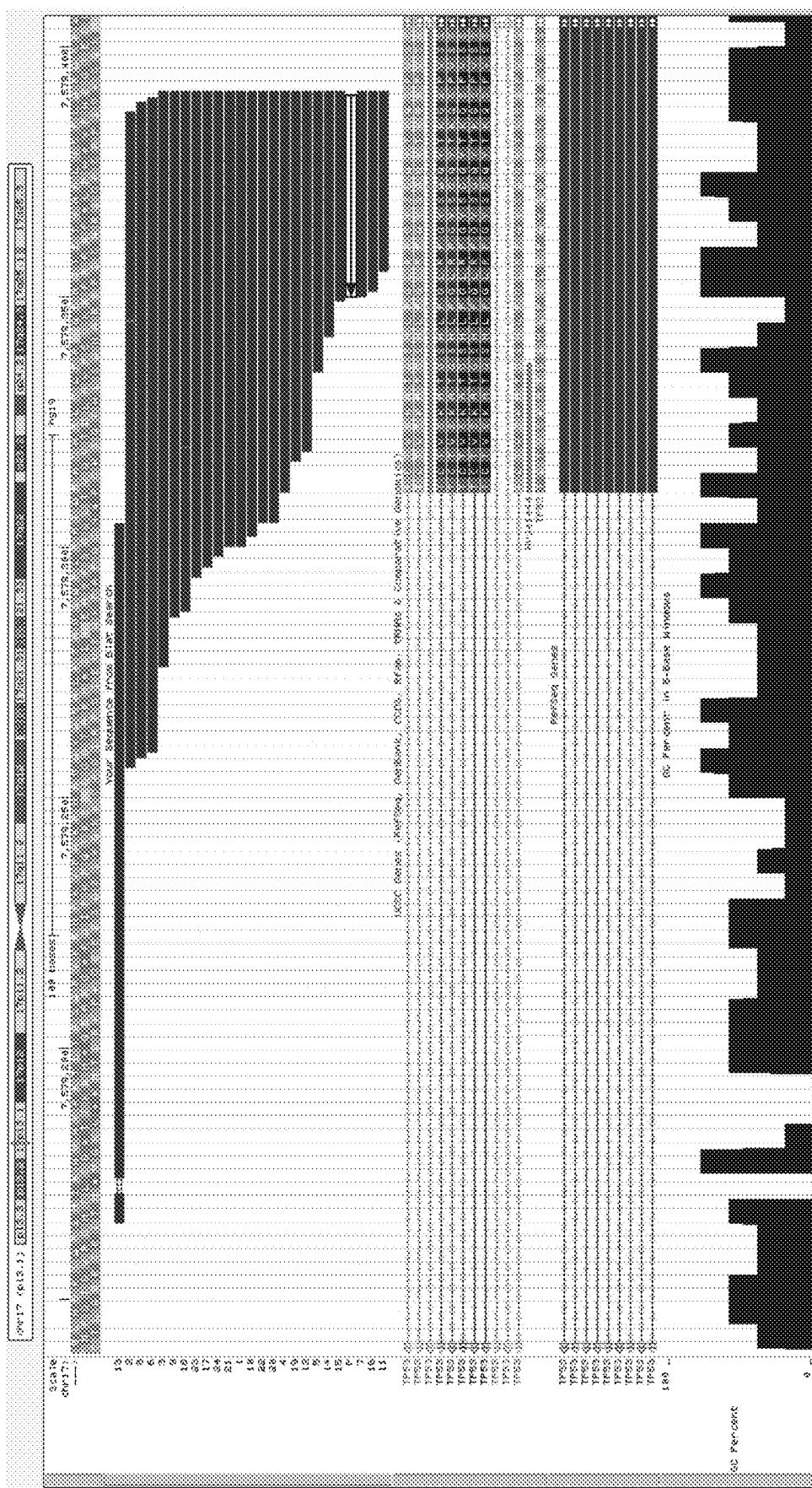


FIG. 11

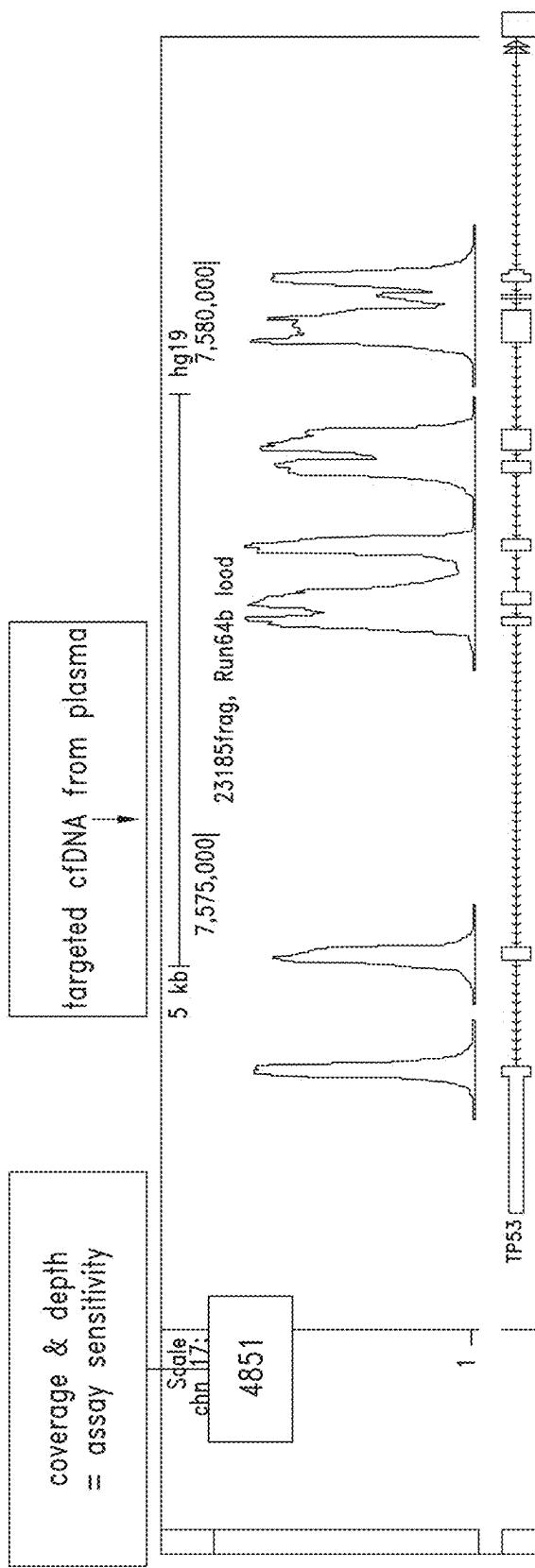
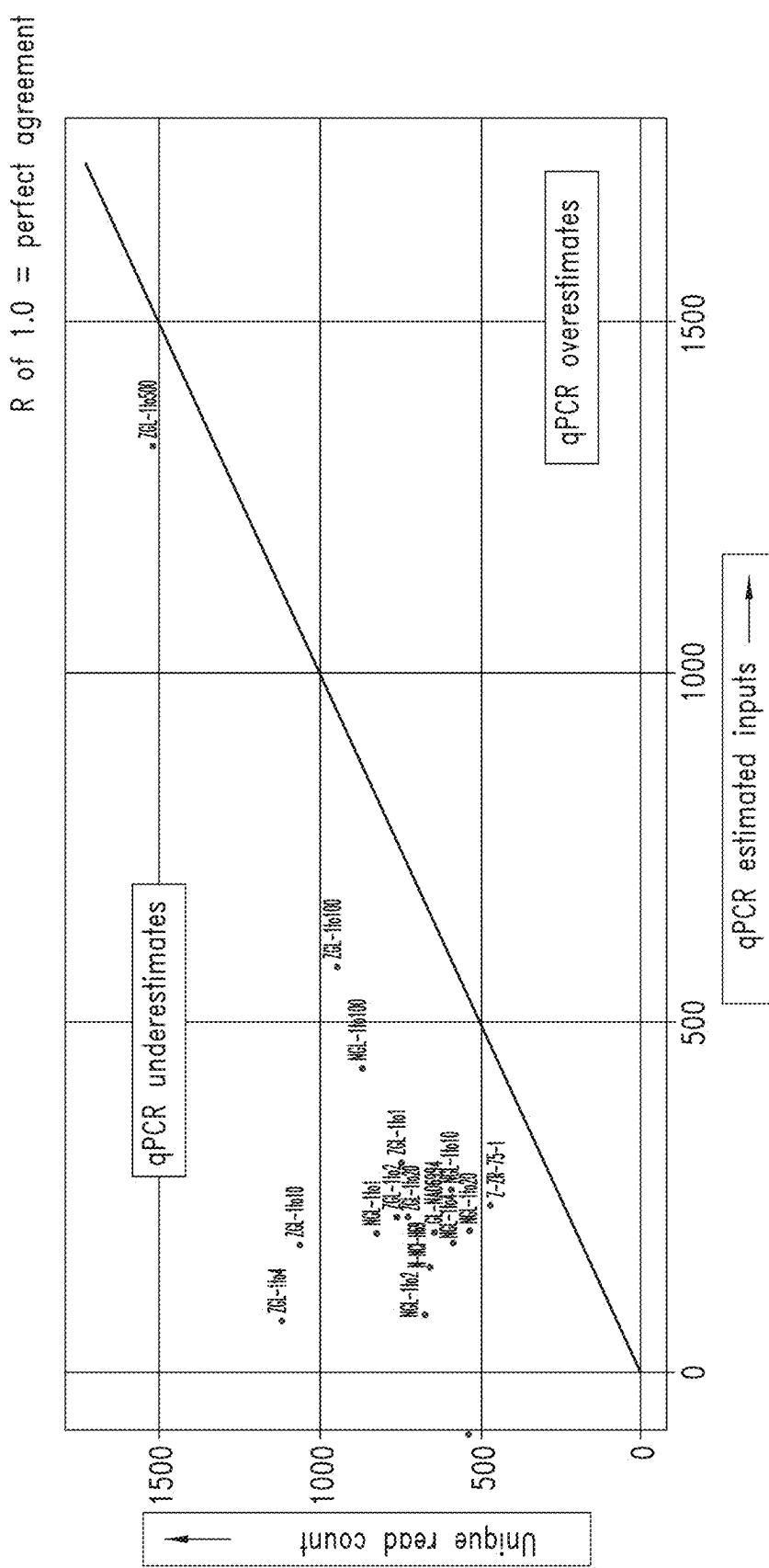
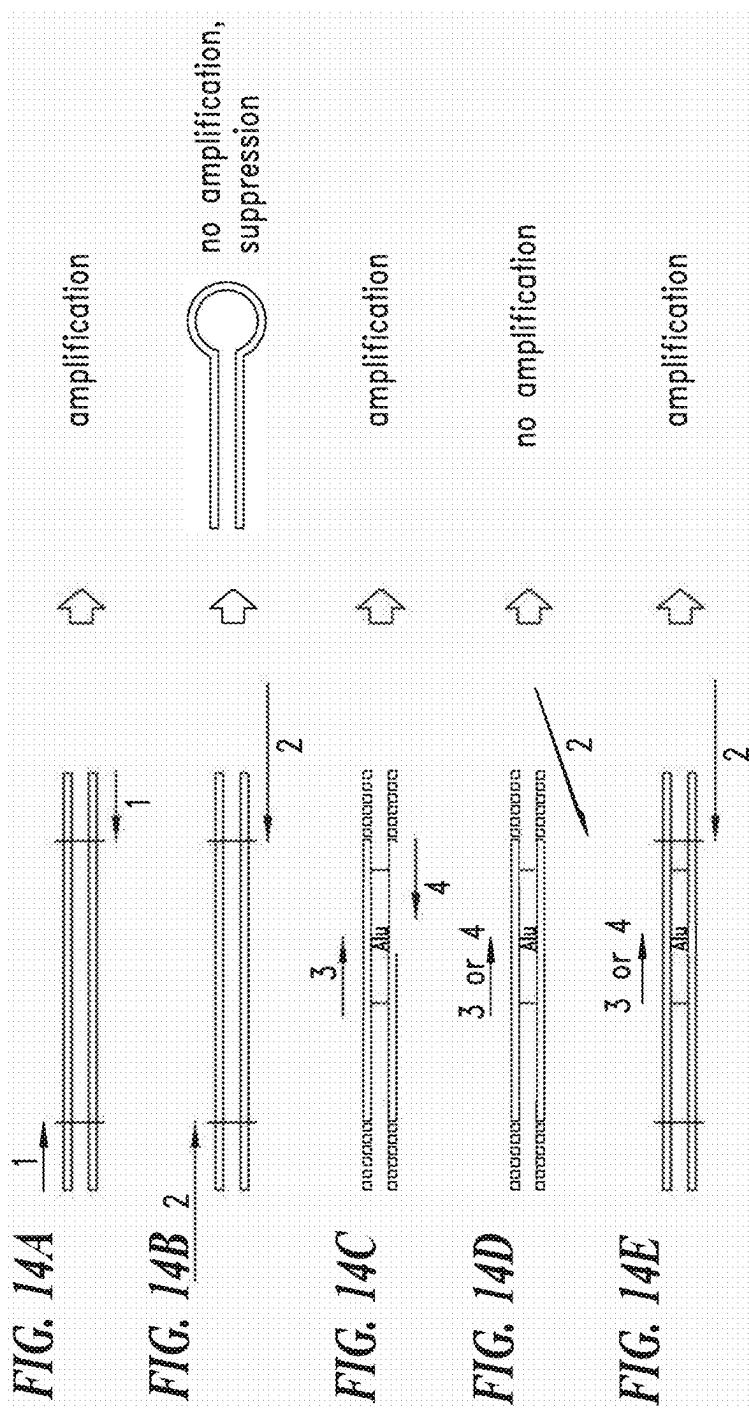


FIG. 12



13  
FIG:



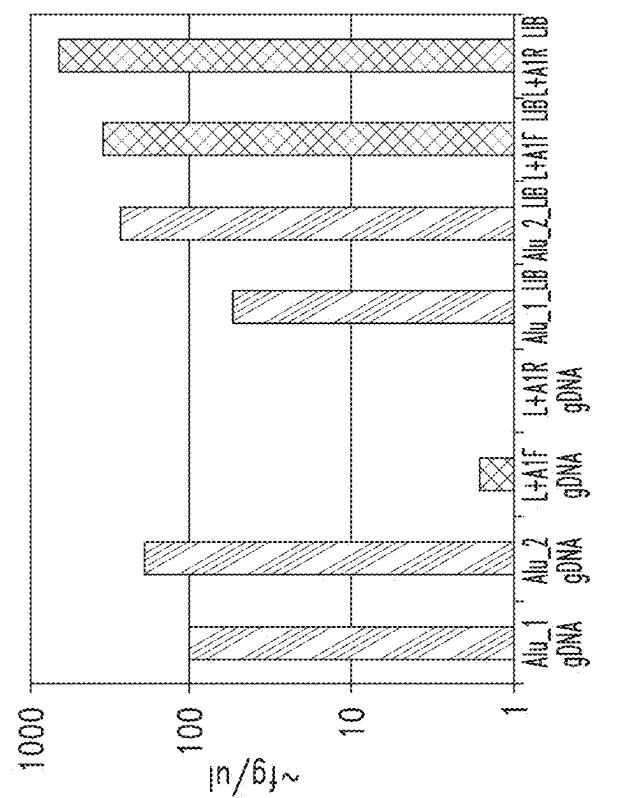


FIG. 15B

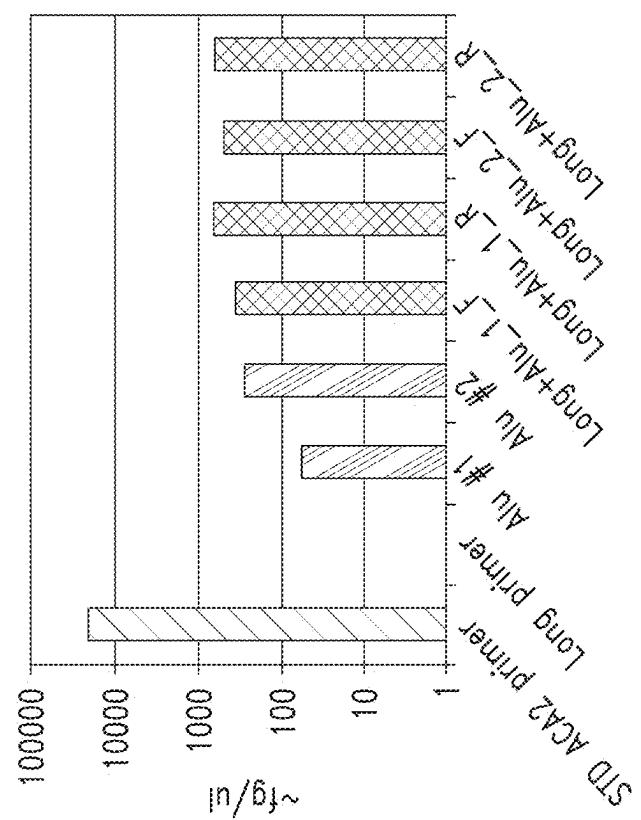


FIG. 15A

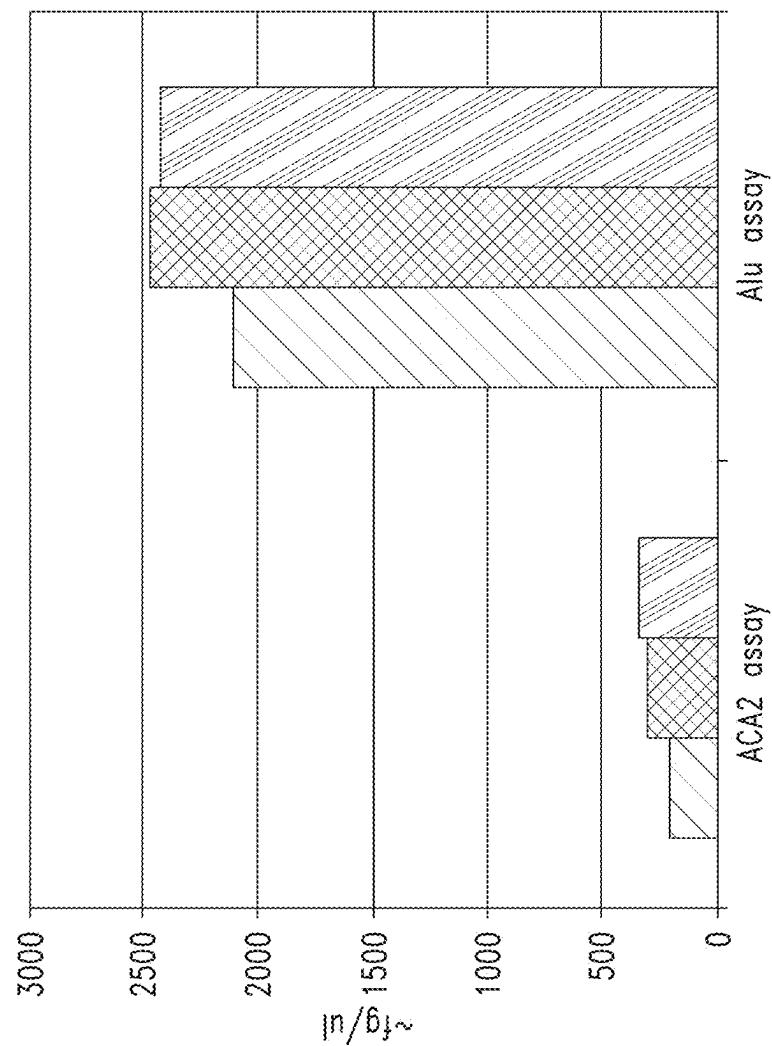
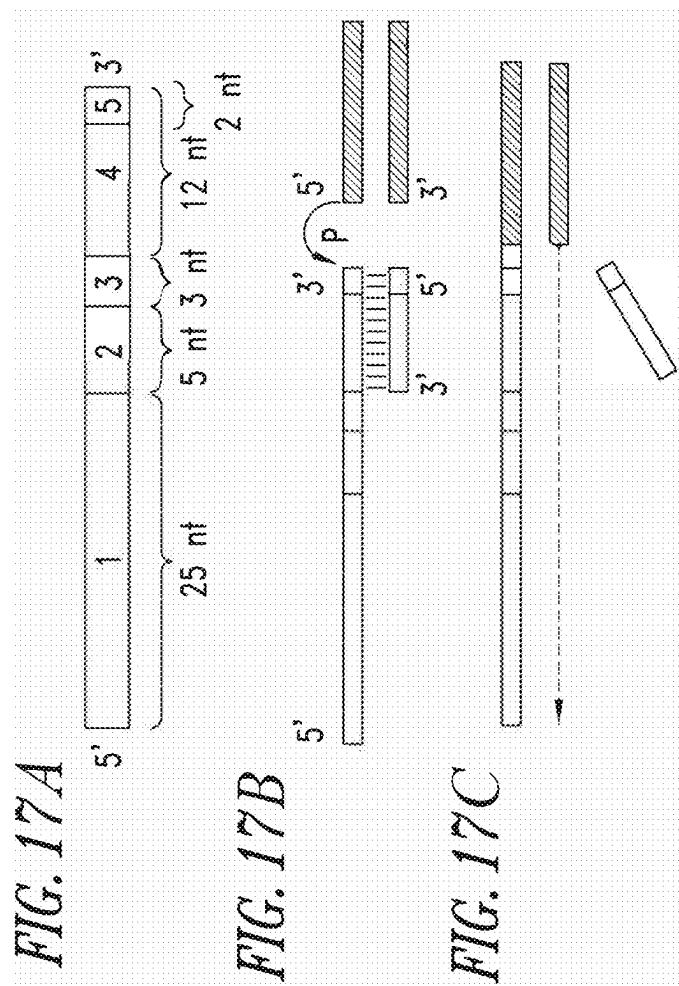
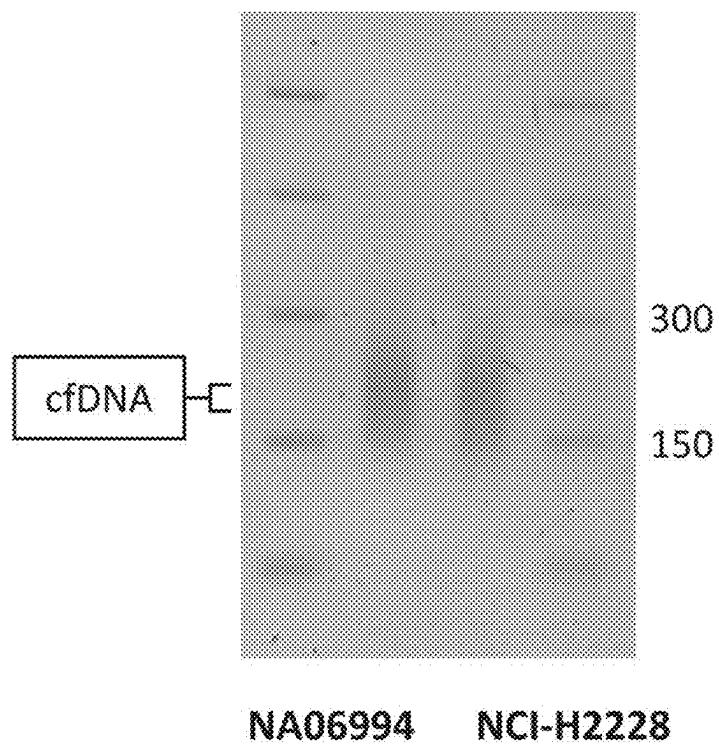


FIG. 16





*FIG. 18*

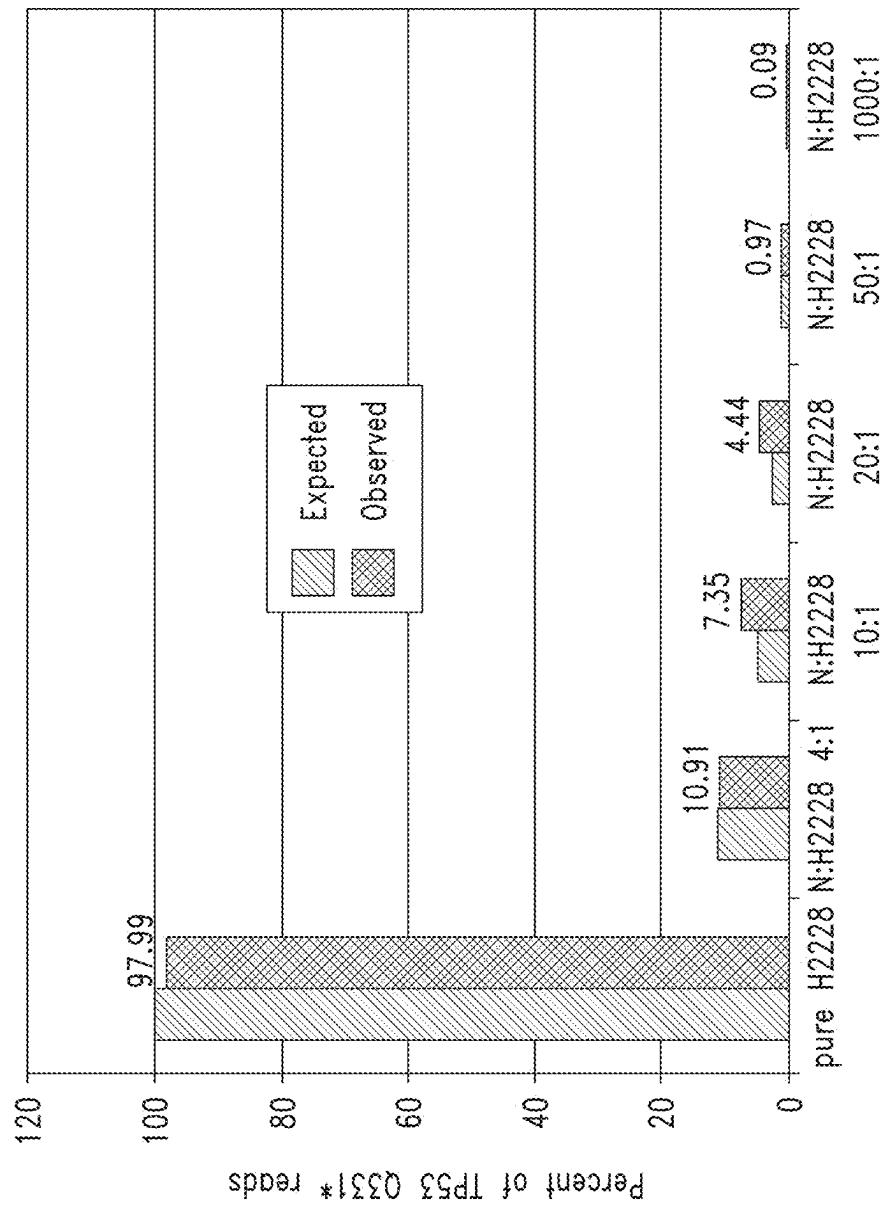
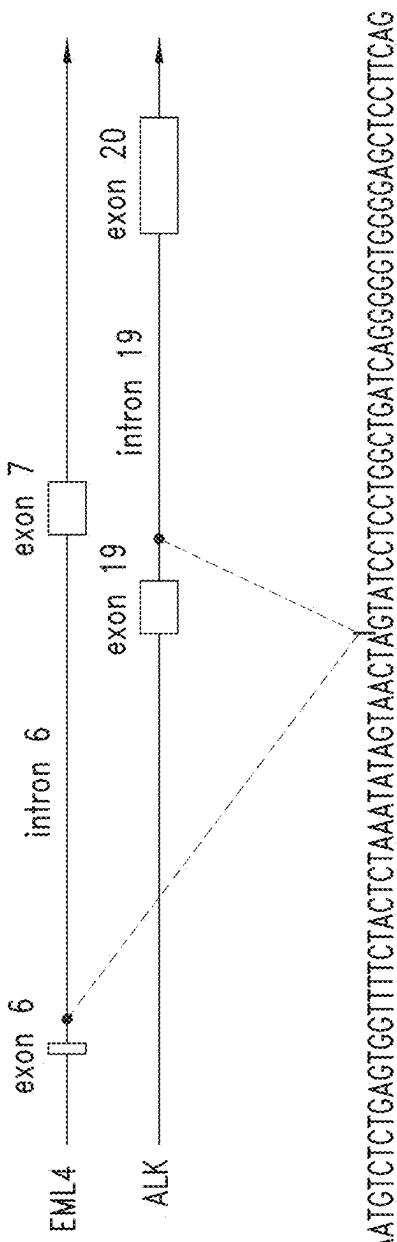


FIG. 19



AATGCTCTGAGTCCTTCACTCTAAATAACTAAGTATCCCTGGCTGGGGAGCTCCCTCAG

FIG. 20

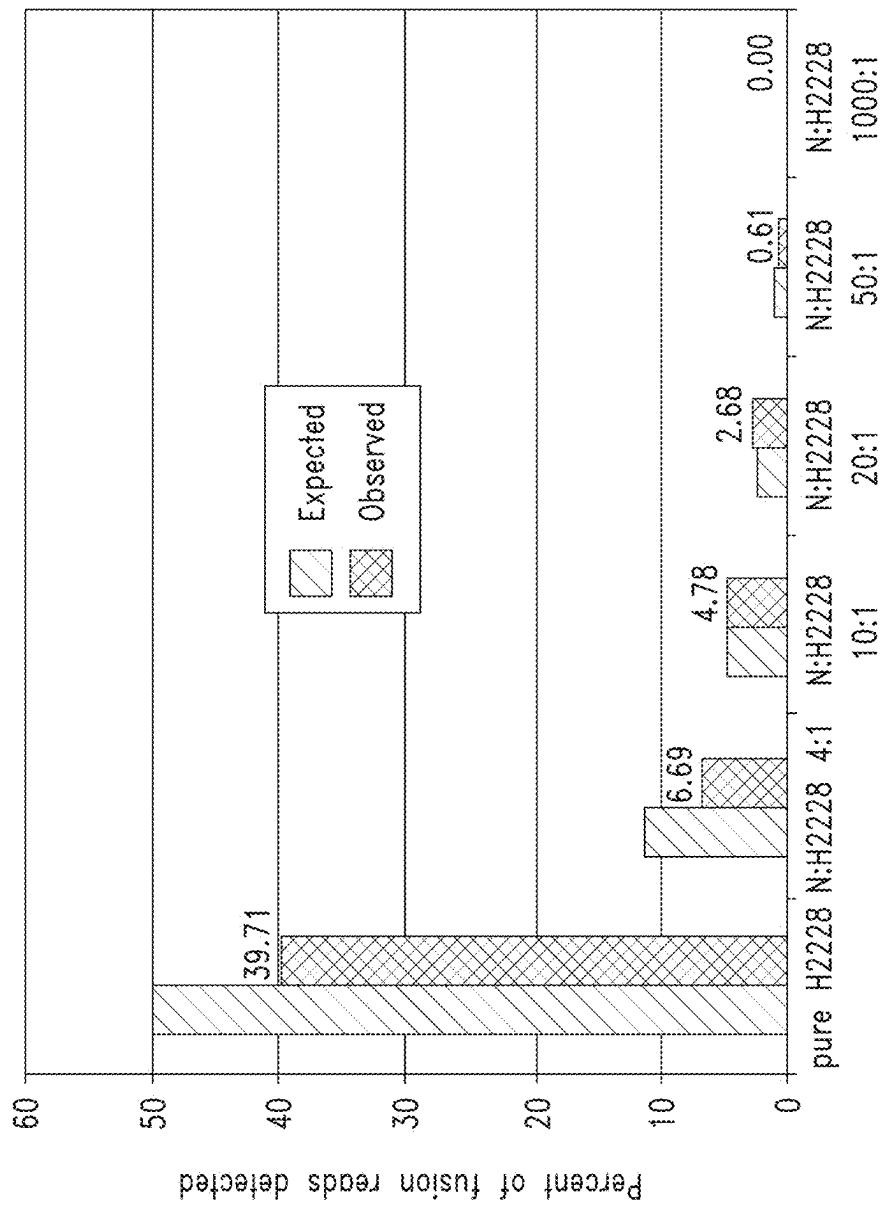


FIG. 21

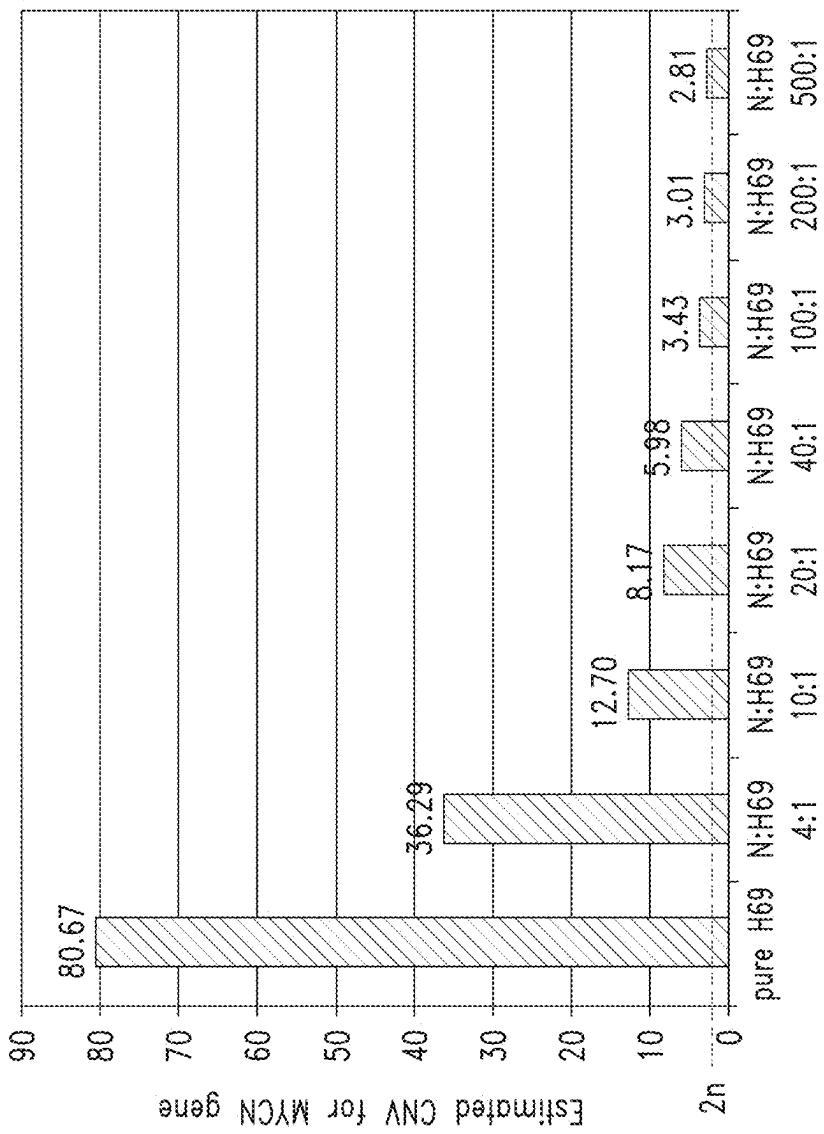


FIG. 22

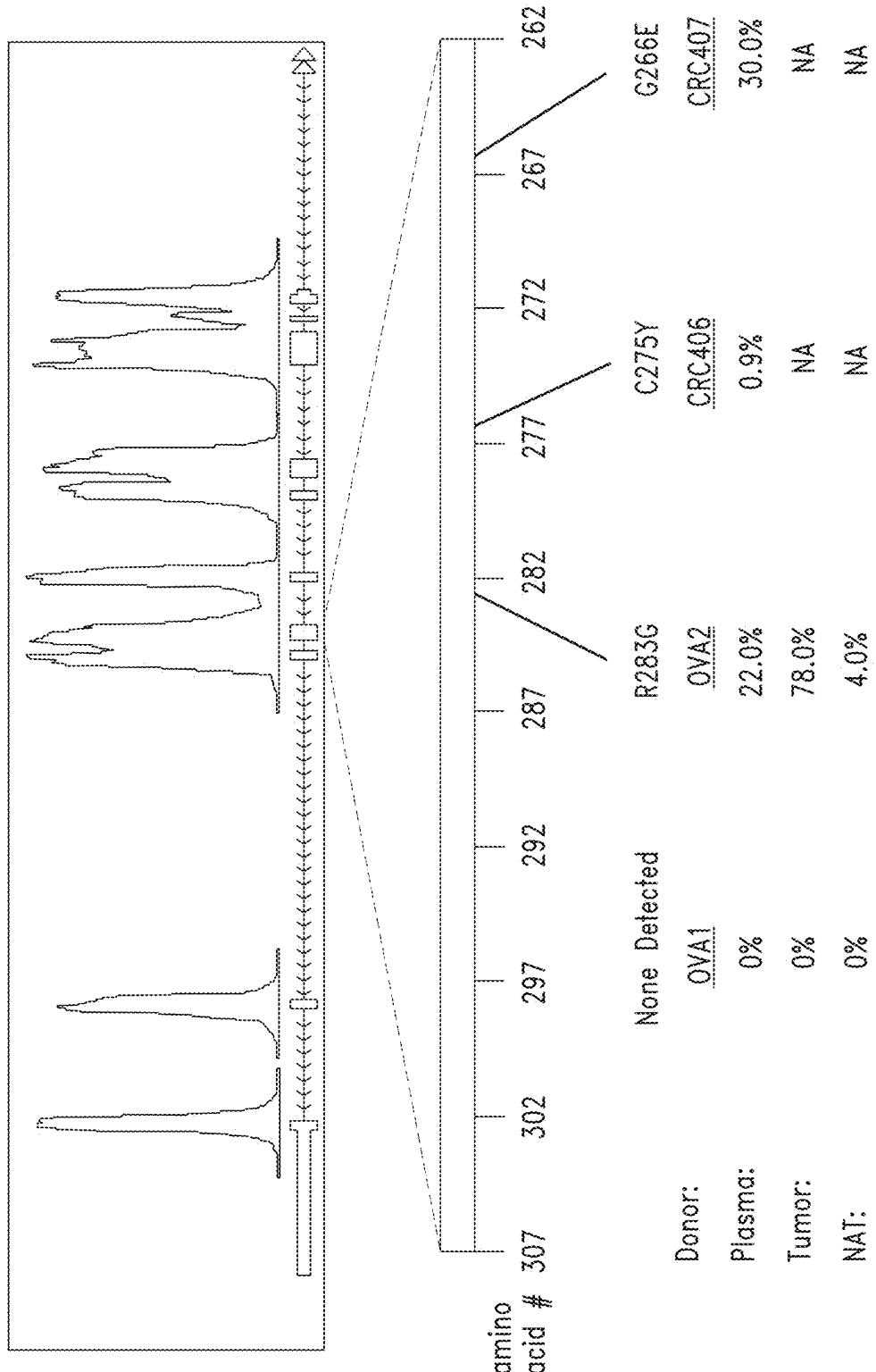


FIG. 23

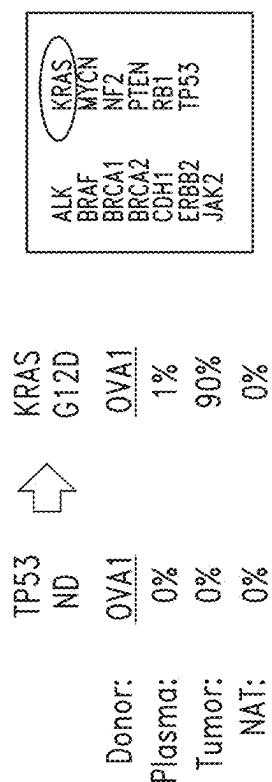


FIG. 24

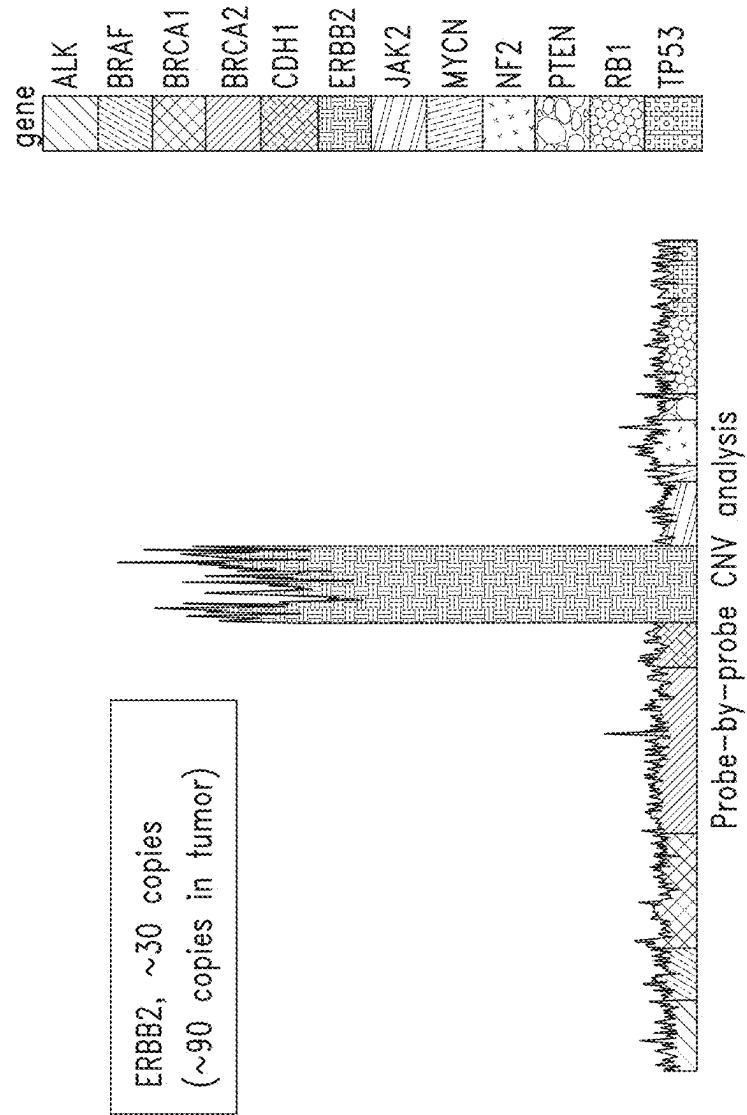


FIG. 25

## METHODS FOR QUANTITATIVE GENETIC ANALYSIS OF CELL FREE DNA

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation of U.S. patent application Ser. No. 18/983,944, filed Dec. 17, 2024, which is a continuation of U.S. patent application Ser. No. 15/727,887, filed Oct. 9, 2017, now allowed, which is a continuation of U.S. patent application Ser. No. 14/466,741, filed Aug. 22, 2014, now abandoned, each of which is incorporated herein by reference in its entirety.

### STATEMENT REGARDING THE SEQUENCE LISTING

[0002] The text of the computer readable sequence listing filed herewith, titled “42436-304\_SEQUENCE\_LISTING”, created Apr. 29, 2025, having a file size of 2,480,219,120 bytes, is hereby incorporated by reference in its entirety.

### BACKGROUND OF THE INVENTION

#### Technical Field

[0003] The invention relates generally to compositions and methods for the quantitative genetic analysis of cell free DNA (cfDNA). In particular, the present invention relates to improved targeted sequence capture compositions and methods for the genetic characterization and analysis of cfDNA.

#### Description of the Related Art

[0004] It is becoming increasing clear that most, if not all, of the most common human cancers are diseases of the human genome. The emerging picture is that somatic mutations accumulate during an individual's lifetime, some of which increase the probability that the cell in which they are harbored can develop into a tumor (Vogelstein et al., *Science* 339(6127): 1546-1558 (2013)). With just the wrong combination of accumulated mutational events, a precancerous growth loses constraints that keep uncontrolled proliferation in check and the resulting cell mass becomes a cancer. The constellations of mutations that are necessary and sufficient to cause cancer are often collectively referred to as “driver mutations.” One of the themes that have emerged from recent and intensive molecular analysis is that cancer, once thought of as a single, tissue-specific disease, is in fact a group of related diseases, each with a unique molecular pathology. The human genome project laid the groundwork for genome-wide analysis of cancers.

[0005] For example, the introduction of next-generation sequencing technologies (2004-present) has accelerated the discovery pace of causal genomic factors that underlie the diagnosis of NSCLC, making it clear that NSCLC is really a family of related diseases, each of which may be responsive to a different targeted therapy.

[0006] The art lacks reliable and robust molecular analysis methods for the analysis of genetic diseases. Traditionally, molecular diagnostics have consisted of antibody-based tests (immunohistochemistry), in-situ hybridization with DNA probes (fluorescence in situ hybridization), and hybridization or PCR-based tests that query specific nucleotide sequences. Until recently, DNA sequencing as a molecular diagnostic tool has been generally limited to the coding exons of one or two genes. While DNA sequencing

has been used in the diagnosis and treatment of solid cancers, one of the most significant drawbacks of these methods is that they require direct access to tumor tissues. Such material is often difficult to obtain from the initial biopsy used to diagnose the disease and virtually impossible to obtain in multiple repetitions over time. Similarly, biopsies are not possible in patients with inaccessible tumors and not practical in individuals suffering from metastatic disease.

[0007] Thus, the vast potential of molecular diagnostics for genetic diseases; fetal testing; paternity testing; predicting response to drug treatment; diagnosing or monitoring a medical condition; mendelian disorders; genetic mosaicism; pathogen screening; microbiome profiling; and organ transplant monitoring; has yet to be realized. To date, existing molecular diagnostics approaches lack efficient solutions to clone and amplify individual DNA molecules, as well as solutions to efficiently target sequencing to specific genomic loci, with sensitivity sufficient to discriminate true positive test results from false positive signals that arise during sample processing.

### BRIEF SUMMARY OF THE INVENTION

[0008] The invention relates generally to compositions and methods for improved compositions and methods for the genetic analysis of cfDNA.

[0009] In various embodiments, a method for genetic analysis of cell-free DNA (cfDNA) is provided, comprising: treating cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA; ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate cfDNA library clones; determining the number of genome equivalents in the cfDNA clone library; and performing a quantitative genetic analysis of one or more target genetic loci in the cfDNA library clones.

[0010] In a particular embodiment, the method further comprises isolating cfDNA from a biological sample of a subject.

[0011] In an additional embodiment, the cfDNA is isolated from a biological sample selected from the group consisting of: amniotic fluid, blood, plasma, serum, semen, lymphatic fluid, cerebral spinal fluid, ocular fluid, urine, saliva, stool, mucous, and sweat.

[0012] In a certain embodiment, the one or more adaptors comprise a plurality of adaptor species.

[0013] In a particular embodiment, the one or more adaptors each comprise a primer binding site for amplification of the cfDNA library.

[0014] In a further embodiment, the one or more adaptors each comprise one or more unique read codes.

[0015] In an additional embodiment, the one or more adaptors each comprise one or more sample codes for sample multiplexing.

[0016] In another embodiment, the one or more adaptors each comprise one or more sequences for DNA sequencing.

[0017] In a particular embodiment, qPCR is performed on the cfDNA clone library and a qPCR measurement is compared to standards of known genome equivalents to determine the genome equivalents of the cfDNA clone library.

[0018] In another particular embodiment, qPCR is performed with a primer that binds to an Alu sequence and a primer that binds to a sequence in an adaptor.

- [0019] In a certain embodiment, the quantitative genetic analysis is performed on a plurality of genetic loci in the cfDNA library clones.
- [0020] In a further embodiment, the quantitative genetic analysis is performed on a plurality of genetic loci in a plurality of cfDNA clone libraries.
- [0021] In an additional embodiment, the quantitative genetic analysis comprises hybridizing one or more capture probes to a target genetic locus to form capture probe-cfDNA clone complexes.
- [0022] In a particular embodiment, the quantitative genetic analysis comprises isolating the capture probe-cfDNA clone complexes.
- [0023] In a certain embodiment, the quantitative genetic analysis comprises amplification of the cfDNA clone sequence in the isolated hybridized capture probe-cfDNA clone complexes.
- [0024] In a further embodiment, the quantitative genetic analysis comprises DNA sequencing to generate a plurality of sequencing reads.
- [0025] In another embodiment, the quantitative genetic analysis comprises bioinformatic analysis of the plurality of sequencing reads.
- [0026] In a particular embodiment, bioinformatics analysis is used: to quantify the number of genome equivalents analyzed in the cfDNA clone library; to detect genetic variants in a target genetic locus; to detect mutations within a target genetic locus; to detect genetic fusions within a target genetic locus; and to measure copy number fluctuations within a target genetic locus.
- [0027] In an additional embodiment, the subject does not have a genetic disease.
- [0028] In a certain embodiment, the subject has not been diagnosed with a genetic disease.
- [0029] In another certain embodiment, the subject has been diagnosed with a genetic disease.
- [0030] In another embodiment, the quantitative genetic analysis is used to identify or detect one or more genetic lesions that cause or associated with the genetic disease.
- [0031] In a certain embodiment, the genetic lesion comprises a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion.
- [0032] In a particular embodiment, the genetic lesion comprises a genomic rearrangement that fuses the 3' coding region of the ALK gene to another gene.
- [0033] In a particular embodiment, the 3' coding region of the ALK gene is fused to the EML4 gene.
- [0034] In another embodiment, the genetic disease is cancer.
- [0035] In a further embodiment, the subject is pregnant.
- [0036] In an additional embodiment, the quantitative genetic analysis is used to identify or detect one or more genetic variants or genetic lesions of one or more target genetic loci in fetal cfDNA.
- [0037] In a particular embodiment, the subject is a transplant recipient.
- [0038] In an additional embodiment, the quantitative genetic analysis is used to identify or detect donor cfDNA in the subject.
- [0039] In various embodiments, a method of predicting, diagnosing, or monitoring a genetic disease in a subject is provided, comprising: isolating or obtaining cfDNA from a biological sample of a subject; treating the cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA; ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate a cfDNA clone library; determining the number of genome equivalents in the cfDNA clone library; and performing a quantitative genetic analysis of one or more target genetic loci associated with the genetic disease in the cfDNA clone library, wherein the identification or detection of one or more genetic lesions in the one or more target genetic loci is prognostic for, diagnostic of, or monitors the progression of the genetic disease.
- [0040] In an additional embodiment, the cfDNA is isolated from a biological sample selected from the group consisting of: amniotic fluid, blood, plasma, serum, semen, lymphatic fluid, cerebral spinal fluid, ocular fluid, urine, saliva, stool, mucous, and sweat.
- [0041] In a certain embodiment, the genetic lesion comprises a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion.
- [0042] In a particular embodiment, the genetic lesion comprises a genomic rearrangement that fuses the 3' coding region of the ALK gene to another gene.
- [0043] In a further embodiment, the 3' coding region of the ALK gene is fused to the EML4 gene.
- [0044] In a particular embodiment, the genetic disease is cancer.
- [0045] In various embodiments, a companion diagnostic for a genetic disease is provided comprising: isolating or obtaining cfDNA from a biological sample of a subject; treating the cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA; ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate a cfDNA clone library; determining the number of genome equivalents in the cfDNA clone library; and performing a quantitative genetic analysis of one or more biomarkers associated with the genetic disease in the cfDNA clone library, wherein detection of, or failure to detect, at least one of the one or more biomarkers indicates whether the subject should be treated for the genetic disease.
- [0046] In a particular embodiment, the cfDNA is isolated from a biological sample selected from the group consisting of: amniotic fluid, blood, plasma, serum, semen, lymphatic fluid, cerebral spinal fluid, ocular fluid, urine, saliva, stool, mucous, and sweat.
- [0047] In an additional embodiment, the biomarker is a genetic lesion.
- [0048] In a particular embodiment, the genetic lesion comprises a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion.
- [0049] In an additional embodiment, the genetic lesion comprises a genomic rearrangement that fuses the 3' coding region of the ALK gene to another gene.
- [0050] In a further embodiment, the 3' coding region of the ALK gene is fused to the EML4 gene.
- [0051] In a certain embodiment, the genetic disease is cancer.

#### BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

- [0052] FIG. 1 shows the expected versus observed variant frequencies as a function of admix dilution in the absence of

unique read filtering. In the absence of unique read filtering, random base changes at these four selected positions occurred with measurable, non-zero frequencies; thus, demonstrating a lack of sensitivity to detect the particular single nucleotide variants (SNV).

[0053] FIG. 2 shows that unique read filtering performed on the data generated in FIG. 1. The left hand panel shows the data from FIG. 1 on the BRAF I326T SNV without unique read filtering. The right hand panel shows that using unique read filtering of the same data increased the assay sensitivity and allowed the discrimination of true signal from error-prone noise.

[0054] FIG. 3 shows capture probe performance as a function of length and wash temperature. The y-axis shows the total number of reads associated with each capture probe. The bars in the bar chart are broken into two categories, where open bars correspond to on-target reads that align to the intended capture probe targets and solid bars show off-target reads that are associated with a capture probe but that map to unintended regions of the genome. Overall, the 40-mer and 60-mer capture probes perform substantially the same with 44° C. and 47° C. washes. With the 50° C. wash, the 40-mer capture probes perform erratically. These data validate the use of 40-mer capture probes at wash temperatures in the range of about 44° C. to about 47° C.

[0055] FIG. 4A-FIG. 4B shows a schematic for the targeted and oriented sequencing of intron 19 of the ALK gene. In the “wild-type” reference sequence, antisense-oriented ALK capture probes identify sequences from intron 19 (FIG. 4A). In the case of pathogenic fusion genes, some ALK capture probes will identify junction sequences associated with the gene fusion event (FIG. 4B).

[0056] FIG. 5 shows a schematic for high density capture probe placement for complete sequencing of target regions. Each capture probe captures a collection of sequences that provide cumulative coverage at each base position. Here, coverage is represented by a line, and the amplitude of the line denotes depth of coverage derived from a particular capture probe. Overlapping coverage from adjacent capture probes provides complete sequencing of target regions in both possible directions. In addition, the head-to-head placement of opposite strand capture probes ensures that all capture probe binding sites are sequenced.

[0057] FIG. 6 shows a representative example of the size distribution of fragmented DNA used in library construction.

[0058] FIG. 7 shows the performance of high-density 40-mer capture probes in a representative experiment. The y-axis shows the total number of reads, which are broken out as on-target reads, off-target reads, and unmappable reads. The x-axis enumerates each of the 105 capture probes used in this experiment for sequence capture.

[0059] FIG. 8 shows a representative example of the cumulative coverage of a target region using high density 40-mer capture probes. Shown here is the cumulative coverage of TP53 coding exons.

[0060] FIG. 9A-FIG. 9C show representative examples of cfDNA libraries. FIG. 9A shows a representative example of the size distribution of cell-free DNA (cfDNA) libraries. The dominant band is consistent with a collection of  $170 \pm 10$  bp fragments ligated to 90 bp of adaptors. FIG. 9B shows a published gel image of cfDNA and a representative cfDNA library generated using the methods disclosed and/or contemplated herein. The qualitative “ladder” appearance is conserved in the library, but the library is shifted to higher

mass by the addition of 90 bp of adaptor sequences. FIG. 9C shows a representative example of genomic, plasma-derived cfDNA libraries from Ovarian cancer patients (OvC) and “healthy donors” (HD).

[0061] FIG. 10 shows the unique read counts across eight cfDNA libraries derived from four plasma samples. Fragmentation (frag) prior to library construction with this sample 23407 increased the library yield by more than two-fold.

[0062] FIG. 11 shows the representative read coverage of cfDNA across a region of the TP53 gene. Twenty four 131 bp reads captured by the “TP53\_NM\_000546\_chr17:7579351:region\_3:280nt:41:80:r” capture probe (SEQ ID NO: 201) were chosen at random and aligned using the BLAT algorithm within the UCSC genome browser. Twenty one reads map to the target region, and they do so in a pattern of overlapping coverage. The probe used to capture these reads is marked with an arrow.

[0063] FIG. 12 shows an overview of targeted DNA sequencing of the coding regions of the TP53 gene from a cfDNA genomic library. The coverage (horizontal axis) extends across all 10 coding regions and includes intronic regions involved in mRNA splicing. The sequencing depth (vertical axis) reaches a maximum of 4851 and is uniform across all coding exons.

[0064] FIG. 13 shows a plot of unique read counts versus qPCR estimated genome equivalents in an ACA2-based assay. qPCR measurements are shown on the X-axis versus read counts on the Y-axis. Perfect agreement between these measurements is shown as the diagonal. There is very poor, if any, correlation between measurements, especially at lower genomic inputs. These data show that the ACA2-based qPCR assay chronically underestimates library complexity and is inadequate for measuring genome equivalents.

[0065] FIG. 14A-FIG. 14E show a schematic of the core elements of a qPCR genome equivalent measurement assay that couples an genomic repeat specific primer (e.g., Alu) and a long adaptor-specific primer. FIG. 14A shows standard library amplification using a single, 25 nt primer named ACA2 (primer 1). FIG. 14B shows longer, 58 nt versions of the ACA2 primer (primer 2) do not amplify genomic libraries because of stem-loop suppression. FIG. 14C shows forward and reverse primers directed to a consensus human Alu repeat element (primers 3 and 4) recognize 1000's of loci and readily amplify genomic DNA. FIG. 14D shows a single Alu primer alone, either forward or reverse (primer 3 or primer 4), coupled with the long ACA2 primer (primer 2) do not amplify genomic DNA. FIG. 14E shows the same primer pair as in FIG. 14D readily amplifies genomic cfDNA library clones that contain Alu sequences.

[0066] FIG. 15A-FIG. 15B show proof-of-concept data for an Alu plus adaptor-based qPCR assay of genome equivalents. FIG. 15A shows amplification of 10 µg of a standard genomic library with various PCR primers. The x-axis specifies PCR primers used for amplification and the Y-axis (log scale) indicates the PCR signal measured in units of fg/µL. The standard ACA2 primer produced a strong signal, as expected. The ACA2 long primer failed to produce signal owing to PCR suppression. The two Alu primer pairs both produced signal at 1% the amount of ACA2, suggesting that 1% of clones possess an amplifiable Alu sequence. The combination of any Alu primer with the long ACA2 primer also produced signal in ~1% of clones.

[0067] FIG. 15B shows validation against 10 µg of genomic DNA (left four samples) or 10 µg of library DNA (right four clones). Alu primer pairs amplify comparable signal from genomic DNA or a genomic library. In contrast, primer pairs consisting of an Alu primer and a long ACA2 primer amplify genomic DNA poorly (L+A1F) or not at all (L+A1R). These same pairs exhibit amplification of library that slightly exceeds the signal from Alu primer pairs.

[0068] FIG. 16 shows a direct comparison of ACA2 primer qPCR assay with the Alu-ACA2 long-primer qPCR assay. The Alu ACA2 long-primer qPCR assay shows an 8-fold increase in detectable genome equivalents, which is more consistent with unique read counts derived from sequencing data.

[0069] FIG. 17A-FIG. 17C show a representative example of adaptor structure and function for high sensitivity, quantitative genetic assays that provide accurate determinations of genome equivalents analyzed. FIG. 17A shows fine structure of the adaptor ligation strand. Details relating to each numbered element are provided in Example 4. FIG. 17B shows the duplex formed between 45 nt ligation strands and 12 nt partner oligo strand creates a blunt-end ligation substrate compatible with end-repaired cfDNA fragments (solid bars). FIG. 17C shows, following ligation, the complement to the ligation strand is created by a DNA polymerase-mediated fill-in reaction.

[0070] FIG. 18 shows a representative example of the size distribution of two DNA samples (NA06994 & NCI-H2228) processed to mimic cfDNA.

[0071] FIG. 19 shows a representative example of the sensitivity of detection of the TP53 point mutation Q331\* in tumor sample DNA (H2228) admixed with normal genomic DNA (N). The most sensitive detection corresponds to ~1 mutant copy of TP53 among 1000 normal copies of the gene.

[0072] FIG. 20 shows the precise determination of the junction sequence for the EML4-ALK fusion gene harbored in cell line NCI-H2228 using the compositions and methods contemplated herein.

[0073] FIG. 21 shows the detection of the EML4-ALK fusion gene tumor sample DNA (H2228) admixed with normal genomic DNA (N). Because the fusion is present as a heterozygote in the NCI-H2228 cell line, the most sensitive detection corresponds to one gene fusion among ~100 normal chromosomal copies of the ALK gene (50 genome equivalents).

[0074] FIG. 22 shows the detection of the MYCN gene amplification in admixtures of cell line NCI-H69 (H69) diluted into normal human DNA (N). The threshold value of two normal diploid copies is shown as a dashed red line.

[0075] FIG. 23 shows the DNA mutations detected in the TP53 gene of three different cancer patients. The canonical gene model is shown at the top of the figure. The peaks represent DNA sequence coverage (X-axis) and depth (Y-axis). Sequencing depth was >4000 genome equivalents for all sample analyzed. An expanded view of exon 7 below the gene model shows where all detected mutations were localized. The frequency of mutant detection in cfDNA (plasma), tumor tissue, and normal adjacent tissue is shown, where available (NA—not available). OVA1 and OVA2 are ovarian cancer patients; CRC406 and CRC407 are colorectal cancer patients. No mutations in TP53 were found in any of the OVA1 samples.

[0076] FIG. 24 shows the DNA sequencing of a larger, thirteen gene panel (boxed). The sequencing identified a KRAS mutation in cfDNA and tumor from ovarian cancer patient OVA1.

[0077] FIG. 25 shows the DNA sequencing of a larger, twelve gene panel. The sequencing identified an ERBB2 gene amplification in the plasma of colorectal cancer patient CRC407.

## DETAILED DESCRIPTION OF THE INVENTION

### A. Overview

[0078] The present invention contemplates, in part, compositions and methods for the quantitative genetic analysis of the genetic state of an individual using cell-free DNA (cfDNA). As used herein, the term “genetic state” refers to the sequence of one or more target genome sequences in the genome in relation to a non-causal normal sequence or in relation to a sequence that is causal for a genetic condition or disease. In one embodiment, analyzing the genetic state refers to identifying, quantifying, or monitoring a genetic variant in a target genetic locus, wherein the variant varies with respect to a reference sequence (e.g., a normal or mutated sequence). The present inventors have provided solutions to the molecular diagnostic problems of genetic conditions or diseases associated with lack of sensitivity to discriminate true positives from false positives, inefficient cloning and amplification of individual DNA molecules, and inefficient targeted sequencing to specific genomic loci. The solutions contemplated herein comprise compositions and methods for reliable and robust quantitative genetic analysis with sensitivity sufficient to discriminate true positive test results from false positive signals that arise during sample processing.

[0079] Next-generation sequencing technology has afforded the opportunity to add broad genomic surveys to molecular diagnosis in a variety of scenarios including cancers, fetal diagnostics, paternity testing, pathogen screening and organ transplant monitoring. In the context of genetic diseases, next-generation sequencing information is being used in a clinical setting to identify mutations within genes that are likely to alter gene function, to identify the gain or loss of genetic material within cells, and to identify genomic rearrangements that are not found in normal, healthy cells. The results of these broad diagnostic surveys are often used to guide patient treatment.

[0080] However, the potential benefits of DNA sequencing in diagnosis and treatment of the genetic state of an individual or genetic conditions or diseases is outweighed by the need to directly access affected tissues to obtain samples. Such material is often difficult to obtain from the initial biopsy used to diagnose the disease and virtually impossible to obtain in multiple repetitions over time. Similarly, in cancer patients, biopsies are not possible in patients with inaccessible tumors and not practical in individuals suffering from metastatic disease. In contrast, the present inventors’ approach derives from the fact that all tissues require access to the vasculature to survive and as a consequence these masses deposit DNA into bodily fluids. One major depot of bodily fluid in which the DNA of diseased cells is found is the plasma of human blood.

[0081] In contrast to testing methods that rely on shallow, genome-wide sequence coverage, molecular diagnostics

contemplated herein for the genetic state of an individual; genetic diseases; mendelian disorders; genetic mosaicism; fetal testing; paternity testing; predicting response to drug treatment; diagnosing or monitoring a medical condition; pathogen screening; microbiome profiling; and organ transplant monitoring leverage the availability of cfDNA to provide deep sequence coverage of select target genes. In addition, the cfDNA-based cancer diagnostics contemplated herein possess the ability to detect a variety of genetic changes including somatic sequence variations that alter protein function, large-scale chromosomal rearrangements that create chimeric gene fusions, and copy number variation that includes loss or gain of gene copies. Using the contemplated compositions and methods, these changes are detectable and quantifiable in the face of significant dilution, or admixture, of normal sequences within cfDNA that are contributed by the normal turnover processes that happen within healthy tissues. The compositions and methods contemplated herein also successfully address the major challenges associated with detecting rare genetic changes causal of disease; namely, that cfDNA is highly fragmented, that cfDNA levels vary substantially between different individuals, and that the degree of admixture of diseased versus normal sequences is highly variable among patients, even within individuals suffering from the same molecular disease and stage.

[0082] In various embodiments, compositions and methods for genetic analysis of comprise interrogating the DNA fraction within biological fluid samples and stool samples. The methods contemplated herein provide a novel comprehensive framework address molecular genetic analysis using cfDNA available from a variety of biological sources. Cloning of purified cfDNA introduces tagged cfDNA sequences that inform downstream analysis and enable amplification of the resulting clone libraries. Hybrid capture with target specific oligonucleotides is used to retrieve specific sequences for subsequent analysis. Independent measurements of the number of genomes present in the library are applied to each sample, and these assays provide a means to estimate the assay's sensitivity. The assays contemplated herein provide reliable, reproducible, and robust methods for the analysis, detection, diagnosis, or monitoring of genetic states, conditions, or disease.

[0083] The practice of particular embodiments of the invention will employ, unless indicated specifically to the contrary, conventional methods of chemistry, biochemistry, organic chemistry, molecular biology, microbiology, recombinant DNA techniques, genetics, immunology, and cell biology that are within the skill of the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, e.g., Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (3rd Edition, 2001); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis et al., *Molecular Cloning: A Laboratory Manual* (1982); Ausubel et al., *Current Protocols in Molecular Biology* (John Wiley and Sons, updated July 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford, 1985); Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); *Transcription and Translation* (B. Hames & S. Higgins, Eds., 1984); Perbal, *A Practical Guide to Molecul-*

*lar Cloning* (1984); and Harlow and Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998).

## B. Definitions

[0084] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, preferred embodiments of compositions, methods and materials are described herein. For the purposes of the present invention, the following terms are defined below.

[0085] The articles "a," "an," and "the" are used herein to refer to one or to more than one (i.e. to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element.

[0086] The use of the alternative (e.g., "or") should be understood to mean either one, both, or any combination thereof of the alternatives.

[0087] The term "and/or" should be understood to mean either one, or both of the alternatives.

[0088] As used herein, the term "about" or "approximately" refers to a quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length that varies by as much as 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% or 1% to a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length. In one embodiment, the term "about" or "approximately" refers a range of quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length  $\pm$  15%, +10%, +9%,  $\pm$  8%,  $\pm$  7%,  $\pm$  6%,  $\pm$  5%,  $\pm$  4%,  $\pm$  3%,  $\pm$  2%, or  $\pm$  1% about a reference quantity, level, value, number, frequency, percentage, dimension, size, amount, weight or length.

[0089] Throughout this specification, unless the context requires otherwise, the words "comprise", "comprises" and "comprising" will be understood to imply the inclusion of a stated step or element or group of steps or elements but not the exclusion of any other step or element or group of steps or elements. In particular embodiments, the terms "include," "has," "contains," and "comprise" are used synonymously.

[0090] By "consisting of" is meant including, and limited to, whatever follows the phrase "consisting of." Thus, the phrase "consisting of" indicates that the listed elements are required or mandatory, and that no other elements may be present.

[0091] By "consisting essentially of" is meant including any elements listed after the phrase, and limited to other elements that do not interfere with or contribute to the activity or action specified in the disclosure for the listed elements. Thus, the phrase "consisting essentially of" indicates that the listed elements are required or mandatory, but that no other elements are optional and may or may not be present depending upon whether or not they affect the activity or action of the listed elements.

[0092] Reference throughout this specification to "one embodiment," "an embodiment," "a particular embodiment," "a related embodiment," "a certain embodiment," "an additional embodiment," or "a further embodiment" or combinations thereof means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the

present invention. Thus, the appearances of the foregoing phrases in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0093] As used herein, the term "isolated" means material that is substantially or essentially free from components that normally accompany it in its native state. In particular embodiments, the term "obtained" or "derived" is used synonymously with isolated.

[0094] As used herein, the term "DNA" refers to deoxyribonucleic acid. In various embodiments, the term DNA refers to genomic DNA, recombinant DNA, synthetic DNA, or cDNA. In one embodiment, DNA refers to genomic DNA or cDNA. In particular embodiments, the DNA comprises a "target region." DNA libraries contemplated herein include genomic DNA libraries and cDNA libraries constructed from RNA, e.g., an RNA expression library. In various embodiments, the DNA libraries comprise one or more additional DNA sequences and/or tags.

[0095] A "target genetic locus" or "DNA target region" refers to a region of interest within a DNA sequence. In various embodiments, targeted genetic analyses are performed on the target genetic locus. In particular embodiments, the DNA target region is a region of a gene that is associated with a particular genetic state, genetic condition, genetic diseases; fetal testing; genetic mosaicism, paternity testing; predicting response to drug treatment; diagnosing or monitoring a medical condition; microbiome profiling; pathogen screening; or organ transplant monitoring.

[0096] As used herein, the terms "circulating DNA," "circulating cell-free DNA" and "cell-free DNA" are often used interchangeably and refer to DNA that is extracellular DNA, DNA that has been extruded from cells, or DNA that has been released from necrotic or apoptotic cells.

[0097] A "subject," "individual," or "patient" as used herein, includes any animal that exhibits a symptom of a condition that can be detected or identified with compositions contemplated herein. Suitable subjects include laboratory animals (such as mouse, rat, rabbit, or guinea pig), farm animals (such as horses, cows, sheep, pigs), and domestic animals or pets (such as a cat or dog). In particular embodiments, the subject is a mammal. In certain embodiments, the subject is a non-human primate and, in preferred embodiments, the subject is a human.

### C. Methods of Genetic Analysis of Cell Free DNA

[0098] In various embodiments, a method for genetic analysis of cfDNA is provided.

[0099] In particular embodiments, a method for genetic analysis of cfDNA comprises: generating and amplifying a cfDNA library, determining the number of genome equivalents in the cfDNA library; and performing a quantitative genetic analysis of one or more genomic target loci.

[0100] A method for genetic analysis of cfDNA comprises treating cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA and ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate cfDNA library clones; determining the number of genome equivalents of cfDNA library clones; and performing a quantitative genetic analysis of one or more target genetic loci in the cfDNA library clones.

#### 1. Generating a cfDNA Library

[0101] In particular embodiments, methods of genetic analysis contemplated herein comprise generating a cfDNA library comprising treating cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA and ligating one or more adaptors to each end of the end-repaired cfDNA to generate the cfDNA library.

##### (a) Cell-Free DNA (cfDNA)

[0102] The methods and compositions contemplated herein are designed to efficiently analyze, detect, diagnose, and/or monitor genetic states, genetic conditions, genetic diseases, genetic mosaicism, fetal diagnostics, paternity testing, microbiome profiling, pathogen screening, and organ transplant monitoring using cell-free DNA (cfDNA) as an analyte. The size distribution of cfDNA ranges from about 150 bp to about 180 bp fragments. Fragmentation may be the result of endonucleolytic and/or exonucleolytic activity and presents a formidable challenge to the accurate, reliable, and robust analysis of cfDNA. Another challenge for analyzing cfDNA is its short half-life in the blood stream, on the order of about 15 minutes. Without wishing to be bound to any particular theory, the present invention contemplates, in part, that analysis of cfDNA is like a "liquid biopsy" and is a real-time snapshot of current biological processes.

[0103] Moreover, because cfDNA is not found within cells and may be obtained from a number of suitable sources including, but not limited to, biological fluids and stool samples, it is not subject to the existing limitations that plague next generation sequencing analysis, such as direct access to the tissues being analyzed.

[0104] Illustrative examples of biological fluids that are suitable sources from which to isolate cfDNA in particular embodiments include, but are not limited to amniotic fluid, blood, plasma, serum, semen, lymphatic fluid, cerebral spinal fluid, ocular fluid, urine, saliva, mucous, and sweat.

[0105] In particular embodiments, the biological fluid is blood or blood plasma.

[0106] In certain embodiments, commercially available kits and other methods known to the skilled artisan can be used to isolate cfDNA directly from the biological fluids of a patient or from a previously obtained and optionally stabilized biological sample, e.g., by freezing and/or addition of enzyme chelating agents including, but not limited to EDTA, EGTA, or other chelating agents specific for divalent cations.

##### (b) Generating End-Repaired cfDNA

[0107] In particular embodiments, generating a cfDNA library comprises the end-repair of isolated cfDNA. The fragmented cfDNA is processed by end-repair enzymes to generate end-repaired cfDNA with blunt ends, 5'-overhangs, or 3'-overhangs. In some embodiments, the end-repair enzymes can yield for example. In some embodiments, the end-repaired cfDNA contains blunt ends. In some embodiments, the end-repaired cfDNA is processed to contain blunt ends. In some embodiments, the blunt ends of the end-repaired cfDNA are further modified to contain a single base pair overhang. In some embodiments, end-repaired cfDNA containing blunt ends can be further processed to contain adenine (A)/thymine (T) overhang. In some embodiments, end-repaired cfDNA containing blunt ends can be further processed to contain adenine (A)/thymine (T) overhang as the single base pair overhang. In some embodiments, the end-repaired cfDNA has non-templated 3' overhangs. In

some embodiments, the end-repaired cfDNA is processed to contain 3' overhangs. In some embodiments, the end-repaired cfDNA is processed with terminal transferase (TdT) to contain 3' overhangs. In some embodiments, a G-tail can be added by TdT. In some embodiments, the end-repaired cfDNA is processed to contain overhang ends using partial digestion with any known restriction enzymes (e.g., with the enzyme Sau3A, and the like).

(c) Ligating Adaptor Molecules to End-Repaired cfDNA

[0108] In particular embodiments, generating a cfDNA library comprises ligating one or more adaptors to each end of the end-repaired cfDNA. The present invention contemplates, in part, an adaptor module designed to accommodate large numbers of genome equivalents in cfDNA libraries. Adaptor modules are configured to measure the number of genome equivalents present in cfDNA libraries, and, by extension, the sensitivity of sequencing assays used to identify sequence mutations.

[0109] As used herein, the term "adaptor module" refers to a polynucleotide comprising at least five elements: (i) a first element comprising a PCR primer binding site for the single-primer library amplification; (ii) a second element comprising a 5 nucleotide read code that acts to uniquely identify each sequencing read; (iii) a third element comprising a 3 nucleotide sample code to identify different samples and enable sample multiplexing within a sequencing run; (iv) a fourth element comprising a 12 nucleotide anchor sequence that enables calibration of proper base calls in sequencing reads and acts as an anchor for hybridization to a partner oligonucleotide; and (v) a fifth element comprising the two 3' terminal nucleotides of Element 4 (FIG. 17 and Tables 12-16). The adaptor module is hybridized to a partner oligonucleotide that is complementary to Element 4 to form an adaptor suitable for ligating to the ends of cfDNA, optionally end-repaired blunt-ended cfDNA.

[0110] In particular embodiments, an adaptor module comprises one or more PCR primer sequences, one or more read codes, one or more sample codes, one or more anchor sequences, and two or more 3'nucleotides that are efficient ligation substrates. In additional embodiments, the adaptor module further comprises one or more sequencing primer binding sites.

[0111] In particular embodiments, an adaptor module comprises a first element that comprises one or more PCR primer binding sequences for single-primer amplification of a cfDNA library. In one embodiment, the PCR primer binding sequence is about 12 to about 40 nucleotides, about 18 to about 40 nucleotides, about 20 to about 35 nucleotides, or about 20 to about 30 nucleotides. In another embodiment, the PCR primer binding sequence is about 12 nucleotides, about 13 nucleotides, about 14 nucleotides, about 15 nucleotides, about 16 nucleotides, about 17 nucleotides, about 18 nucleotides, about 19 nucleotides, about 20 nucleotides, about 21 nucleotides, about 22 nucleotides, about 23 nucleotides, about 24 nucleotides, about 25 nucleotides, about 26 nucleotides, about 27 nucleotides, about 28 nucleotides, about 29 nucleotides, about 30 nucleotides, about 31 nucleotides, about 32 nucleotides, about 33 nucleotides, about 34 nucleotides, about 35 nucleotides, about 36 nucleotides, about 37 nucleotides, about 38 nucleotides, about 39 nucleotides, or about 40 nucleotides or more.

[0112] In one embodiment, the PCR primer binding sequence is about 25 nucleotides.

[0113] In particular embodiments, an adaptor module comprises a second element that comprises one or more read code sequences. As used herein, the term "read code" refers to a polynucleotide that is used to identify unique sequencing reads. In one embodiment, the read code is a random sequence of nucleotides. In one embodiment, the read code is about 1 nucleotide, about 2 nucleotides, about 3 nucleotides, about 4 nucleotides, about 5 nucleotides, about 6 nucleotides, about 7 nucleotides, about 8 nucleotides, about 9 nucleotides, about 10 nucleotides, or more.

[0114] By way of a non-limiting example, a 5 nucleotide codes consists of 256 possible unique sequences where each code chosen is 2 nucleotides different from every other code in the set. This feature enables unique and distinct reads to be differentiated from reads that appear to be unique owing to a sequencing error in the code region. In particular embodiments, codes that have been empirically determined to interfere with adaptor function, owing to particular sequence combinations, may be excluded from use, e.g., seven codes of the 256 had an overrepresentation of G nucleotides and were excluded.

[0115] In other embodiments, each read code of 5, 6, 7, 8, 9, 10 or more nucleotides may differ by 2, 3, 4, or 5 nucleotides from every other read code.

[0116] In one embodiment, the read code is about 5 nucleotides and differs from every other read code by 2 nucleotides.

[0117] In particular embodiments, an adaptor module comprises a third element that comprises one or more sample code sequences. As used herein, the term "sample code" refers to a polynucleotide that is used to identify the sample. The sample code is also useful in establishing multiplex sequencing reactions because each sample code is unique to the sample and thus, can be used to identify a read from a particular sample within a multiplexed sequencing reaction.

[0118] In one embodiment, the sample code comprises sequence that is about 1, about 2 nucleotides, about 3 nucleotides, about 4 nucleotides, or about 5 nucleotides, or more. In another embodiment, each sample code of 2, 3, 4, 5 or more nucleotides may differ from every other sample code by 2, 3, 4, or 5 nucleotides.

[0119] In one embodiment, the sample code is about three nucleotides and differs from every other sample code used in other samples by two nucleotides.

[0120] In particular embodiments, an adaptor module comprises a fourth element that comprises one or more anchor sequences. As used herein, an "anchor sequence" refers to a nucleotide sequence of at least 8 nucleotides, at least 10 nucleotides, at least 12 nucleotides, at least 14 nucleotides, or at least 16 nucleotides that hybridizes to a partner oligonucleotide and that comprises the following three properties: (1) each anchor sequence is part of a family of four anchor sequences that collectively represent each of the four possible DNA bases at each site within extension; this feature, balanced base representation, is useful to calibrate proper base calling in sequencing reads in particular embodiments; (2) each anchor sequence is composed of only two of four possible bases, and these are specifically chosen to be either and equal number of A+C or an equal number of G+T; an anchor sequence formed from only two bases reduces the possibility that the anchor sequence will participate in secondary structure formation that would preclude proper adaptor function; and (3) because each anchor

sequence is composed of equal numbers of A+C or G+T, each anchor sequence shares roughly the same melting temperature and duplex stability as every other anchor sequence in a set of four.

[0121] In particular embodiments, an adaptor module comprises a fifth element that is comprised of the two 3' terminal nucleotides of Element 4. These two bases at the 3' end of each anchor are chosen based on an empirical determination that shows that these two nucleotides are efficient substrates for ligation to the cfDNA. In particular embodiments, Element 5 comprises the sequences selected from the group consisting of: AA, CC, TT and GG. In particular embodiments, Element 5 does not comprise the dinucleotide combination CG or TG as the inventors have determined that these combinations are not efficient ligation substrates.

[0122] In particular embodiments, a ligation step comprises ligating an adaptor module to the end-repaired cfDNA to generate a “tagged” cfDNA library. In some embodiments, a single adaptor module is employed. In some embodiments, two, three, four or five adaptor modules are employed. In some embodiments, an adaptor module of identical sequence is ligated to each end of the fragmented end-repaired DNA.

[0123] In one embodiment, a plurality of adaptor species is ligated to an end-repaired cfDNA library. Each of the plurality of adaptors may comprise one or more primer binding site for amplification of the cfDNA library, one or more read code sequences, one or more sequences for sample multiplexing, and one or more sequences for DNA sequencing.

[0124] Ligation of one or more adaptors contemplated herein may be carried out by methods known to those of ordinary skill in the art. In particular embodiments, one or more adaptors contemplated herein are ligated to end-repaired cfDNA that comprises blunt ends. In certain embodiments, one or more adaptors contemplated herein are ligated to end-repaired cfDNA that comprises complementary ends appropriate for the ligation method employed. In certain embodiments, one or more adaptors contemplated herein are ligated to end-repaired cfDNA that comprises a 3' overhang.

## 2. cfDNA Library Amplification

[0125] In particular embodiments, methods of genetic analysis contemplated herein comprise amplification of a cfDNA library to generate a cfDNA clone library or a library of cfDNA clones. Each molecule of the cfDNA library comprises an adaptor ligated to each end of an end-repaired cfDNA, and each adaptor comprises one or more PCR primer binding sites. In one embodiment, different adaptors are ligated to different ends of the end-repaired cfDNA.

[0126] In a preferred embodiment, the same adaptor is ligated to both ends of the cfDNA. Ligation of the same adaptor to both ends of end-repaired cfDNA allows for PCR amplification with a single primer sequence. In particular embodiments, a portion of the adaptor ligated-cfDNA library will be amplified using standard PCR techniques with a single primer sequence driving amplification. In one embodiment, the single primer sequence is about 25 nucleotides, optionally with a projected Tm of  $\geq 55^\circ$  C. under standard ionic strength conditions.

[0127] In particular embodiments, picograms of the initial cfDNA library are amplified into micrograms of cfDNA clones, implying a 10,000-fold amplification. The amount of

amplified product can be measured using methods known in the art, e.g., quantification on a Qubit 2.0 or Nanodrop instrument.

## 3. Determining the Number of Genome Equivalents

In various embodiments, a method for genetic analysis of cfDNA comprises determining the number of genome equivalents in the cfDNA clone library. As used herein, the term “genome equivalent” refers to the number of genome copies in each library. An important challenge met by the compositions and methods contemplated herein is achieving sufficient assay sensitivity to detect and analyze rare genetic mutations or differences in genetic sequence. To determine assay sensitivity value on a sample-by-sample basis, the numbers of different and distinct sequences that are present in each sample are measured, by measuring the number of genome equivalents that are present in a sequencing library. To establish sensitivity, the number of genome equivalents must be measured for each sample library.

[0128] The number of genome equivalents can be determined by qPCR assay or by using bioinformatics-based counting after sequencing is performed. In the process flow of clinical samples, qPCR measurement of genome equivalents is used as a QC step for cfDNA libraries. It establishes an expectation for assay sensitivity prior to sequence analysis and allows a sample to be excluded from analysis if its corresponding cfDNA clone library lacks the required depth of genome equivalents. Ultimately, the bioinformatics-based counting of genome equivalents is also used to identify the genome equivalents—and hence the assay sensitivity and false negative estimates—for each given cfDNA clone library.

[0129] The empirical qPCR assay and statistical counting assays should be well correlated. In cases where sequencing fails to reveal the sequence depth in a cfDNA clone library, reprocessing of the cfDNA clone library and/or additional sequencing may be required.

[0130] In one embodiment, the genome equivalents in a cfDNA clone library are determined using a quantitative PCR (qPCR) assay. In a particular embodiment, a standard library of known concentration is used to construct a standard curve and the measurements from the qPCR assay are fit to the resulting standard curve and a value for genome equivalents is derived from the fit. Surprisingly, the present inventors have discovered that a qPCR “repeat-based” assay comprising one primer that specifically hybridizes to a common sequence in the genome, e.g., a repeat sequence, and another primer that binds to the primer binding site in the adaptor, measured an 8-fold increase in genome equivalents compared to methods using just the adaptor specific primer (present on both ends of the cfDNA clone). The number of genome equivalents measured by the repeat-based assays provides a more consistent library-to-library performance and a better alignment between qPCR estimates of genome equivalents and bioinformatically counted tag equivalents in sequencing runs.

[0131] Illustrative examples of repeats suitable for use in the repeat-based genome equivalent assays contemplated herein include, but not limited to: short interspersed nuclear elements (SINEs), e.g., Alu repeats; long interspersed nuclear elements (LINEs), e.g., LINE1, LINE2, LINE3; microsatellite repeat elements, e.g., short tandem repeats (STRs), simple sequence repeats (SSRs); and mammalian-wide interspersed repeats (MTRs).

[0132] In one embodiment, the repeat is an Alu repeat.

#### 4. Quantitative Genetic Analysis

[0133] In various embodiments, a method for genetic analysis of cfDNA comprises quantitative genetic analysis of one or more target genetic loci of the cfDNA library clones. Quantitative genetic analysis comprises one or more of, or all of, the following steps: capturing cfDNA clones comprising a target genetic locus; amplification of the captured targeted genetic locus; sequencing of the amplified captured targeted genetic locus; and bioinformatic analysis of the resulting sequence reads.

##### (a) Capture of Target Genetic Locus

[0134] The present invention contemplates, in part, a capture probe module designed to retain the efficiency and reliability of larger probes but that minimizes uninformative sequence generation in a cfDNA clone library. A “capture probe module” refers to a polynucleotide that comprises a capture probe sequence and a tail sequence. In particular embodiments, the capture probe module sequence or a portion thereof serves as a primer binding site for one or more sequencing primers.

[0135] In particular embodiments, a capture probe module comprises a capture probe. As used herein a “capture probe” refers to a region capable of hybridizing to a specific DNA target region. Because the average size of cfDNA is about 150 to about 170 bp and is highly fragmented the compositions and methods contemplated herein comprise the use of high density and relatively short capture probes to interrogate DNA target regions of interest.

[0136] One particular concern with using high density capture probes is that generally capture probes are designed using specific “sequence rules.” For example, regions of redundant sequence or that exhibit extreme base composition biases are generally excluded in designing capture probes. However, the present inventors have discovered that the lack of flexibility in capture probe design rules does not substantially impact probe performance. In contrast, capture probes chosen strictly by positional constraint provided on-target sequence information; exhibit very little off-target and unmappable read capture; and yield uniform, useful, on-target reads with only few exceptions. Moreover, the high redundancy at close probe spacing more than compensates for occasional poor-performing capture probes.

[0137] In particular embodiments, a target region is targeted by a plurality of capture probes, wherein any two or more capture probes are designed to bind to the target region within 10 nucleotides of each other, within 15 nucleotides of each other, within 20 nucleotides of each other, within 25 nucleotides of each other, within 30 nucleotides of each other, within 35 nucleotides of each other, within 40 nucleotides of each other, within 45 nucleotides of each other, or within 50 nucleotides or more of each other, as well as all intervening nucleotide lengths.

[0138] In one embodiment, the capture probe is about 25 nucleotides, about 26 nucleotides, about 27 nucleotides, about 28 nucleotides, about 29 nucleotides, about 30 nucleotides, about 31 nucleotides, about 32 nucleotides, about 33 nucleotides, about 34 nucleotides, about 35 nucleotides, about 36 nucleotides, about 37 nucleotides, about 38 nucleotides, about 39 nucleotides, about 40 nucleotides, about 41

nucleotides, about 42 nucleotides, about 43 nucleotides, about 44 nucleotides, or about 45 nucleotides.

[0139] In one embodiment, the capture probe is about 100 nucleotides, about 200 nucleotides, about 300 nucleotides, about 400 nucleotides, or about 100 nucleotides. In another embodiment, the capture probe is from about 100 nucleotides to about 500 nucleotides, about 200 nucleotides to about 500 nucleotides, about 300 nucleotides to about 500 nucleotides, or about 400 nucleotides to about 500 nucleotides, or any intervening range thereof.

[0140] In a particular embodiment, the capture probe is not 60 nucleotides.

[0141] In another embodiment, the capture probe is substantially smaller than 60 nucleotides but hybridizes comparably, as well as, or better than a 60 nucleotide capture probe targeting the same DNA target region.

[0142] In a certain embodiment, the capture probe is 40 nucleotides.

[0143] In certain embodiments, a capture probe module comprises a tail sequence. As used herein, the term “tail sequence” refers to a polynucleotide at the 5' end of the capture probe module, which in particular embodiments can serve as a primer binding site. In particular embodiments, a sequencing primer binds to the primer binding site in the tail region.

[0144] In particular embodiments, the tail sequence is about 5 to about 100 nucleotides, about 10 to about 100 nucleotides, about 5 to about 75 nucleotides, about 5 to about 50 nucleotides, about 5 to about 25 nucleotides, or about 5 to about 20 nucleotides. In certain embodiments, the third region is from about 10 to about 50 nucleotides, about 15 to about 40 nucleotides, about 20 to about 30 nucleotides or about 20 nucleotides, or any intervening number of nucleotides.

[0145] In particular embodiments, the tail sequence is about 30 nucleotides, about 31 nucleotides, about 32 nucleotides, about 33 nucleotides, about 34 nucleotides, about 35 nucleotides, about 36 nucleotides, about 37 nucleotides, about 38 nucleotides, about 39 nucleotides, or about 40 nucleotides.

[0146] In various embodiments, the capture probe module comprises a specific member of a binding pair to enable isolation and/or purification of one or more captured fragments of a tagged and or amplified cfDNA library that hybridizes to the capture probe. In particular embodiments, the capture probe module is conjugate to biotin or another suitable hapten, e.g., dinitrophenol, digoxigenin.

[0147] In various embodiments, the capture probe module is hybridized to a tagged and optionally amplified cfDNA library to form a complex. In some embodiments, the multifunctional capture probe module substantially hybridizes to a specific genomic target region in the cfDNA library.

[0148] Hybridization or hybridizing conditions can include any reaction conditions where two nucleotide sequences form a stable complex; for example, the tagged cfDNA library and capture probe module forming a stable tagged cfDNA library capture probe module complex. Such reaction conditions are well known in the art and those of skill in the art will appreciate that such conditions can be modified as appropriate, e.g., decreased annealing temperatures with shorter length capture probes, and within the scope of the present invention. Substantial hybridization can occur when the second region of the capture probe complex exhibits 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%

91%, 90%, 89%, 88%, 85%, 80%, 75%, or 70% sequence identity, homology or complementarity to a region of the tagged cfDNA library.

[0149] In particular embodiments, the capture probe is about 40 nucleotides and has an optimal annealing temperature of about 44° C. to about 47° C.

[0150] In certain embodiments, the methods contemplated herein comprise isolating a tagged cfDNA library-capture probe module complex. In particular embodiments, methods for isolating DNA complexes are well known to those skilled in the art and any methods deemed appropriate by one of skill in the art can be employed with the methods of the present invention (Ausubel et al., *Current Protocols in Molecular Biology*, 2007-2012). In particular embodiments, the complexes are isolated using biotin-streptavidin isolation techniques.

[0151] In particular embodiments, removal of the single stranded 3'-ends from the isolated tagged cfDNA library—capture probe module complex is contemplated. In certain embodiments, the methods comprise 3'-5' exonuclease enzymatic processing of the isolated tagged DNA library-mutifunctional capture probe module complex to remove the single stranded 3' ends.

[0152] In certain other embodiments, the methods comprise performing 5'-3' DNA polymerase extension of multifunctional capture probe utilizing the isolated tagged DNA library fragments as template.

[0153] In certain other embodiments, the methods comprise creating a hybrid capture probe-isolated tagged cfDNA target molecule through the concerted action of a 5' FLAP endonuclease, DNA polymerization and nick closure by a DNA ligase.

[0154] A variety of enzymes can be employed for the 3'-5' exonuclease enzymatic processing of the isolated tagged cfDNA library-mutifunctional capture probe module complex. Illustrative examples of suitable enzymes, which exhibit 3'-5' exonuclease enzymatic activity, that can be employed in particular embodiments include, but are not limited to: T4 or Exonucleases I, III, V (see also, Shevelev IV, Hübscher U., "The 3' 5' exonucleases," *Nat Rev Mol Cell Biol.* 3(5):364-76 (2002)). In particular embodiments, the enzyme comprising 3'-5' exonuclease activity is T4 polymerase. In particular embodiments, an enzyme which exhibits 3'-5' exonuclease enzymatic activity and is capable of primer template extension can be employed, including for example T4 or Exonucleases I, III, V. Id.

[0155] In some embodiments, the methods contemplated herein comprise performing sequencing and/or PCR on the 3'-5' exonuclease enzymatically processed complex discussed supra and elsewhere herein. In particular embodiments, a tail portion of a capture probe molecule is copied in order to generate a hybrid nucleic acid molecule. In one embodiment, the hybrid nucleic acid molecule generated comprises the target region capable of hybridizing to the capture probe module and the complement of the capture probe module tail sequence.

[0156] In a particular embodiment, genetic analysis comprises a) hybridizing one or more capture probe modules to one or more target genetic loci in a plurality of cfDNA library clones to form one or more capture probe module-cfDNA library clone complexes; b) isolating the one or more capture probe module-cfDNA library clone complexes from a); c) enzymatically processing the one or more isolated capture probe module-cfDNA library clone complexes from

step b); d) performing PCR on the enzymatically processed complex from c) wherein the tail portion of the capture probe molecule is copied in order to generate amplified hybrid nucleic acid molecules, wherein the amplified hybrid nucleic acid molecules comprise a target sequence in the target genomic locus capable of hybridizing to the capture probe and the complement of the capture probe module tail sequence; and e) performing quantitative genetic analysis on the amplified hybrid nucleic acid molecules from d).

[0157] In a particular embodiment, methods for determining copy number of a specific target genetic locus are contemplated comprising: a) hybridizing one or more capture probe modules to one or more target genetic loci in a plurality of cfDNA library clones to form one or more capture probe module-cfDNA library clone complexes; b) isolating the one or more capture probe module-cfDNA library clone complexes from a); c) enzymatically processing the one or more isolated capture probe module-cfDNA library clone complexes from step b); d) performing PCR on the enzymatically processed complex from c) wherein the tail portion of the capture probe molecule is copied in order to generate amplified hybrid nucleic acid molecules, wherein the amplified hybrid nucleic acid molecules comprise a target sequence in the target genetic locus capable of hybridizing to the capture probe and the complement of the capture probe module tail sequence; e) performing PCR amplification of the amplified hybrid nucleic acid molecules in d); and f) quantitating the PCR reaction in e), wherein the quantitation allows for a determination of copy number of the specific target region.

[0158] In one embodiment, the enzymatic processing of step c) comprises performing 3'-5' exonuclease enzymatic processing on the one or more capture probe module-cfDNA library clone complexes from b) using an enzyme with 3'-5' exonuclease activity to remove the single stranded 3' ends; creating one or more hybrid capture probe module-cfDNA library clone molecules through the concerted action of a 5' FLAP endonuclease, DNA polymerization and nick closure by a DNA ligase; or performing 5'-3' DNA polymerase extension of the capture probe using the isolated cfDNA clone in the complex as a template.

[0159] In one embodiment, the enzymatic processing of step c) comprises performing 5'-3' DNA polymerase extension of the capture probe using the isolated cfDNA clone in the complex as a template.

[0160] In particular embodiments, PCR can be performed using any standard PCR reaction conditions well known to those of skill in the art. In certain embodiments, the PCR reaction in e) employs two PCR primers. In one embodiment, the PCR reaction in e) employs a first PCR primer that hybridizes to a repeat within the target genetic locus. In a particular embodiment, the PCR reaction in e) employs a second PCR primer that hybridizes to the hybrid nucleic acid molecules at the target genetic locus/tail junction. In certain embodiments, the PCR reaction in e) employs a first PCR primer that hybridizes to the target genetic locus and a second PCR primer hybridizes to the amplified hybrid nucleic acid molecules at the target genetic locus/tail junction. In particular embodiments, the second primer hybridizes to the target genetic locus/tail junction such that at least one or more nucleotides of the primer hybridize to the target genetic locus and at least one or more nucleotides of the primer hybridize to the tail sequence.

[0161] In certain embodiments, the amplified hybrid nucleic acid molecules obtained from step e) are sequenced and the sequences aligned horizontally, i.e., aligned to one another but not aligned to a reference sequence. In particular embodiments, steps a) through e) are repeated one or more times with one or more capture probe modules. The capture probe modules can be the same or different and designed to target either cfDNA strand of a target genetic locus. In some embodiments, when the capture probes are different, they hybridize at overlapping or adjacent target sequences within a target genetic locus in the tagged cfDNA clone library. In one embodiment, a high density capture probe strategy is used wherein a plurality of capture probes hybridize to a target genetic locus, and wherein each of the plurality of capture probes hybridizes to the target genetic locus within about 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200 bp of any other capture probe that hybridizes to the target genetic locus in a tagged cfDNA clone library, including all intervening distances.

[0162] In some embodiments, the method can be performed using two capture probe modules per target genetic locus, wherein one hybridizes to the "Watson" strand (non-coding or template strand) upstream of the target region and one hybridizes to the "Crick" strand (coding or non-template strand) downstream of the target region.

[0163] In particular embodiments, the methods contemplated herein can further be performed multiple times with any number of capture probe modules, for example 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more capture probe modules per target genetic locus any number of which hybridize to the Watson or Crick strand in any combination. In some embodiments, the sequences obtained can be aligned to one another in order to identify any of a number of differences.

[0164] In certain embodiments, a plurality of target genetic loci are interrogated, e.g., 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 2500, 3000, 3500, 4000, 4500, 5000, 10000, 50000, 100000, 500000 or more in a single reaction, using one or more capture probe modules.

#### (b) Sequencing

[0165] In particular embodiments, the quantitative genetic analysis comprises sequencing a plurality of hybrid nucleic acid molecules, as discussed elsewhere herein, supra, to generate sufficient sequencing depths to obtain a plurality of unique sequencing reads. A unique read is defined as the single consensus read from a "family" of reads that all share the same read code and sequence start point within cfDNA. Each capture probe yields a set of unique reads that are computationally distilled from total reads by grouping into families. The unique reads for a given sample are then computed as the average of all the unique reads observed on a probe-by-probe basis. Cases where there is an obvious copy number change are excluded from the data set used to compute the average. Unique reads are important because each unique read must be derived from a unique cfDNA clone. Each unique read represents the input and analysis of a haploid equivalent of genomic DNA. The sum of unique reads is the sum of haploid genomes analyzed. The number of genomes analyzed, in turn, defines the sensitivity of the sequencing assay. By way of a non-limiting example, if the average unique read count is 100 genome equivalents, then

that particular assay has a sensitivity of being able to detect one mutant read in 100, or 1%. Any observation less than this is not defensible.

[0166] In particular embodiments, the quantitative genetic analysis comprises multiplex sequencing of hybrid nucleic acid molecules derived from a plurality of samples.

[0167] In various embodiments, the quantitative genetic analysis comprises obtaining one or more or a plurality of tagged DNA library clones, each clone comprising a first DNA sequence and a second DNA sequence, wherein the first DNA sequence comprises a sequence in a targeted genetic locus and the second DNA sequence comprises a capture probe sequence; performing a paired end sequencing reaction on the one or more clones and obtaining one or more sequencing reads or performing a sequencing reaction on the one or more clones in which a single long sequencing read of greater than about 100, 200, 300, 400, 500 or more nucleotides is obtained, wherein the read is sufficient to identify both the first DNA sequence and the second DNA sequence; and ordering or clustering the sequencing reads of the one or more clones according to the probe sequences of the sequencing reads.

#### (c) Bioinformatics Analysis

[0168] In various embodiments, the quantitative genetic analysis further comprises bioinformatic analysis of the sequencing reads. Bioinformatic analysis excludes any purely mental analysis performed in the absence of a composition or method for sequencing. In certain embodiments, bioinformatics analysis includes, but is not limited to: sequence alignments; genome equivalents analysis; single nucleotide variant (SNV) analysis; gene copy number variation (CNV) analysis; and detection of genetic lesions. In particular embodiments, bioinformatics analysis is useful to quantify the number of genome equivalents analyzed in the cfDNA clone library; to detect the genetic state of a target genetic locus; to detect genetic lesions in a target genetic locus; and to measure copy number fluctuations within a target genetic locus.

[0169] Sequence alignments may be performed between the sequence reads and one or more human reference DNA sequences. In particular embodiments, sequencing alignments can be used to detect genetic lesions in a target genetic locus including, but not limited to detection of a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion. Detection of genetic lesions that are causal or prognostic indicators may be useful in the diagnosis, prognosis, treatment, and/or monitoring of a particular genetic condition or disease.

[0170] Also contemplated herein, are methods for sequence alignment analysis that can be performed without the need for alignment to a reference sequence, referred to herein as horizontal sequence analysis. Such analysis can be performed on any sequences generated by the methods contemplated herein or any other methods. In particular embodiments, the sequence analysis comprises performing sequence alignments on the reads obtained by the methods contemplated herein.

[0171] In one embodiment, the genome equivalents in a cfDNA clone library are determined using bioinformatics-based counting after sequencing is performed. Each sequencing read is associated with a particular capture probe, and the collection of reads assigned to each capture

probe is parsed into groups. Within a group, sets of individual reads share the same read code and the same DNA sequence start position within genomic sequence. These individual reads are grouped into a “family” and a single consensus representative of this family is carried forward as a “unique read.” All of the individual reads that constituted a family are derived from a single ligation event and thus, they are amplification-derived “siblings” of one another. Each unique read is considered a unique ligation event and the sum of unique reads is considered equivalent to the number of genome equivalents analyzed.

[0172] As the number of unique clones approaches the total number of possible sequence combinations, probability dictates that the same code and start site combinations will be created by independent events and that these independent events will be inappropriately grouped within single families. The net result will be an underestimate of genome equivalents analyzed, and rare mutant reads may be discarded as sequencing errors because they overlap with wild-type reads bearing the same identifiers.

[0173] In particular embodiments, to provide an accurate analysis for cfDNA clone libraries, the number of genome equivalents analyzed is about  $\frac{1}{10}$ , about  $\frac{1}{12}$ , about  $\frac{1}{14}$ , about  $\frac{1}{16}$ , about  $\frac{1}{18}$ , about  $\frac{1}{20}$ , about  $\frac{1}{25}$  or less the number of possible unique clones. It should be understood that the procedure outlined above is merely illustrative and not limiting.

[0174] In some embodiments, the number of genome equivalents to be analyzed may need to be increased. To expand the depth of genome equivalents, at least two solutions are contemplated. The first solution is to use more than one adaptor set per sample. By combining adaptors, it is possible to multiplicatively expand the total number of possible clones and therefore, expand the comfortable limits of genomic input. The second solution is to expand the read code by 1, 2, 3, 4, or 5 or more bases. The number of possible read codes that differ by at least 2 bases from every other read code scales as  $4^{(n-1)}$  where n is the number of bases within a read code. Thus, in a non-limiting example, if a read code is 5 nucleotides and  $4^{(5-1)}=256$ ; therefore, the inclusion of additional bases expands the available repertoire by a factor of four for each additional base.

[0175] In one embodiment, quantitative genetic analysis comprises bioinformatic analysis of sequencing reads to identify rare single nucleotide variants (SNV).

[0176] Next-generation sequencing has an inherent error rate of roughly 0.02-0.02%, meaning that anywhere from  $\frac{1}{200}$  to  $\frac{1}{500}$  base calls are incorrect. To detect variants and other mutations that occur at frequencies lower than this, for example at frequencies of 1 per 1000 sequences, it is necessary to invoke molecular annotation strategies. By way of a non-limiting example, analysis of 5000 unique molecules using targeted sequence capture technology would generate—at sufficient sequencing depths of >50,000 reads—a collection of 5000 unique reads, with each unique read belonging to a “family” of reads that all possess the same read code. A SNV that occurs within a family is a candidate for being a rare variant. When this same variant is observed in more than one family, it becomes a very strong candidate for being a rare variant that exists within the starting sample. In contrast, variants that occur sporadically within families are likely to be sequencing errors and variants that occur within one and only one family are either

rare or the result of a base alteration that occurred ex vivo (e.g., oxidation of a DNA base or PCR-introduced errors).

[0177] In one embodiment, the methods of detecting SNVs comprise introducing 10-fold more genomic input (genomes or genome equivalents) as the desired target sensitivity of the assay. In one non-limiting example, if the desired sensitivity is 2% (2 in 100), then the experimental target is an input of 2000 genomes.

[0178] In particular embodiments, bioinformatics analysis of sequencing data is used to detect or identify SNV associated with a genetic state, condition or disease, genetic mosaicism, fetal testing, paternity testing, predicting response to drug treatment, diagnosing or monitoring a medical condition, microbiome profiling, pathogen screening, and monitoring organ transplants.

[0179] In various embodiments, a method for copy number determination analysis is provided comprising obtaining one or more or a plurality of clones, each clone comprising a first DNA sequence and a second DNA sequence, wherein the first DNA sequence comprises a sequence in a targeted genetic locus and the second DNA sequence comprises a capture probe sequence. In related embodiments, a paired end sequencing reaction on the one or more clones is performed and one or more sequencing reads are obtained. In another embodiment, a sequencing reaction on the one or more clones is performed in which a single long sequencing read of greater than about 100 nucleotides is obtained, wherein the read is sufficient to identify both the first DNA sequence and the second DNA sequence. The sequencing reads of the one or more clones can be ordered or clustered according to the probe sequence of the sequencing reads.

[0180] Copy number analyses include, but are not limited to analyses, that examine the number of copies of a particular gene or mutation that occurs in a given genomic DNA sample and can further include quantitative determination of the number of copies of a given gene or sequence differences in a given sample. In particular embodiments, copy number analysis is used to detect or identify gene amplification associated with genetic states, conditions, or diseases, fetal testing, genetic mosaicism, paternity testing, predicting response to drug treatment, diagnosing or monitoring a medical condition, microbiome profiling, pathogen screening, and monitoring organ transplants.

[0181] In particular embodiments, bioinformatics analysis of sequencing data is used to detect or identify one or more sequences or genetic lesions in a target locus including, but not limited to detection of a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion. Detection of genetic lesions that are causal or prognostic indicators may be useful in the diagnosis, prognosis, treatment, and/or monitoring of a particular genetic condition or disease. In one embodiment, genetic lesions are associated with genetic states, conditions, or diseases, fetal testing, genetic mosaicism, paternity testing, predicting response to drug treatment, diagnosing or monitoring a medical condition, microbiome profiling, pathogen screening, and monitoring organ transplants.

#### D. Clinical Applications of Quantitative Genetic Analysis

[0182] In various embodiments, the present invention contemplates a method of detecting, identifying, predicting, diagnosing, or monitoring a condition or disease in a subject.

[0183] In particular embodiments, a method of detecting, identifying, predicting, diagnosing, or monitoring a genetic state, condition or disease in a subject comprises performing a quantitative genetic analysis of one or more target genetic loci in a cfDNA clone library to detect or identify a change in the sequence at the one or more target genetic loci.

[0184] In one embodiment, a method of detecting, identifying, predicting, diagnosing, or monitoring a genetic state, condition or disease comprises isolating or obtaining cfDNA from a biological sample of a subject; treating the cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA; ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate a cfDNA clone library; determining the number of genome equivalents in the cfDNA clone library; and performing a quantitative genetic analysis of one or more target genetic loci in a cfDNA clone library to detect or identify a change in the sequence at the one or more target genetic loci.

[0185] In particular embodiments, a method of detecting, identifying, predicting, diagnosing, or monitoring a genetic state, or genetic condition or disease selected from the group consisting of: genetic diseases; genetic mosaicism; fetal testing; paternity testing; paternity testing; predicting response to drug treatment; diagnosing or monitoring a medical condition; microbiome profiling; pathogen screening; and organ transplant monitoring comprising isolating or obtaining cfDNA from a biological sample of a subject; treating the cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA; ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate a cfDNA clone library; determining the number of genome equivalents in the cfDNA clone library; and performing a quantitative genetic analysis of one or more target genetic loci in a cfDNA clone library to detect or identify a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion in the sequence at the one or more target genetic loci.

[0186] Illustrative examples of genetic diseases that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include, but are not limited to cancer, Alzheimer's disease (APOE1), Charcot-Marie-Tooth disease, Leber hereditary optic neuropathy (LHON), Angelman syndrome (UBE3A, ubiquitin-protein ligase E3A), Prader-Willi syndrome (region in chromosome 15),  $\beta$ -Thalassaemia (HBB,  $\beta$ -Globin), Gaucher disease (type I) (GBA, Glucocerebrosidase), Cystic fibrosis (CFTR Epithelial chloride channel), Sickle cell disease (HBB,  $\beta$ -Globin), Tay-Sachs disease (HEXA, Hexosaminidase A), Phenylketonuria (PAH, Phenylalanine hydrolyase), Familial hypercholesterolemia (LDLR, Low density lipoprotein receptor), Adult polycystic kidney disease (PKD1, Polycystin), Huntington disease (HDD, Huntingtin), Neurofibromatosis type I (NF1, NF1 tumour suppressor gene), Myotonic dystrophy (DM, Myotonin), Tuberous sclerosis (TSC1, Tuberin), Achondroplasia (FGFR3, Fibroblast growth factor receptor), Fragile X syndrome (FMR1, RNA-binding protein), Duchenne muscular dystrophy (DMD, Dystrophin), Haemophilia A (F8C, Blood coagulation factor VIII), Lesch-Nyhan syndrome (HPRT1, Hypoxanthine guanine ribosyltransferase 1), and Adrenoleukodystrophy (ABCD1).

[0187] Illustrative examples of cancers that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include, but are not limited to: B cell cancer, e.g., multiple myeloma, melanomas, breast cancer, lung cancer (such as non-small cell lung carcinoma or NSCLC), bronchus cancer, colorectal cancer, prostate cancer, pancreatic cancer, stomach cancer, ovarian cancer, urinary bladder cancer, brain or central nervous system cancer, peripheral nervous system cancer, esophageal cancer, cervical cancer, uterine or endometrial cancer, cancer of the oral cavity or pharynx, liver cancer, kidney cancer, testicular cancer, biliary tract cancer, small bowel or appendix cancer, salivary gland cancer, thyroid gland cancer, adrenal gland cancer, osteosarcoma, chondrosarcoma, cancer of hematological tissues, adenocarcinomas, inflammatory myofibroblastic tumors, gastrointestinal stromal tumor (GIST), colon cancer, multiple myeloma (MM), myelodysplastic syndrome (MDS), myeloproliferative disorder (MPD), acute lymphocytic leukemia (ALL), acute myelocytic leukemia (AML), chronic myelocytic leukemia (CML), chronic lymphocytic leukemia (CLL), polycythemia Vera, Hodgkin lymphoma, non-Hodgkin lymphoma (NHL), soft-tissue sarcoma, fibrosarcoma, myxosarcoma, liposarcoma, osteogenic sarcoma, chordoma, angiosarcoma, endothelioma, lymphangiosarcoma, lymphangiomyomatoma, synovioma, mesothelioma, Ewing's tumor, leiomyosarcoma, rhabdomyosarcoma, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, medullary carcinoma, bronchogenic carcinoma, renal cell carcinoma, hepatoma, bile duct carcinoma, choriocarcinoma, seminoma, embryonal carcinoma, Wilms' tumor, bladder carcinoma, epithelial carcinoma, glioma, astrocytoma, medulloblastoma, craniopharyngioma, ependymoma, pinealoma, hemangioblastoma, acoustic neuroma, oligodendroglioma, meningioma, neuroblastoma, retinoblastoma, follicular lymphoma, diffuse large B-cell lymphoma, mantle cell lymphoma, hepatocellular carcinoma, thyroid cancer, gastric cancer, head and neck cancer, small cell cancers, essential thrombocythemia, agnogenic myeloid metaplasia, hypereosinophilic syndrome, systemic mastocytosis, familiar hypereosinophilia, chronic eosinophilic leukemia, neuroendocrine cancers, carcinoid tumors, and the like.

[0188] In one embodiment, the genetic lesion is a lesion annotated in the Cosmic database (the lesions and sequence data can be downloaded from the Cancer Gene Census section of the Cosmic website) or a lesion annotated in the Cancer Genome Atlas (the lesions and sequence data can be downloaded from The Cancer Genome Atlas website of the National Cancer Institute and National Human Genome Research Institute).

[0189] Illustrative examples of genes that harbor one or more genetic lesions associated with cancer that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include, but are not limited to ABCB1, ABCC2, ABCC4, ABCG2, ABL1, ABL2, AKT1, AKT2, AKT3, ALDH4A1, ALK, APC, AR, ARAF, ARFRP1, ARID1A, ATM, ATR, AURKA, AURKB, BCL2, BCL2A1, BCL2L1, BCL2L2, BCL6, BRAF, BRCA1, BRCA2, Clorf144, CARD11, CBL, CCND1, CCND2, CCND3, CCNE1, CDH1, CDH2, CDH20, CDH5, CDK4, CDK6, CDK8, CDKN2A, CDKN2B, CDKN2C, CEBPA, CHEK1, CHEK2, CRKL,

CRLF2, CTNNB1, CYP1B1, CYP2C19, CYP2C8, CYP2D6, CYP3A4, CYP3A5, DNMT3A, DOT1L, DPYD, EGFR, EPHA3, EPHA5, EPHA6, EPHA7, EPHB1, EPHB4, EPHB6, EPHX1, ERBB2, ERBB3, ERBB4, ERCC2, ERG, ESR1, ESR2, ETV1, ETV4, ETV5, ETV6, EWSR1, EZH2, FANCA, FBXW7, FGFR3A, FGFR1, FGFR2, FGFR3, FGFR4, FLT1, FLT3, FLT4, FOXP4, GATA1, GNA11, GNAQ, GNAS, GPR124, GSTP1, GUCY1A2, HOXA3, HRAS, HSP90AA1, IDH1, IDH2, IGF1R, IGF2R, IKBKE, IKZF1, INHBA, IRS2, ITPA, JAK1, JAK2, JAK3, JUN, KDR, KIT, KRAS, LRP1B, LRP2, LTK, MAN1B1, MAP2K1, MAP2K2, MAP2K4, MCL1, MDM2, MDM4, MEN1, MET, MITF, MLH1, MLL, MPL, MRE11A, MSH2, MSH6, MTHFR, MTOR, MUTYH, MYC, MYCL1, MYCN, NF1, NF2, NKX2-1, NOTCH1, NPM1, NQO1, NRAS, NRP2, NTRK1, NTRK3, PAK3, PAX5, PDGFRA, PDGFRB, PIK3CA, PIK3R1, PKHD1, PLCG1, PRKDC, PTCH1, PTEN, PTPN11, PTPRD, RAF1, RARA, RB1, RET, RICTOR, RPTOR, RUNX1, SLC19A1, SLC22A2, SLC01B3, SMAD2, SMAD3, SMAD4, SMARCA4, SMARCB1, SMO, SOD2, SOX10, SOX2, SRC, STK11, SULTIA1, TBX22, TET2, TGFBR2, TMRSS2, TNFRSF14, TOP1, TP53, TPMT, TSC1, TSC2, TYMS, UGT1A1, UMPs, USP9X, VHL, and WT1.

[0190] In particular embodiments, the genetic lesion comprises a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, a change in copy number, or a gene fusion.

[0191] In one embodiment, the genetic lesion is a gene fusion that fuses the 3' coding region of the ALK gene to another gene.

[0192] In one embodiment, the genetic lesion is a gene fusion that fuses the 3' coding region of the ALK gene to the EML4 gene.

[0193] Illustrative examples of conditions suitable for fetal testing that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include but are not limited to: Down Syndrome (Trisomy 21), Edwards Syndrome (Trisomy 18), Patau Syndrome (Trisomy 13), Klinefelter's Syndrome (XXY), Triple X syndrome, XYY syndrome, Trisomy 8, Trisomy 16, Turner Syndrome (XO), Robertsonian translocation, DiGeorge Syndrome and Wolf-Hirschhorn Syndrome.

[0194] Illustrative examples of alleles suitable for paternity testing that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include but are not limited to 16 or more of: D20S1082, D6S474, D12ATA63, D22S1045, D10S1248, D1S1677, D11S4463, D4S2364, D9S1122, D2S1776, D10S1425, D3S3053, D5S2500, D1S1627, D3S4529, D2S441, D17S974, D6S1017, D4S2408, D9S2157, Amelogenin, D17S1301, DIGATA113, D18S853, D20S482, and D14S1434.

[0195] Illustrative examples of genes suitable for predicting the response to drug treatment that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include, but are not limited to, one or more of the following genes: ABCB1 (ATP-binding cassette, sub-family B (MDR/TAP), member 1), ACE (angiotensin I converting enzyme), ADH1A (alcohol dehydrogenase 1A (class I), alpha polypeptide), ADH1B (alcohol dehydrogenase 1B (class I), beta polypeptide),

ADH1C (alcohol dehydrogenase 1C (class I), gammopoly-peptide), ADRB1 (adrenergic, beta-1-, receptor), ADRB2 (adrenergic, beta-2-, receptor, surface), AHR (aryl hydro-carbon receptor), ALDH1A1 (aldehyde dehydrogenase 1 family, member A1), ALOX5 (arachidonate 5-lipoxy-genase), BRCA1 (breast cancer 1, early onset), COMT (catechol-O-methyltransferase), CYP2A6 (cytochrome P450, family 2, subfamily A, polypeptide 6), CYP2B6 (cytochrome P450, family 2, subfamily B, polypeptide 6), CYP2C9 (cytochrome P450, family 2, subfamily C, poly-peptide 9), CYP2C19 (cytochrome P450, family 2, subfamily C, polypeptide 19), CYP2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6), CYP2J2 (cytochrome P450, family 2, subfamily J, polypeptide 2), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), CYP3A5 (cytochrome P450, family 3, subfamily A, polypeptide 5), DPYD (dihydropyrimidine dehydrogenase), DRD2 (dop-amine receptor D2), F5 (coagulation factor V), GSTP1 (glutathione S-transferase pi), HMGCR (3-hydroxy-3-meth-ylglutaryl-Coenzyme A reductase), KCNH2 (potassium voltage-gated channel, subfamily H (eag-related), member 2), KCNJ11 (potassium inwardly-rectifying channel, subfamily J, member 11), MTHFR (5,10-methylenetetrahydro-folate reductase (NADPH)), NQO1 (NAD(P)H dehydroge-nase, quinone 1), P2RY1 (purinergic receptor P2Y, G-protein coupled, 1), P2RY12 (purinergic receptor P2Y, G-protein coupled, 12), PTGIS (prostaglandin 12 (prosta-cyclin) synthase), SCN5A (sodium channel, voltage-gated, type V, alpha (long QT syndrome 3)), SLC19A1 (solute carrier family 19 (folate transporter), member 1), SLC01B1 (solute carrier organic anion transporter family, member 1B1), SULT1A1 (sulfotransferase family, cytosolic, 1A, phenol-preferring, member 1), TPMT (thiopurine S-meth-ylyltransferase), TYMS (thymidylate synthetase), UGT1A1 (UDP glucuronosyltransferase 1 family, polypeptide A1), VDR (vitamin D (1,25-dihydroxyvitamin D3) receptor), VKORC1 (vitamin K epoxide reductase complex, subunit 1).

[0196] Illustrative examples of medical conditions that can be detected, identified, predicted, diagnosed, or moni-tored with the compositions and methods contemplated herein include, but are not limited to: stroke, transient ischemic attack, traumatic brain injury, heart disease, heart attack, angina, atherosclerosis, and high blood pressure.

[0197] Illustrative examples of pathogens that can be screened for with the compositions and methods contemplated herein include, but are not limited to: bacteria fungi, and viruses.

[0198] Illustrative examples of bacterial species that can be screened for with the compositions and methods contemplated herein include, but are not limited to: a *Mycobacte-rium* spp., a *Pneumococcus* spp., an *Escherichia* spp., a *Campylobacter* spp., a *Corynebacterium* spp., a *Clostridium* spp., a *Streptococcus* spp., a *Staphylococcus* spp., a *Pseudomonas* spp., a *Shigella* spp., a *Treponema* spp., or a *Salmonella* spp.

[0199] Illustrative examples of fungal species that can be screened for with the compositions and methods contemplated herein include, but are not limited to: an *Aspergillus* spp., a *Blastomyces* spp., a *Candida* spp., a *Coccidioides* spp., a *Cryptococcus* spp., dermatophytes, a *Tinea* spp., a *Trichophyton* spp., a *Microsporum* spp., a *Fusarium* spp., a

*Histoplasma* spp., a *Mucoromycotina* spp., a *Pneumocystis* spp., a *Sporothrix* spp., an *Exserophilum* spp., or a *Cla-dosporium* spp.

[0200] Illustrative examples of viruses that can be screened for with the compositions and methods contemplated herein include, but are not limited to: Influenza A such as H1N1, H1N2, H3N2 and H5N1 (bird flu), Influenza B, Influenza C virus, Hepatitis A virus, Hepatitis B virus, Hepatitis C virus, Hepatitis D virus, Hepatitis E virus, Rotavirus, any virus of the Norwalk virus group, enteric adenoviruses, parvovirus, Dengue fever virus, Monkey pox, Mononegavirales, Lyssavirus such as rabies virus, Lagos bat virus, Mokola virus, Duvenhage virus, European bat virus 1 & 2 and Australian bat virus, Ephemerovirus, Vesiculovirus, Vesicular Stomatitis Virus (VSV), Herpesviruses such as Herpes simplex virus types 1 and 2, varicella zoster, cytomegalovirus, Epstein-Bar virus (EBV), human herpesviruses (HHV), human herpesvirus type 6 and 8, Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), Harvey murine sarcoma virus (Ha-MuSV), murine mammary tumor virus (MuMTV), gibbon ape leukemia virus (GaLV), feline leukemia virus (FLV), spumavirus, Friend murine leukemia virus, Murine Stem Cell Virus (MSCV) and Rous Sarcoma Virus (RSV), HIV (human immunodeficiency virus; including HIV type 1, and HIV type 2), visna-maedi virus (VMV) virus, the caprine arthritis-encephalitis virus (CAEV), equine infectious anemia virus (EIAV), feline immunodeficiency virus (FIV), bovine immune deficiency virus (BIV), and simian immunodeficiency virus (SIV), papilloma virus, murine gamma-herpesvirus, Arenaviruses such as Argentine hemorrhagic fever virus, Bolivian hemorrhagic fever virus, Sabia-associated hemorrhagic fever virus, Venezuelan hemorrhagic fever virus, Lassa fever virus, Machupo virus, Lymphocytic choriomeningitis virus (LCMV), Bunyaviridae such as Crimean-Congo hemorrhagic fever virus, Hantavirus, hemorrhagic fever with renal syndrome causing virus, Rift Valley fever virus, Filoviridae (filovirus) including Ebola hemorrhagic fever and Marburg hemorrhagic fever, Flaviviridae including Kaysanur Forest disease virus, Omsk hemorrhagic fever virus, Tick-borne encephalitis causing virus and Paramyxoviridae such as Hendra virus and Nipah virus, variola major and variola minor (smallpox), alphaviruses such as Venezuelan equine encephalitis virus, eastern equine encephalitis virus, western equine encephalitis virus, SARS-associated coronavirus (SARS-CoV), West Nile virus, and any encephalitis causing virus.

[0201] Illustrative examples of genes suitable for monitoring an organ transplant in a transplant recipient that can be detected, identified, predicted, diagnosed, or monitored with the compositions and methods contemplated herein include, but are not limited to, one or more of the following genes: HLA-A, HLA-B, HLA-C, HLA-DR, HLA-DP, and HLA-DQ.

[0202] In particular embodiments, a bioinformatic analysis is used to quantify the number of genome equivalents analyzed in the cfDNA clone library; detect genetic variants in a target genetic locus; detect mutations within a target genetic locus; detect genetic fusions within a target genetic locus; or measure copy number fluctuations within a target genetic locus.

#### E. Companion Diagnostics

[0203] In various embodiments, a companion diagnostic for a genetic disease is provided, comprising: isolating or obtaining cfDNA from a biological sample of a subject; treating the cfDNA with one or more end-repair enzymes to generate end-repaired cfDNA; ligating one or more adaptors to each end of the end-repaired cfDNA to generate a cfDNA library; amplifying the cfDNA library to generate a cfDNA clone library; determining the number of genome equivalents in the cfDNA clone library; and performing a quantitative genetic analysis of one or more biomarkers associated with the genetic disease in the cfDNA clone library, wherein detection of, or failure to detect, at least one of the one or more biomarkers indicates whether the subject should be treated for the genetic disease.

[0204] As used herein, the term "companion diagnostic" refers to a diagnostic test that is linked to a particular anti-cancer therapy. In a particular embodiment, the diagnostic methods comprise detection of genetic lesion in a biomarker associated with in a biological sample, thereby allowing for prompt identification of patients should or should not be treated with the anti-cancer therapy.

[0205] Anti-cancer therapy includes, but is not limited to surgery, radiation, chemotherapeutics, anti-cancer drugs, and immunomodulators.

[0206] Illustrative examples of anti-cancer drugs include, but are not limited to: alkylating agents such as thiotepa and cyclophosphamide (CYTOXANT<sup>TM</sup>); alkyl sulfonates such as busulfan, improsulfan and piposulfan; aziridines such as benzodopa, carboquone, meturedopa, and uredopa; ethylenimines and methylamelamines including altretamine, triethylenemelamine, trietylenephosphoramide, triethylenethiophosphoramide and trimethylololomelamine resueme; nitrogen mustards such as chlorambucil, chlornaphazine, chlorophosphamide, estramustine, ifosfamide, mechlorethamine, mechlorethamine oxide hydrochloride, melphalan, novembichin, phenesterine, prednimustine, trofosfamide, uracil mustard; nitrosoureas such as carmustine, chlorozotocin, fotemustine, lomustine, nimustine, raniimustine; antibiotics such as aclacinomysins, actinomycin, authramycin, azaserine, bleomycins, cactinomycin, calicheamicin, carabacin, carminomycin, carzinophilin, chromomycins, dactinomycin, daunorubicin, detorubicin, 6-diazo-5-oxo-L-norleucine, doxorubicin and its pegylated formulations, epirubicin, esorubicin, idarubicin, marcellomycin, mitomycins, mycophenolic acid, nogalamycin, olivomycins, peplomycin, pofiriomycin, puromycin, quelamycin, rodorubicin, streptonigrin, streptozocin, tubercidin, ubenimex, zinostatin, zorubicin; anti-metabolites such as methotrexate and 5-fluorouracil (5-FU); folic acid analogues such as denopterin, methotrexate, pteropterin, trimetrexate; purine analogs such as fludarabine, 6-mercaptopurine, thioguanine; pyrimidine analogs such as ancytabin, azacitidine, 6-azauridine, carmofur, cytarabine, dideoxyuridine, doxifluridine, encycitabine, floxuridine, 5-FU; androgens such as calusterone, dromostanolone propionate, epitostanol, mepitiostane, testolactone; anti-adrenals such as aminoglutethimide, mitotane, trilostane; folic acid replenisher such as frolinic acid; aceglatone; aldophosphamide glycoside; aminolevulinic acid; amsacrine; bestrabucil; bisantrene; edatraxate; defofamine; demecolcine; diaziquone; elformithine; elliptinium acetate; etoglucid; gallium nitrate; hydroxyurea; lentinan; lonidamine; mitoguazone; mitoxantrone; moperidol; nitrocrine; pentostatin; phenacet; pirarubicin; podophyllinic

acid; 2-ethylhydrazide; procarbazine; PSK®; razoxane; sizofiran; spirogermanium; tenuazonic acid; triaziquone; 2, 2',2"-trichlorotriethylamine; urethan; vindesine; dacarbazine; mannomustine; mitobronitol; mitolactol; pipobroman; gacytosine; arabinoside ("Ara-C"); cyclophosphamide; thiotepa; taxoids, e.g., paclitaxel (TAXOL®, Bristol-Myers Squibb Oncology, Princeton, N.J.) and doxetaxel (TAXOTERE®, Rhne-Poulenc Rorer, Antony, France); chlorambucil; gemcitabine; 6-thioguanine; mercaptopurine; methotrexate; platinum analogs such as cisplatin and carboplatin; vinblastine; platinum; etoposide (VP-16); ifosfamide; mitomycin C; mitoxantrone; vincristine; vinorelbine; navelbine; novantrone; teniposide; aminopterin; xeloda; ibandronate; CPT-11; topoisomerase inhibitor RFS 2000; difluoromethylomithine (DMFO); retinoic acid derivatives such as Targretin™ (bexarotene), Panretin™ (alitretinoin); ONTAK™ (denileukin diftitox); esperamicins; capecitabine; and pharmaceutically acceptable salts, acids or derivatives of any of the above. Also included in this definition are anti-hormonal agents that act to regulate or inhibit hormone action on cancers such as anti-estrogens including for example tamoxifen, raloxifene, aromatase inhibiting 4(5)-imidazoles, 4-hydroxytamoxifen, trioxifene, keoxifene, LY117018, onapristone, and toremifene (Fareston); and anti-androgens such as flutamide, nilutamide, bicalutamide, leuprolide, and goserelin; and pharmaceutically acceptable salts, acids or derivatives of any of the above.

[0207] Illustrative examples of immunomodulators include, but are not limited to: cyclosporine, tacrolimus, tresperimus, pimecrolimus, sirolimus, verolimus, laflimus, laquinimod and imiquimod, as well as analogs, derivatives, salts, ions and complexes thereof.

[0208] All publications, patent applications, and issued patents cited in this specification are herein incorporated by reference as if each individual publication, patent application, or issued patent were specifically and individually indicated to be incorporated by reference.

[0209] Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to one of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims. The following examples are provided by way of illustration only and not by way of limitation. Those of skill in the art will readily recognize a variety of noncritical parameters that could be changed or modified to yield essentially similar results.

## EXAMPLES

### Example 1: Accurate Detection of Rare Mutations Using Targeted Sequence Capture Technology

#### Purpose

[0210] The purpose of this experiment was to provide a direct proof-of-principle demonstration of rare variant detection using targeted sequence capture technology.

#### Background

[0211] Target sequence capture technology provides quantitative, sequence-based genetic analysis of nucleic acids and can be exploited to perform a combined mutational and copy number analysis of drug metabolism genes. The present inventors used targeted sequence capture technology and subsequent genetic analysis to detect rare sequence variants.

[0212] Genomic DNA inputs play a central role in rare variant detection, but quantitative analysis and control of genomic inputs places bounds on the estimated sensitivity of rare variant analysis. A genomic qPCR assay was used by the present inventors to estimate genomic inputs.

[0213] One experimental goal for rare variant analysis is to introduce 10-fold more genomic input as the target sensitivity of the assay. In other words, to measure variants with a sensitivity of 1% (1 in 100), then the experimental target is to input 1000 genomes. Downstream of sequencing, bioinformatics analysis reveals the number of unique reads, and this has the desirable quality of being both an orthogonal and a more direct measure of genomic inputs.

#### Summary

[0214] A cell line (ZR75-30) with known SNVs was admixed with a germ line DNA sample (NA12878) in a dilution series ranging from 1-to-1 through 1-to-1000. Target regions corresponding to known sequence differences were retrieved using targeted sequence capture technology and sequenced. Sequence variants that occur at a frequency of less than 1 per 1000 sequences were detected.

#### Methods

[0215] Capture probes: The following table shows a collection of 62 capture probes that were used in this experiment.

TABLE 1

60 base probe sequences used in the admix proof-of-concept study		
Target	SEQ ID NO:	Probe Sequence
BRAF	1	TAAACATTGGAAAGGTTCTAATTAACCAGGAGATCCAAAAGAAAGCGGT TCAAGTAGCA
	2	GATCTCAGTTTTGGTTAACTATGTATTTGGTATATGAAGCTCTGG GTTTGACACA
MYCN	3	GACAGATAAGCATACATATTAACATGGATATATGTGAATTCATTCAA ATGGTTCTCA
	4	AGCTCTTAGCCTTGGGGGATGACACTCTTGAGCGGACGTGGGACGCC TCGCTCTTA

TABLE 1-continued

<u>60 base probe sequences used in the admix proof-of-concept study</u>	
Target	SEQ ID NO: Probe Sequence
BRAF	5 AAGCCCCCACCGCCGCTTTCCAAAATAAACACCAGCCAGCGCCGAG CCCAGTCG 6 GCCTCCCTCCCCCTCCCGCCGAAGCGGCCGCTGGGCCCGCTCT CGTTATAAG
CDH1	7 GGTGGCAGCCAGGGGGCGCACTCTGCTCTGGCTGGGCCCTTCTCCC ATGTTTCTT 8 TTACACAACTTGGGCTGGACAACACTTGGGTCAAAGAACCTAAG AGTCTTCTG
EPHX2	9 TGATGAACTTGGCTGGATGGGCACAGGTAGGGTGCTTGTGCTTCA GTCAGATGAA 10 AATGAAAGAAAAGGAGGCCAGATTGCTACTCCTGGTCCCTGCCACACACT AGGTACCTA
BRCA1	11 ATTGACAATACCTACATAAAACTCTTCCAGAATGTTGTAAGTCTTAGT CATTAGGGAG 12 GGATTCCACCAACACTGTATTGATCATGTACCCATTTCCTCTAACCTAAC TTTATTGGTC
BRCA2	13 CAAAGGGGAAAAACCATCAGGACATTATTTAACACGGAAATATCTAACT GAAAGGCAA 14 CAGGCAGACCAACAAAGTCTTGTCCACCTTTAAAACAAATCACAT TTTCACAGAG
MYCN	15 CCCAGGCCAGCGGTCCGAACCCCTGGCGCATCCACGAAACTTGGCCAT AGCAGCGGGC
MYC_r1_F1	16 CGACTCATCTCAGCATTTAAAGTGATAAAAAAATAAAAGGCAAGTG GACTTCGGTG
MYC_r1_R1	17 CTCTGGCGCGACTGCGCGCTGCGCCAGGTTCCGACCAAGACCCCTT AACTCAAGAC
MYC_r2_F1	18 TTCTACTGCGACGAGGAGGAGAACTTCTACAGCAGCAGCAGAGCGA GCTGCAGCCC
MYC_r2_F2	19 ACCGAGCTGCTGGAGGAGACATGGTGAACCAGAGTTTACATCGGACCC GGACGACGAG
MYC_r2_F3	20 GCCGCCCTCAGAGTGCATGACCCCTCGTGGCTTCCCTACCCCT CAACGACAGC
MYC_r2_F4	21 GGCGGCTAGGGACAGGGCGGGTGGCAGCAGCTGAATTCTTCAG ATATCCTCGC
MYC_r2_R1	22 AGACGAGCTTGGCGCGCCGAGAACGCCGCTCCACATACAGTCTGGATG ATGATGTTTT
MYC_r2_R2	23 AGGAGAGCAGAGAATCCGAGGACGGAGAGAAGGCCTGGAGTCTGCGAG GCGCAGGACT
MYC_r2_R3	24 TAAGAGTGGCCGTTAAATAAGCTGCCAATGAAAATGGAAAGGTATCCA GCCGCCCACT
MYC_r2_R4	25 TTGTATTGTACAGCATTAATCTGGTAATTGATTATTTAATGTAACCTT GCTAAAGGAG
MYC_r3_F1	26 GAGGCCACAGCAAACCTCCTCACAGCCACTGGTCTCAAGAGGTGCCAC GTCTCCACAC
MYC_r3_F2	27 AGAGGAGGAACGAGCTAAACGGAGCTTTTGCCTGCGTACCGAGATC CCGGAGTTGG
MYC_r3_F3	28 TCCAACTTGACCCCTTGGCAGCAGGAGTAGTCCTCGAGTGGAGGGAGG CGCTGCGTAG
MYC_r3_R1	29 GCTTGGACGGACAGGATGTATGCTGGTTTTAAGGATAACTACCTT GGGGCCCTT

TABLE 1-continued

<u>60 base probe sequences used in the admix proof-of-concept study</u>		
Target	SEQ ID NO:	Probe Sequence
MYC_r3_R2	30	GCATTGATCATGCATTGAAACACAAGTTCATAGGTGATTGCTCAGGACATTCTGTTAGA
MYC_r3_R3	31	CGCCCGCGCCCTCCCAGCCGGTCCAGCCGGAGCCATGGGCCGGAGCCGCAGTGACCA
ERBB2r1r	32	CTCTGGCCCCGCCGGCGGGACCTCGGCGGGGCATCCACAGGGCAGGTTCCCGCCGCT
ERBB2r2f	33	GGCATGACTTGGAGTGAGTTGGATGGGTGGCCAGGTCTGAGAACGGTCCCGCCAGTAGCAGTG
ERBB2r2r	34	GCAGGGCACCTTCTGCCACCCACCTGTAAACAGAGGGCTCAGCCAGCTGGAGGCAG
ERBB2r3f	35	CCCAAGATCTCCAAGTACTGGGAACCCCAAGGGAGGCCCTGGGGGTGGCAGTGTCTTA
ERBB2r3r	36	CTAATGCACACAAAGCCTCCCCCTGGTTAGCAGTGGCCCTGGTCAGCTCTAACAA
ERBB2r4f	37	CTGCTCCTCTTTAGAAGGCAGGAGGGCCCAAGGGAAGCAGAAGGTGACAGGGAA
ERBB2r4r	38	TGGGCCAGTGGCGGGCAGGCAGTGGTTGTAAGTTGGAGTTGGCTGAGCTGGCTGGCTGAGGCT
ERBB2r5f	39	TCTGCTGCTGTTGTGCCTCTCTGTACTAACCGTCCTCTCGCTGTTAGACATCTCT
ERBB2r5r	40	CCCACCCCTCCATGTCACCTGTATGACACCTGCATTCCACCCGGCCCACTCC
ERBB2r6f	41	TGGGCCAGGTAGTCTCCCTAGAAGGTGATGCTGATGAGGGCTGGTGCAC
ERBB2r6r	42	GGTCCCCACCCCTTGACCTGGGGTAGAGCACATTGGCACAAAGCAAGGCACATA
ERBB2r7f	43	CACCCCTGCTGGTACTGCCCTATTGCCCTGGCACACCAGGGCAAAACAGCACAGTGAAA
ERBB2r7r	44	CCATTACAGAAACAAACCTCCCCACCAAATGAGAAAATGTGTTCTCCTGGCACT
ERBB2r8f	45	TTATTCTCTGTGCCTGGCACGGTAATGCTGCTCATGGTGGTGCACGAAGGGCAGGG
ERBB2r8r	46	GAAGGATAGGACAGGGTGGCTGGCCAGGCTGCATGCGCAGAGGGACAGGAACAGC
ERBB2r9f	47	GGGCCGGACCCGTATGCTCATGTGGCTGTTGACCTGTCCTGGTATGAAGGCTGAGACGG
ERBB2r9r	48	TCTGCTCCTGCCATCCCCAAGAGATGCTGCCACATCTGGATCCTCAGGACTCTGTCTGC
TYMSr2f	49	TCACGTCCCAGGGCAGTTTCTTCCCTGAAGAAAGTTGGATGGCATGATGATCTCTCCCA
TYMSr2r	50	GTGTTGAGAACAGACTACTGACTTCTAATAGCAGCGACTTCTTACCTGATGAAACCA
TYMSr3f	51	AAAAAAAGGATGGTCCATATGGGTGGTGTCAAGTGCCCACCTCCTAGCAAGTCAGCAG
TYMSr3r	52	CCCTCACAAAGGTCAAAGCTATACATCAGCTCCTGTGACATTGACTCATCCCGACACCTT
TYMSr4f	53	AACCCACCGAGATCTGCAAACCTTGAGGATGCACAGATGTCTTGTAGCATGGTCAA

TABLE 1-continued

<u>60 base probe sequences used in the admix proof-of-concept study</u>			
Target	SEQ ID NO:	Probe Sequence	
TYMSr4r	54	TGCCTCCCTCAGGTGCCCTGCACAAAACCAGATTGCTTCCCTCTAAGAG TATGGTTAGT	
TYMSr5f	55	GTTTACTTGCCTTAGCTGTGGCTTCAAAACCACCATCCCTCCTTAT CTTCCTCTGC	
TYMSr5r	56	CTCTGCAATTGTTCCCATAATTAAAGAACCTGAAGAGCTCAGTGTGGTA GGCTGGCAAG	
TYMSr6f	57	TTTTAAATGATGTTAAAGAACATTGAAACTAACATACTGTTCTGCTTCT CCCCGGGTT	
TYMSr6r	58	CCTGCCACCCTTCCTAAACTGAAGCCCCACATTGGAGCAGTCAT CTTTATCTTG	
TYMSr7f	59	GGTTGCGCTCCAATCATGTTACATAACCTACGGCAAGGTATCGACAGGAT CATACTCCTG	
TYMSr7r	60	GCACAGTTACATTGCCAGTGGAACATCCTTAAAAATAACTGATA GGTCACGGAC	
TYMSr1f	61	CGTCCCGCCGCGCCACTTGGCCTGCCTCCGCCCCGCGCCACTTCGCC TGCCTCCGTC	
TYMSr1r	62	CTGTAAGCGAGGAGGACATGCGTCCCCTCCCTCGCAGGATTGAGGTTA GGACTAAACG	

[0216] Capture probe modules were pooled from stock plates, combined with partner oligo 138 (SEQ ID NO: 63) (GTGAAAACCAGGATCAACTCCGTGCCAGTCA-CAT/3BioTEG/) and diluted to a final working concentration of 1 nM.

[0217] Genomic Analysis. Commercially-purchased genomic DNA from germ line sample NA12878 and cell line ZR75-30 was fragmented at a concentration of 10-20 ng/ $\mu$ L to a target fragment size of 500 bp on a Covaris sonication instrument. The DNA was purified with a 1:1 concentration of DNA purification beads and end-repaired using the New England Biolabs (NEB) Quick blunt kit at a final concentration of 15-30 ng/ $\mu$ L. The germ line and cell line DNAs were blended at ratios of 1:1, 10:1, 100:1 and 1000:1, respectively. Libraries were constructed, purified and quantified. The sample codes, library quantitation and inputs used for library construction are shown in Table 2.

[0218] Genomic libraries were pooled, denatured, combined with probe, hybridized and washed. The washed capture probe-tagged genomic library complexes were amplified with forward and reverse full-length primers, purified, and size-selected for 225-600 bp fragments on a Pippin-prep instrument. Finally, the captured material was sequenced using a 150-V3 Illumina sequencing kit.

#### Result

[0219] The paired capture probes that target BRAF (in two loci), MYCN and CDH1 were used to analyze the SNVs in these loci. The results are shown in Table 3.

TABLE 2

Adaptors and genomic analysis of libraries used as inputs					
Admix	Adaptor code	genomes/ $\mu$ L	# desired genomes	$\mu$ L into PCR	
1:1	NNNNNNNNCATGGCCGCAGG (SEQ ID NO: 64)	55	200	4	
10:1	NNNNNNNNNATCTTAGTGGCA (SEQ ID NO: 65)	66	200	3	
100:1	NNNNNNNNCGAACCTCGGAG (SEQ ID NO: 66)	64	1000	16	
1000:1	NNNNNNNNNGACTCCGATCCC (SEQ ID NO: 67)	77	10000	130	

TABLE 3

Mutation	Sample NA:ZR	Bioinformatic summary									
		Unique reads by tag and position					No unique filtering				
		Total # reads	change to A	change to C	change to G	change to T	Total # reads	change to A	change to C	change to G	change to T
BRAF-I326T-A-G	1 to 1	186		0	55	0	34728		6	11308	10
BRAF-I326T-A-G	10 to 1	171		0	10	0	46464		6	2615	2
BRAF-I326T-A-G	100 to 1	733		0	3	0	51540		10	89	2
BRAF-I326T-A-G	1000 to 1	3020		0	3	0	54565		6	19	7
BRAF-E26D-C-A	1 to 1	171	66		0	0	29192	12388		5	2
BRAF-E26D-C-A	10 to 1	183	12		0	1	38085	2573		1	4
BRAF-E26D-C-A	100 to 1	631	3		0	0	38913	110		1	2
BRAF-E26D-C-A	1000 to 1	2367	1		0	0	46623	13		2	2
CDH1-E243 <sup>a</sup> -G-T	1 to 1	180	0	0		54	23846	2	1		8556
CDH1-E243 <sup>a</sup> -G-T	10 to 1	208	0	0		12	42307	1	2		2658
CDH1-E243 <sup>a</sup> -G-T	100 to 1	728	0	0		4	42440	6	0		129
CDH1-E243 <sup>a</sup> -G-T	1000 to 1	2727	0	0		0	53632	5	1		7
MYCN-E270Q-G-C	1 to 1	158	0	47		0	20302	0	5587		0
MYCN-E270Q-G-C	10 to 1	186	0	14		0	35733	1	2663		2
MYCN-E270Q-G-C	100 to 1	566	0	3		0	35393	0	247		5
MYCN-E270Q-G-C	1000 to 1	2101	0	2		0	37223	0	15		2

[0220] Column 3 shows the total number unique read counts, which in turn provide bounds on the sensitivity of the assay. The estimated and measured genomic inputs were well within range of one another. The lightly shaded boxes highlight the SNV where the cell line sequence differed from the germ line sequence. In the absence of unique read filtering—shown on the right portion of the table—random base changes at these four selected positions occurred with measurable, non-zero frequencies. FIG. 1. By requiring that changes occur within unique read families, it became possible to sort true signal from error-prone noise. FIG. 2.

#### Example 2: A Novel Probe Design Effective for Comprehensive Sequencing of Target Regions in Highly Fragmented 2DNA

##### Purpose

[0221] The purpose of these experiments is to develop an assay system to reliably and reproducibly interrogate circulating DNAs.

##### Background

[0222] Analysis of circulating DNA from body fluids represents an exciting, but as yet, unrealized opportunity in molecular diagnostics. Genomic DNA is highly intact. Literature suggest that the average size of circulating DNA is about 150 bp, which correlates well to the size of DNAs wrapped around a single nucleosomal histone complex.

##### Summary

[0223] The technical parameters of targeted sequence capture technology contemplated herein were designed to accommodate highly fragmented DNA and to retain the ability to generate comprehensive sequence coverage of targeted DNA. Capture probe density was increased and the length of capture probe sequences was reduced from 60 nucleotide to 40 nucleotide to minimize uninformative sequence generation in the clone library. The human genome is littered with repetitive sequences and drastic fluctuations in base composition, thus, the suitability of implementing higher capture probe densities and shorter capture probes could not be conceded but required empirical validation of the new assay.

[0224] Conditions were established in which the shorter 40 mer capture probe sequences exhibit reliable and robust assay performance. In a first set of experiments, the assays were used to query two large regions—the coding regions for the tumor suppressor gene TP53 and the long, contiguous, intron 19 of the ALK oncogene, both of which are central to cancer diagnostics. In a second set of experiments, several high density pairwise capture probes that possess shorter 40 nucleotide capture probe sequences were used to interrogate known SNVs that reside in the NCI-H69 cell line.

[0225] The new high density shorter capture probes were successfully used to query short fragmented DNAs and the

results indicated that the assay design is well suited to sequencing of circulating DNAs that are found in the plasma fraction of blood.

**[0226]** Modified 40 mer Capture Probes: The capture probe sequences used to empirically validate the performance of the 40 mer capture probes are shown in Table 4.

TABLE 4

Name	SEQ ID NO: Sequence	Capture probes	
		60 mer Capture Probes	
PLP1_ex2_F	68 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGGGTTTGAGTGGCATGAGCTACCTACTGGATGTGCCTGACTGTTCCCTCTTCTTCCC		
PLP1_ex2_R	69 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTATCTCCAGGATGGAGA GAGGGAAAAAAAAGATGGGTCTGTGGAGGGCA70GGTACTT		
PLP1_ex3_F	70 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGAAAGAAGCCAGGTCTTC AATTATAAGATCCCTGGTCTGTTGTCTACCTGTTAATG		
PLP1_ex3_M	71 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCAAGACTCGGCCAATTT TCCCCCACCCCTGTTATTGCCACAAAATCTGAGGATGATC		
CYP2D6_F	72 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAAGCACCTAGCCCCATTCTGCTGAGCAGGGTGGCAGGTACCCAGACTGGGAGGTAA		
CYP2D6_R	73 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAGTCGGTGGGGCCAGGAT GAGGCCCAAGTCAGTGTACACATGGCTGCTGCCCTCAGCTCT		
chrX_15_F	74 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCCCTGGCCCTCAGCCAGTA CAGAAAGTCATTGTCAAGGCCCTCAGTGGCAGACGTGCTC		
chrX_15_R	75 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAGAATTCTAGGCCAGCTA TAAATCTGTGAAACGCTGCACACAATCTAGCACACAAGA		
chrX_69_F	76 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTTACCTCCCTCAGTTT GTTGCTTCAGAAAACAAGAATCTTCTCATGAAATCATG		
chrX_69_R	77 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCAGGGTATCTATTATCC CCATTTCCTCACAAAGGAAACCAAGATAAAAGTTAAATGG		
KRAS_ex1_F	78 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTTACCTTAAAGACA TCTGCTTCTGCCAAAATTATGTGCTGAACCTAAACTTAC		
KRAS_ex1_R	79 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTTCCAGTAAATTACTCT TACCAATGCAACAGACTTAAAGAAGTTGTGTTTACAATGC		
KRAS_ex2_F	80 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTAAATGACATAACAGTTA TGATTTCAGAAAACAGATCTGTATTTCAGTGTACT		
KRAS_ex2_R	81 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGACAGGTTGAAAGATA TTTGTGTTACTAATGACTGTGCTATAACTTTTTCTTCC		
MYC_r2_F1	82 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTGTGGCGCACTGCGCGTGCAGCTGGTCCGCACCAAAGACCCCTTAACATCAAGAC		
MYC_r2_R1	83 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGGCGGTAGGGGACAGGG GCGGGGTGGGCAGCAGCTCGAATTCTTCCAGATATCCTCGC		
MYC_r2_F3	84 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACACCGAGCTGTGGGAGGA GACATGGTGAACAGAGTTCTGCGACCCGACGAGCAG		
MYC_r2_R3	85 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAGGAGAGCAGAGAATCCG AGGACGGAGAGAACGGCGTGGAGTCCTGCGACGGCGCAGGACT		
SRY_r1_F	86 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTGTAAGTTACCGTAAAA AGGAGCATCTAGGTAGGTCTTGTAGCCAATGTTACCGGATT		
SRY_r1_M3	87 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACATGGCATTCTTCCAGG AGGCACAGAAATTACAGGCCATGCACAGAGAGAAATACCGGA		
VHL_r3_F	88 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTGTTGTCCTTGTAC TGAGACCCTAGTCGCCACTGAGGATTGGTTGCCCTTC		
VHL_r3_R	89 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACATCAAGACTCATCAGTAC CATAAAAGCTGAGATGAAACAGTGTAAAGTTCAACAGAAAT		

TABLE 4 - continued

Capture probes	
Name	SEQ ID NO: Sequence
UGT1A1_r_4F	90 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC TCAGAGATGTAACGTGTCAGACATCCTCCCTATTTGCATCTCA
UGT1A1_r_4R	91 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC ATGAATGCCATGACCAAACATTCTCTGTATCTCTCTTCTT
TNFRSF14_r3_F	92 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC AGGGGCCTCCCGCAGACTTGCAGTCCCCTCTCTGGCG
TNFRSF14_r3_R	93 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC AGGCTGCCAACGCGAGGTGGCGGCTGTGCTGGCCTT
RUNX1_r4_F	94 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC TTGCCATGAAACGTGTTCAAGCATAGTTGACAGATAACG
RUNX1_r4_R	95 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC ATAACATCCCTGATGTCCTGCATTGTCCTTGACTGGTGT
RHD_r5_F	96 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACACCCCTGAGGCTCAGA CCTTGGAGCAGGAGTGATTCTGGCCAACCCCTCTCG
RHD_r5_R	97 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTCCAGCTGTGAAAC TCCTGTTAGACCCAAGTGCTGCCAACGGCAGCGCCCTGCTC
PTEN_r5_F	98 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTACTGTGTTAATTAAAAAAT TCAAGAGTTTTTTCTATTCTGAGGTATCTTTACCA
PTEN_r5_R	99 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACCCAAAATCTGTTTCAA TAAATTCTCAGATCCAGGAAGAGGAAAGGAAAACATCAA
EP300_r18_F	100 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTACTCCATCTCCGTA AAATAGTGAGACTTGAGTAATGTTGATGTCACTTGTCTTC
EP300_r18_R	101 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACAGTCACCACTATTAT TCTAGGTATCCAGAAAAGTTAAAGTCAAATCTGAAACACAT
VHL_r1_F	102 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACGCCCCGCGTCCGACCCG CGGATCCCGGGCGTCCGCCCGGGTGGATCGGGAG
VHL_r1_R	103 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACCCATACGGCAGCACGAC GCGCGACTGCGATTGCGAGATGACCTGGAGGGCTCGCG
VHL_r1_M1	104 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTAGAGGGGCTTCAGACCG TGCTATCGCCCTGCTGGTCGGGCTAAGCGCCGGCCGT
VHL_r1_M2	105 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGGCGCCGAGGAGGAGATG GAGGCCGGCGGCCGCGGCCCGTGCTGCGCTCGGTGAACCTCG
VHL_r2_F	106 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGGGTGGGCCACCGTGC CAGCCACGGGTGTGGCTCTTAACACCTTGCTGTCCGA
VHL_r2_R	107 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACAGGGTCTATCCTGTAC TTACCCACAACACCTTATCTTTTAAAGTAAAACGTCAGT
VHL_r3_F	108 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTTGTGCTCCCTGTAC TGAGACCCCTAGTCTGCCACTGAGGATTGGTTTGGCCCTTC
VHL_r3_R	109 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTCAAGACTCATCAGTAC CATCAAAGCTGAGATGAAACAGTGTAAAGTTCAACAGAAAT
40 mers	
PLP1_ex2_F_40	110 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCAACCTACTGGATGTGCCTG ACTGTTCCCTCTCTTCTT
PLP1_ex2_R_40	111 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGGGGGGGGGGGGGG CTGTGTGGAGGGCAGGTACTT
PLP1_ex3_F_40	112 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTTAATAAGATTCCCTGGT CTCGTTGTCTACCTGTAAATG

TABLE 4 - continued

Capture probes		
Name	SEQ ID NO: Sequence	
PLP1_ex3_M_40	113 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCCCCACCCCTGTTATTG CCACAAAATCCTGAGGATGATC	
CYP2D6_F_40	114 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCTGAGCAGGAGGGCA GGTACCCCAACTGGGAGGTAA	
CYP2D6_R_40	115 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCCAGTCTGTTCACAC ATGGCTGCTGCCTCTCAGCTCT	
chrX_15_F_40	116 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGAAAGTCATTGTCAAGG CCTTCAGTTGGCAGACGTGCTC	
chrX_15_R_40	117 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAATCTGTGAAACGCTGC CACACAATCTTAGCACACAAGA	
chrX_69_F_40	118 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGCTGCAAACACAGA ATCTTCTCTCCATGAAATCATG	
chrX_69_R_40	119 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACATTTCTCACAAAGGAAA CCAAGATAAAAGGTTAACATGG	
KRAS_ex1_F_40	120 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGCTTCTGCCAAAATTA ATGTGCTGAACCTAACATTAC	
KRAS_ex1_R_40	121 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCCAAATGCAACAGACTTTA AAGAAGTTGTGTTTACATGC	
KRAS_ex2_F_40	122 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACATTTGCAGAAAACAGAT CTGTATTCTTCACTGTTACT	
KRAS_ex2_R_40	123 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGTGTTACTAATGACTGT GCTATAACTTTTTCTTCC	
MYC_r2_F_1_40	124 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTGCCAGGTTCCGCAC CAAGACCCCTTAACTCAGAC	
MYC_r2_R_1_40	125 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGGGTGGCAGCAGCTCG AATTCTCCAGATATCCTCGC	
MYC_r2_F_3_40	126 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCATGGTAACCAGAGATT CATCTGCGACCCGGACGACGAG	
MYC_r2_R_3_40	127 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGACGGAGAGAAGGGCGCTG GAGCTTGCAGGGCGCAGGACT	
SRY_r1_F_40	128 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGAGCATCTAGGTAGGTCT TTGTAGCCAATGTTACCCGATT	
SRY_r1_M3_40	129 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCACAGAAATTACAGGCC ATGCACAGAGAGAAATACCCGA	
VHL_r3_F_40	130 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAGACCCTAGTCTGCCACT GAGGATTGGTTTGCCTTC	
VHL_r3_R_40	131 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTAAAAGCTGAGATGAAA CAGTGTAAAGTTCAACAGAAAT	
UGT1A1_r_4F_40	132 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAGAGATGTAACGTGCTGAC ATCCTCCCTATTTCATCTCA	
UGT1A1_r_4R_40	133 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGAATGCCATGACCAAAGT ATTCTTCTGTATCTCTTCTT	
TNFRSF14_r3_F_40	134 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGGCCTCCGCAGACTTG CGAAGTTCCCACTCTCTGGCG	
TNFRSF14_r3_R_40	135 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCTGCCAAGCGGAGGCT GGGCCGGCTGTGCTGCCCTT	
RUNX1_r4_F_40	136 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCCATGAAACGTGTTCA AGCATAGTTTGACAGATAACG	

TABLE 4 -continued

Capture probes			
Name	SEQ ID NO: Sequence		
RUNX1_r4_R_40	137 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAACATCCCTGATGTCCTGC ATTTGTCCTTGACTGGTCTT		
RHD_r5_F_40	138 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTGGAGCAGGAGTGTGA TTCTGGCCAACCACCCCTCTCG		
RHD_r5_R_40	139 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTGGTAGACCCAAGTGCT GCCCAAGGGCAGGCCCTGCTC		
PTEN_r5_F_40	140 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAAGAGTTTTTTCTTA TTCTGAGGTTATCTTTTACCA		
PTEN_r5_R_40	141 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAATTCTCAGATCCAGGAA GAGGAAAGGAAAAACATCAAAA		
EP300_r18_F_40	142 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACATAGTGAGACTTGAGTAA TGTTTGATGTCACTTGCTTTC		
EP300_r18_R_40	143 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACTAGGTATCCCAGAAAAGT TAAAGTCAAATCTGAAACACAT		
VHL_r1_F_40	144 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGATCCCGCGCGTCCGGC CCGGGTGGCTGGATCGGGAG		
VHL_r1_R_40	145 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCGGACTGCGATTGCAGA AGATGACCTGGGAGGGCTCGCG		
VHL_r1_M1_40	146 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACCTATCGTCCCTGCTGGGT CGGGCCTAACGCGCCGGCCCGT		
VHL_r1_M2_40	147 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGGCCGGCGCCGGCC CGTGCTCGCCTCGGTGAACCTCG		
VHL_r2_F_40	148 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACGCCACCGGTGTGGCTTT TAACAAACCTTGCTTGCTCCGA		
VHL_r2_R_40	149 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACACCAACAAACCTTATCT TTTTAAAAAAGTAAACGTCACT		
VHL_r3_F_40	150 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACAGACCCTAGTCTGCCACT GAGGATTGGTTTTGCCCTTC		
VHL_r3_R_40	151 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCACACTAAAAGCTGAGATGAAA CAGTGTAAGTTCAACAGAAAT		

[0227] The performance of 40 mer capture probes was compared to that of 60 mer capture probes. The 40 mer was designed from the 60 mer by removing 20 nucleotides from the 5' end of the 60 mer. Although the 3' end of both capture probe sets are identical with respect to the sequences that are copied from captured genomic clones, the probe sequence signature (Read 2 of the paired end read) is different between the 60 mer and 40 mer probe sets. This design is useful

because it allows the capture probes to be multiplexed during sequencing and their performance subsequently analyzed during downstream bioinformatics deconvolution.

[0228] Genomic samples: A pool of 12 genomic DNA samples (chosen from a Coriell human panel of 112 human genomic DNAs) was used as the target DNA. The 12 samples were broken into four sets of four samples each, as shown in detail in Table 5.

TABLE 5

Sample Summary						
Capture probes	Wash temp.	Sample Code	Samples			
60 mer	50 C.	AAT GM20291	M Americas	AFRICAN ANCESTRY IN SOUTHWEST USA		
		CTA GM19373	M W African	LUHYA IN WEBUYE, KENYA		
		GGG HG00428	F AsianE	HAN CHINESE SOUTH		
		TCC HG01624	M Euro	IBERIAN POPULATIONS IN SPAIN		

TABLE 5-continued

Sample Summary					
Capture probes	Wash temp.	Sample Code		Samples	
40 mer	50 C.	AAT	GM20291	M Americas	AFRICAN ANCESTRY IN SOUTHWEST USA
		CTA	GM19373	M WAfrican	LUHYA IN WEBUYE, KENYA
		GGG	HG00428	F AsianE	HAN CHINESE SOUTH
		TCC	HG01624	M Euro	IBERIAN POPULATIONS IN SPAIN
60 mer	47 C.	AGA	HG02489	M Americas	AFRICAN CARIBBEAN IN BARBADOS
		CCT	HG01108	F Americas	PUERTO RICAN IN PUERTO RICO
		GAC	GM19011	F AsianE	JAPANESE IN TOKYO, JAPAN
		TTG	GM18946	F AsianE	JAPANESE IN TOKYO, JAPAN
40 mer	47 C.	AGA	HG02489	M Americas	AFRICAN CARIBBEAN IN BARBADOS
		CCT	HG01108	F Americas	PUERTO RICAN IN PUERTO RICO
		GAC	GM19011	F AsianE	JAPANESE IN TOKYO, JAPAN
		TTG	GM18946	F AsianE	JAPANESE IN TOKYO, JAPAN
60 mer	44 C.	ATC	NA13783	F NA13783	GM13783
		CAG	HG03700	F Asians	PUNJABI IN LAHORE, PAKISTAN
		GCA	HG03367	M WAfrican	ESAN FROM NIGERIA
		TGT	NA22991	F NA22991	GM22991
40 mer	44 C.	ATC	NA13783	F NA13783	GM13783
		CAG	HG03700	F Asians	PUNJABI IN LAHORE, PAKISTAN
		GCA	HG03367	M WAfrican	ESAN FROM NIGERIA
		TGT	NA22991	F NA22991	GM22991

[0229] Hybridization, washing and sequencing: Six different hybridization conditions were used to hybridize the 60 mer and 40 mer probes to the genomic target DNA:

- [0230] 1) 60mer probes washed at 50° C.
- [0231] 2) 40mer probes washed at 50° C.
- [0232] 3) 60mer probes washed at 47° C.
- [0233] 4) 40mer probes washed at 47° C.
- [0234] 5) 60mer probes washed at 44° C.
- [0235] 6) 40mer probes washed at 44° C.

For each experiment, the capture probe oligos were combined with partner oligo; the final concentration of duplex capture probe was 1 nM for each capture probe.

[0236] Each hybridization reaction had ~2.5 µg of genomic library in 40 µL total volume. Each sample was heated to 98° C. for 2 min then cooled on ice. 20 µL of capture probe and 90 µL of hybridization buffer were added and the hybridization mix was incubated for 24 hours starting at 80° C. and decreasing one degree every 48 minutes to 50° C. The complexes were bound to 20 µL of streptavidin beads in 1 mL total volume of TEzero buffer+0.05% Tween20 (TT). The beads were washed 3 times, 5 min each with 200 µL of TT, and once at 45° C. for 5 min in wash buffer. The beads were then washed with TEzero and each reaction was resuspended in 20 µL TEzero. The complexes were then PCR amplified with full length forward (ACA2\_FLFP; SEQ ID NO: 152; AATGATACGGCGACCACCGAGATCTACACGT-CATGCAGGACCAGAGAAATTCA (CA) and full length reverse (CAC3\_FLRP; SEQ ID NO: 153; CAAGCAGAACGACGGCAT-ACGAGATGTGACTGGCACGGAGTT-GATCCTGGTTTCA (C) primers.

[0237] Following amplification and purification, the resulting product masses were measured and equal masses were pooled for sequencing.

#### Results—Modified 40mer Primers

[0238] The capture probe performance as a function of length and wash temperature is shown graphically in FIG. 3. Overall, the 40 mer capture probes performed as well as the

60 mer capture probes with 44° C. and 47° C. washes. With the 50° C. wash, the 40 mer capture probes exhibit sporadic behavior. These data empirically validate the use of 40 mer capture probes and wash temperatures in the 44° C. to 47° C. range when using these reagents.

#### Methods—High Density 40 Mers

[0239] High Density 40 mers: In general, sequence capture probes are designed using specific “rules.” For example, regions of redundant sequence or that exhibit extreme base composition biases are generally avoided. One key implication of the requirement for high probe density and close spacing of probes along target regions is that there is little or no latitude to move probes in order to accommodate any such probe design rules. In this study, probes were designed based solely on their position relative to one another with no consideration of probe binding sequences; thus, use of this high density approach required empirically validating that the hybridization and processing methods would accommodate such a collection of probes.

[0240] The human ALK gene encodes a protein kinase important in early development, but normal ALK gene expression is essentially undetectable in normal adults. Oncogenic ALK fusions are created when intron 19 of ALK undergoes illegitimate recombination to fuse the kinase encoding portion of ALK to the 5' end of another gene. Such gene fusions often cause ectopic expression of the ALK kinase, which in turn is important in driving the inappropriate cell proliferation observed in pulmonary tumors. In lung cancer, this “other gene” is often EML4, but other fusion partners have also been detected. To create an assay that can detect any possible ALK gene fusion event, 40 nucleotide probes were designed that were placed at 80 nucleotide intervals in intron 19 of ALK. These probes were oriented such that they are antisense relative to the gene (FIG. 4). This means that their 3' terminus extends and copies genic regions that are 5' to their hybridization site. When fusion genes are present, probe extension from probes positioned near the fusion junction copy junction sequences. The DNA sequences resulting from these junction clones

have fusion partner sequences at their 5' end, the fusion junction sequence, and ALK intron 19 sequences at their 3' ends (FIG. 4B).

[0241] Another important diagnostic target in cancer is the TP53 gene. It encodes a tumor suppressor, and it is often inactivated by mutations in cancers. Mutations that can inactivate gene function are scattered throughout the gene, and hence conclusive sequence-based assays for TP53 inactivating mutations must address the entire coding region and untranslated regions (UTRs) of the gene. Because circulat-

ing DNA fragments are short, high density probes were used to interrogate all target regions of the TP53 gene. Unlike ALK, probes for TP53 are placed in both possible orientations (FIG. 5). At high probe densities, the cumulative coverage from multiple probes provides uniform deep sequencing coverage of target regions.

[0242] The collection of 105 probes used in this study is shown in Table 6. In addition to probes that target the fusion-prone region of ALK and the coding regions of TP53, probes that cover known SNVs in the cell line DNA were also included.

TABLE 6

Name_target_region	SEQ_ID NO.: Probe sequence
ALK_chr2:29446208_fusion_f	154 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCGAATGAGGGTGATGTTTCGGCGGCACCTCC TTCAGGT
ALK_chr2:29446288_fusion_f	155 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTTGTAGTCGGTCATGATGGTCGAGGTGCGGAG CTTGCTC
ALK_chr2:29446368_fusion_f	156 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGCAGCTCTGGTGCCTCCGGCGTAACGTGAG GTGGGTG
ALK_chr2:29446448_fusion_f	157 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCTACACAGGCCACTCCTACAGGAAGCCTCCCT GGATCTC
ALK_chr2:29446528_fusion_f	158 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGAAATACTAATAAAATGATTAAAGAAGGTGTTG CTTTAAT
ALK_chr2:29446608_fusion_f	159 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTATATGGAAAATAATTATTGTATTATATAGGG CAGAGTC
ALK_chr2:29446688_fusion_f	160 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CATTAGACCCAAATAGGTCTGCAGATTATTAG AAGAAAT
ALK_chr2:29446768_fusion_f	161 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTGAACCAGCAGACTGTGTTGCAAGTATAACCC CACGTGA
ALK_chr2:29446848_fusion_f	162 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGCCATGGAGCCTAAGGAAGTTTCAGCAAGGCC TAAGGGG
ALK_chr2:29446928_fusion_f	163 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCCAGGAATTGGCCTGCCTTAGTATTCTGCTG TGCTCAG
ALK_chr2:29447008_fusion_f	164 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTTTGAGGGTGCAGCTGGGATCTGGTCAGTGT GTTTCCT
ALK_chr2:29447088_fusion_f	165 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCACATCATGAAAAGATCTCTGAATTGGTGTCTG GGGATCT
ALK_chr2:29447168_fusion_f	166 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTGAGGACCGAGTCACAGGACCTTTGGACTGC AGTTTC
ALK_chr2:29447248_fusion_f	167 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTAACCACTGCCACTCCCCACCCCTCTAGGGTGT CAATGAA
ALK_chr2:29447328_fusion_f	168 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGAGCTCTACCAATGTGAGTGACCAATTACTC CTACATG

TABLE 6-continued

Name_target region	SEQ ID NO:	Probe sequence
ALK_chr2:29447408_fusion_f	169	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAAATTGTGATTCACTGGTAGATTCTGTGTGT AAAGCCC
ALK_chr2:29447488_fusion_f	170	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTATGTGCTCAGTCCCTCCTATGCAATGGAC CGACCGT
ALK_chr2:29447568_fusion_f	171	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTGTAAATTGCCGAGCACGTAGTAACCATGCAA CAAGTGT
ALK_chr2:29447648_fusion_f	172	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTGGGACACAGTGTGCTGCCATCTCCCTCT ACCGGCA
ALK_chr2:29447728_fusion_f	173	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAAGAGCCTTCCTCTGCCCTTTCAAGCCCT GCCCATC
ALK_chr2:29447808_fusion_f	174	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGACCACACTGAGTTCTGTGACCTGCAGGTCA GCTCACC
ALK_chr2:29447888_fusion_f	175	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTTCCCTATCTCTGCCCTGGAGGGTGGTGAGG GCTGGTT
ALK_chr2:29447968_fusion_f	176	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAAACAGGAGCTGCCGGTGGAAGCATGTGGGA GCTAGAA
ALK_chr2:29448048_fusion_f	177	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGACACTGAAGGAGCTCCCACCCCTGATCAG CCAGGAG
ALK_chr2:29448128_fusion_f	178	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGGAACTGCACTGCTCTGGTGGGGGAAGGTT GGGAGCT
ALK_chr2:29448208_fusion_f	179	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CACCCAATTCCAGGGACTAGCATAACGAAGTGAC ACCTTGG
ALK_chr2:29448288_fusion_f	180	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGCCCTGGAGTCCCTGGGCTCTGTGC ACTCACC
MYCNrlf_40	181	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGAAGCACCCCCGGTATTAAACGAACGGGGCG GAAAGAA
MYCNrlr_40	182	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCTAACAAAGGGGACCGCACCCGGGCTCCAGTGC CCCAGGG
MYCNrlf2_40	183	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGGGGGACTGGTGGCCTCACCCCAACCC GGTCATC
MYCNrlr2_40	184	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCGCGCTCCAGCTCTCGCGGGCGAGAACCGC TCCACAT
MYCNrlf3_40	185	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCCACCCGGCGCCAGTGCCTGGATCCCACCG TGGTCTT
MYCNrlr3_40	186	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGGCACGGGCGCTGGCTCGCGCTGTTCACGGG AAAGGGG

TABLE 6-continued

Name_target region	SEQ ID NO: Probe sequence
MYCNr2f_40	187 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAACATGGATATAATATGTGAATTCAATTCAAATG GTTCTCA
MYCNr2r_40	188 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTAAACCAACATTCTTAATGTCAACACAATGTT GTTTAAA
MYCNr2f2_40	189 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCCCTACGTGGAGAGTGAGGATGCACCCCCACAG AAGAAGA
MYCNr2r2_40	190 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CATGACACTCTTGAGCGGACGTGGGACGCCCTCG CTCTTTA
MYCNr2f3_40	191 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTCTCACGCTCAGGGACCACGTGCCGGAGTTGGT AAAGAAT
MYCNr2r3_40	192 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCAGTGCCCTTTCAAAATGACCACCTTGGCGG CCTTCTC
TP53_chr17:7579779:region_1:75nt : -59:-20:f	193 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGCTAGGGGGCTGGGGTGGGGTGGGGTGGTGG GCCTGCC
TP53_chr17:7579838:region_1:75nt : 1:40:f	194 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCAGTTTCATAGGTCTGAAAATGTTCTGACT CAGAGGG
TP53_chr17:7579878:region_1:75nt : 41:+5:r	195 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCTGCCATGGAGGAGCCGCAGTCAGATCCTAGCG TCGAGCC
TP53_chr17:7579932:region_1:75nt : +20:+59:r	196 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTCATGCTGGATCCCCACTTTCCCTTGAGCAGCA GCCAGAC
TP53_chr17:7579640:region_2:23nt : -59:-20:f	197 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAGCCCCCCAGCCCTCCAGGTCCCAGCCCTCCA GGTCCCC
TP53_chr17:7579741:region_2:23nt : +20:+59:r	198 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGCAGAGACCTGTGGAAAGCGAAAATCCATGGG ACTGACT
TP53_chr17:7579252:region_3:280nt : -59:-20:f	199 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGCAGGGGGATACGGCCAGGCATTGAAGTCTCAT GGAAGCC
TP53_chr17:7579311:region_3:280nt : 1:40:f	200 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCGTGCAAGTCACAGACTTGGCTGTCCCAGAAC GCAAGAA
TP53_chr17:7579351:region_3:280nt : 41:80:r	201 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCAGAAAACCTACCAGGGCAGCTACGGTTCCG TCTGGC
TP53_chr17:7579391:region_3:280nt : 81:120:f	202 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGAAGGGACAGAAAGATGACAGGGGCCAGGAGGG GCTGGTG
TP53_chr17:7579431:region_3:280nt : 121:160:r	203 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTGGCCCTGCAACCAGCAGCTCCTACACCGCG GCCCTG
TP53_chr17:7579471:region_3:280nt : 161:200:f	204 ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGGGGGAGCAGCCTCTGGCATTCTGGGAGCTTC ATCTGGA

TABLE 6-continued

Name_target region	SEQ ID NO:	Probe sequence
TP53_chr17:7579511:region_3:280nt: 201:240:r	205	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCCCGGACGATATTGAACAAATGGTCACTGAAG ACCCAGG
TP53_chr17:7579610:region_3:280nt: +20:+59:r	206	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCTGGGGGCTGGGGGCTGAGGACCTGGTCCTC TGACTGC
TP53_chr17:7578327:region_4:185nt: -43:-4:f	207	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGGCAACCAGCCCTGCGTCTCTCCAGCCC CAGCTGC
TP53_chr17:7578370:region_4:185nt: 1:40:f	208	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCATCGCTATCTGAGCAGCGCTCATGGTGGGG CAGCGCC
TP53_chr17:7578410:region_4:185nt: 41:80:r	209	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGCCATCTACAAGCAGTCACAGCACATGACGGAG GTTGTGA
TP53_chr17:7578450:region_4:185nt: 81:120:f	210	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCATGGCGCGACGCGGGTGCCGGGGGTGT GGAAATCA
TP53_chr17:7578490:region_4:185nt: 121:160:r	211	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTTGCCAACTGGCAAGACCTGCCCTGTGCAGC TGTGGGT
TP53_chr17:7578574:region_4:185nt: +20:+59:r	212	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTGCTTTATCTGTTCACTTGTGCCCTGACTTCA ACTCTGT
TP53_chr17:7578117:region_5:114nt: -59:-20:f	213	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGAGGGCCACTGACAACCACCTTAACCCCTCCT CCCAGAG
TP53_chr17:7578176:region_5:114nt: 1:40:f	214	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTCAGGGGCTCATAGGGCACCAACACTAT GTCGAAA
TP53_chr17:7578216:region_5:114nt: 41:80:r	215	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAGGAAATTGCGTGTGGAGTATTGGATGACAG AAACACT
TP53_chr17:7578292:region_5:114nt: +3:+42:r	216	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCAGGGTCCCGAGGCCTCTGATTCTCACTGAT TGCTCTT
TP53_chr17:7577439:region_6:111nt: -59:-20:f	217	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGAGGCAAGCAGAGGCTGGGCACAGCAGGCCAG TGTGCAG
TP53_chr17:7577498:region_6:111nt: 1:40:f	218	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGGAGTCTCCAGTGTGATGATGGTAGGGAT GGGCCTC
TP53_chr17:7577538:region_6:111nt: 41:80:r	219	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CACTACATGTGTAACAGTCCCTGCATGGCGCA TGAACCG
TP53_chr17:7577628:region_6:111nt: +20:+59:r	220	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCTTGCCACAGGTCTCCCAAGGCACACTGGCT CATCTTG
TP53_chr17:7576974:region_7:138nt: -44:-5:f	221	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCTGCAACCTGGTCTCCTCCACCGCTTCTGTC CTGCTTG
TP53_chr17:7577018:region_7:138nt: 1:40:f	222	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTCGCTTAGTGTCCCTGGGGCAGCTCGTGG TGAGGCT

TABLE 6-continued

Name_target region	SEQ ID NO:	Probe sequence
TP53_chr17:7577058:region_7:138nt: 41:80:r	223	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGACCGGCCACAGAGGAAGAGAATCTCCGAAG AAAGGGG
TP53_chr17:7577098:region_7:138nt: 81:120:f	224	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTCTCCCAGGACAGGCACAAACACGCACCTCAA GCTGTTTC
TP53_chr17:7577138:region_7:138nt: 121:+22:r	225	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTCTCTTTCTATCCTGAGTAGTGTAATCTAC TGGGACG
TP53_chr17:7577175:region_7:138nt: +20:+59:r	226	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGACAGGTAGGACCTGATTTCCTACTGCCTCT TGCTTCT
TP53_chr17:7576793:region_8:75nt: -59:-20:f	227	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGCATTTGAGTGTTAGACTGGAAACTTTCCAC TTGATAA
TP53_chr17:7576852:region_8:75nt: 1:40:f	228	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGAAAGGTGAAATATTCTCATCCAGTGTT TCTTCTT
TP53_chr17:7576892:region_8:75nt: 41:+5:r	229	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTAGCACTGCCAACAACACCAGCTCCTCC CCAGCCA
TP53_chr17:7576931:region_8:75nt: +5:+44:r	230	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTGCCCTCAGATTCACTTTATCACCTTCCTGC CTCTTTC
TP53_chr17:7573867:region_9:108nt: -59:-20:f	231	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CATGGCTTCCAACCTAGGAAGGCAGGGAGTAG GGCCAGG
TP53_chr17:7573926:region_9:108nt: 1:40:f	232	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGGAGTGAGCCCTGCTCCCCCTGGCTCCT CCCAGCC
TP53_chr17:7573966:region_9:108nt: 41:80:r	233	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTCCGAGAGCTGAATGAGGCCTTGGAACTCAAGG ATGCCCA
TP53_chr17:7574053:region_9:108nt: +20:+59:r	234	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCATCTTTAACTCAGGTACTGTGTATATACTTA CTTCTCC
TP53_chr17:7572867:region_10:83nt: -59:-20:f	235	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGCAGGGGAGGGAGAGATGGGGTGGGAGGCTG TCAGTGG
TP53_chr17:7572926:region_10:83nt: 1:40:f	236	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTCAGTCTGAGTCAGGCCCTCTGTCTTGAAACA TGAGTTT
TP53_chr17:7572966:region_10:83nt: 41:80:r	237	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTGAAGTCCAAAAGGGTCAGTCTACCTCCCG CCATAAA
TP53_chr17:7573028:region_10:83nt: +20:+59:r	238	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGCACAGACCCCTCACTCATGTGATGTCATCT CTCCTCC
ALDH4A1_chr1:19199369_rs61757683: G:T:f	239	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCAGGGCTTATGTCTCCTTGATGACCTCGG CGACGTC
ALDH4A1_chr1:19199488_rs61757683: G:T:r	240	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCATCATCTCCTCCCTCCCTCTGCCAGGC TGTTGCA

TABLE 6-continued

Name_target region	SEQ ID NO:	Probe sequence
BRCA1_chr17:41223015_rs1799966: T:A,C:f	241	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAATTCTGGCTTCCTCCCTGCACACTTCTTCC ATTGCAT
BRCA1_chr17:41223134_rs1799966: T:A,C:r	242	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTCAGCTCGTGGCAACATACCATCTTCAAC CTCTGCA
BRCA1_chr17:41243921_rs16942:T: C:f	243	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTAATTCTGGCCCTCTCGTAACCCTGAGC CAAATGT
BRCA1_chr17:41244040_rs16942:T: C:r	244	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGTGAAATAAAGGAAGATACTAGTTTGCTGAA AATGACA
BRCA2_chr13:32906670_rs144848: A:C:f	245	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CACTCATTTGATCTGAAGTGGAACCAAATGATA CTGATCC
BRCA2_chr13:32906769_rs144848: A:C:r	246	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CAGTTGAGACATTCAACAGGCCAAGACGGTACA ACTTCCT
CDKN2A_chr9:21970837_rs3731249: C:T:f	247	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGAAAATGAATGCTCTGAGCTTGGAAAGCTCTCA GGGTACA
CDKN2A_chr9:21970956_rs3731249: C:T:r	248	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGGGCCATCGCGATGTCGACCGTACCTGCCGCG GGCTGCG
DPYD_chr1:97981316_rs1801159:T: C:f	249	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCCCATCCAGCTCAAAGCTCTTCGAATCATT GATGTGC
DPYD_chr1:97981435_rs1801159:T: C:r	250	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTGCCAAGCCTGAACTACCCCTCTTACACTCC TATTGAT
EPHX1_chr1:226026327_rs2234922: A:G:f	251	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCACCTGACTGTGCTCTGCCCCCAGGCTG GACATCC
EPHX1_chr1:226026446_rs2234922: A:G:r	252	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CACTCAGGAGTGGATGATCTTATAAAACTCGTA GAAAGAG
MYC_chr8:128750752_G123E:f	253	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCCTCGGGAGACAACGACGGCGGTGGCGGAGC TTCTCCA
MYC_chr8:128750871_G123E:r	254	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CCATACAGTCTGGATGATGATGTTTGATGAA GGTCTCG
RB1_chr13:49039115_rs121913297: G:T:f	255	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CTTTTACTGTTCTCCTCAGACATTCAAACGTGT TTTGATC
RB1_chr13:49039204_rs121913297: G:T:r	256	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTGGAAAGCATACTGCAAAATATTGTTTCAGT CTCTGCA
TNFRSF14_chr1:2491227_rs2234163: G:A:f	257	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CACGTACCCCTCTCAGCCCCCTCTTGGACTCC AGCCATG
TNFRSF14_chr1:2491346_rs2234163: G:A:r	258	ATGTGACTGGCACGGGAGTTGATCCTGGTTTCA CGTGGCGTAAGCGCGGACGCGGCGCAGTGGTCC CCGTCC

[0243] Genomic samples: Three samples of genomic DNA were analyzed:

[0244] 1) Germline sample NA 06994—a normal human sample obtained from the Coriell repository;

[0245] 2) Cancer cell line NCI-H69—a cell line known to harbor a mutation in TP53, an amplification of the MYCN locus, and SNVs in ALDH4A1, BRCA1, BRCA2, CDKN2A, DPYD, EPHX1, MYC, RB1 and TNFRSF14 that were included in the target probe set;

[0246] 3) Cancer cell line ZR-75-1, which was reported to harbor an EML4-ALK fusion gene (Lin et al., *Mol. Cancer Res.* 7(9):1466, 2009).

[0247] DNA sequencing libraries are generally constructed from sheared DNA fragments. Acoustic disruption was used to generate DNA fragments that ranged in size from 200 to >500 bp. Enzymatic fragmentation of the acoustically fragmented DNA was performed in an effort to emulate circulating DNA, which is reputed to be composed of nucleosomal, ~150 bp fragments. Briefly, DNA at 20-40 ng/μL was sonicated on the 200 bp setting, which yields fragments that range in size from 150 bp to 400 bp in a broad

smear. The DNA was further fragmented by the addition of 0.01 and 0.02 μL of DNase enzyme (New England Biolabs recombinant bovine DNase) to 50 μL aliquots of DNA in DNase buffer (10 mM Tris pH 8.0, 2.5 mM MnCl<sub>2</sub>, 0.5 mM CaCl<sub>2</sub>). The DNase reaction was incubated at 37°C. for 10 min and stopped with the addition of 0.5 M EDTA to a final concentration of 25 mM. DNA with an average size of 150 bp was purified by “two-sided” bead selection by first adding 0.9 volumes of beads to 1 volume of DNA. The beads bind unwanted larger fragments and are discarded, and an additional 1.6 volumes of beads are added to the supernatant. The bound material is then purified and quantified. An agarose gel of the resulting, highly fragmented, short DNA used for library construction is shown in FIG. 6.

[0248] Fragmented DNA was end-repaired using the Quick Blunt kit form NEB and blended in the ratios shown in Table 7. Ten nanograms of blended DNA were then ligated to adaptors with the sequences shown in Table 7. For mixes 9 and 15, two ligation reactions with 10 ng each were performed and subsequently pooled. For mix 16, four reactions were done. An estimate of genomic inputs into each library using a qPCR assay is also shown in Table 7.

TABLE 7

Samples and admixture ratios					
Sample	Admix ratio	Code	SEQ ID NO:	Genomic inputs	
1) NA06994 = GL	pure	NNNNNAAGATCTTAGTGGCAC	259	202	
2) NCI-H69 = N	pure	NNNNNCGACAGAACTATTGCC	260	149	
3) ZR-75-1 = Z	pure	NNNNNACTATCTTAGTGGCAC	261	242	
4) GL:N	1:1	NNNNNCTCCAGAACTATTGCC	262	200	
5) GL:N	2:1	NNNNNAGCATCTTAGTGGCAC	263	83	
6) GL:N	4:1	NNNNNCATCAGAACTATTGCC	264	186	
7) GL:N	10:1	NNNNNATAATCTTAGTGGCAC	265	264	
8) GL:N	20:1	NNNNNAAGAAGGTAGACCCTC	266	203	
9) GL:N	100:1	NNNNNTTCTACTCGTGAC	267	436	
10) GL:Z	1:1	NNNNNACTAAGGTAGACCCTC	268	297	
11) GL:Z	2:1	NNNNNGAAGCTACGAGTATCC	269	224	
12) GL:Z	4:1	NNNNNAGCAAGGTAGACCCTC	270	73	
13) GL:Z	10:1	NNNNNCATTGACGTCTAGAGC	271	181	
14) GL:Z	20:1	NNNNNTCACTCTACTCGTGAC	272	224	
15) GL:Z	100:1	NNNNNATAAAGGTAGACCCTC	273	580	
16) GL:N:Z	500:1:1	NNNNNTACCTACTCGTGAC	274	1324	

[0249] Targeted sequencing: One microgram of each of the sixteen DNA libraries shown in Table 7 were pooled and adjusted to a final volume of 160  $\mu$ L. Eight identical 20  $\mu$ L aliquots were denatured at 98° C., cooled on ice, and 20  $\mu$ L of probes (Table 6) at 1 nM/probe and 50  $\mu$ L of CF hybridization buffer were added. The samples were annealed for 24 hours from 80° C. to 50° C., washed, and amplified. Following amplification of the resulting captured and processed fragments, the final sequencing library was size selected using the Pippin Prep™ instrument with a size selection of 175–400 bp. The library was sequenced on an Illumina MiSeq using a 150 read V3 kit.

### Results

[0250] Capture probe performance of high density capture probes that were chosen based on their position with target sequences were monitored. A graphical display of the performance of each capture probe is shown in FIG. 7. These data demonstrate that:

- [0251] 1) all capture probes chosen strictly by positional constraint provided on-target sequence information;
- [0252] 2) most capture probes exhibit very little off-target and unmappable read capture; and
- [0253] 3) the yield of useful, on-target reads was substantially uniform.

Capture probes that captured a high proportion of off-target and unmappable reads were analyzed further. These capture probes were generally positioned in regions of low sequence complexity/high sequence redundancy. Here, however, such capture probes had no significant detrimental impact on the sequencing depth because the high level of probe redundancy (high density probes) means that all regions are covered by reads derived from several probes. The net effect was excellent uniformity of coverage. See, e.g., FIG. 8, probe coverage for the TP53 gene using the 40 mer capture probes.

### Conclusion

[0254] Taken together, these data demonstrate that capture probe length can be reduced from 60 nucleotides to 40 nucleotides with little or no discernible loss of probe performance (once capture wash temperatures are adjusted). They also show that probe design can follow positional constraints and can generally ignore sequence context or composition. Even though this methodology produces the occasional poor-performing probe, the high redundancy at close probe spacing more than compensates for individual probe deficiencies.

### Example 3: Genetic Analysis of Circulating DNA

#### Purpose

[0255] The purpose of this example was to benchmark the genetic analysis of cfDNA using an efficient cloning procedure for cfDNA and target retrieval system.

#### Background

[0256] While there is tremendous enthusiasm in the scientific and health-care community for “liquid biopsies”—analysis of circulating DNA (cfDNA) for markers associated with potential disease states, there is remarkably little practical information about this potential analyte.

### Summary

[0257] Plasma samples collected from healthy donors and individuals suffering from either ovarian or colon cancers were used to perform the genetic analysis of circulating DNA. The amount and the overall character of circulating cfDNA can vary widely from individual to individual. Surprisingly, the present inventors found that cfDNA is readily clonable with an efficiency indistinguishable from highly purified and fragmented genomic DNA; that the fragment size was remarkably consistent, with an average clone insert size of 170 $\pm$ 10 bp (in 7% samples); and that the genome representation from such samples was uniform and comparable to experiments performed using purified gDNA. It was further established that by counting unique reads, the depth of representation in each library provided an estimate of minor allele frequency for tumor markers present in the cfDNA of diseased patients. This study established that construction and target retrieval systems contemplated herein were effectively applied to the quantitative genetic analysis of cfDNA.

### Methods

[0258] DNA purification: Eight sets of plasma samples were purchased from Proteogenex, Inc., Culver City, CA (Table 8). Circulating DNA was extracted from the samples (on two separate occasions) using the Circulating Nucleic Acid Purification kit from Qiagen. Samples were passed through DNA mini-columns using centrifugation. The specimen IDs and yield of DNA are shown in Table 8.

TABLE 8

Plasma samples and cfDNA yields				
Sample ID	Patient diagnosis	Specimen type	DNA yield (ng per mL of input)	mL of Volume input
D5930P	Healthy donor	plasma	4 mL	11
D5942P	Healthy donor	plasma	4 mL	68
023407P	Colorectal cancer	plasma	4 mL	10
023406P	Colorectal cancer	plasma	4 mL	63
023185P	Colorectal cancer	plasma	4 mL	171
023149P	Colorectal cancer	plasma	4 mL	36
032667P	Ovarian cancer	plasma	4 mL	24
032676P	Ovarian cancer	plasma	4 mL	13

[0259] Library construction: Purified DNA from 4 mL of plasma was collected in 100  $\mu$ L elution buffer. For the four samples collected from colon cancer patients (CRC), the DNA was split in half and one 50  $\mu$ L aliquot from each patient was sonicated to 200 bp. One 50  $\mu$ L aliquot of untreated cfDNA and one 50  $\mu$ L fragmented cfDNA from each patient (the entire sample from each patient) was end repaired by adding (per sample):

- [0260] 6  $\mu$ L of 10x quick blunt buffer (New England Biolabs (NEB))
- [0261] 0.6  $\mu$ L of 10 mM dNTPs
- [0262] 2.4  $\mu$ L of quick blunt enzyme mix
- [0263] 1.2  $\mu$ L of PreCR enzyme mix.
- [0264] Samples were incubated at 20° C. for 30 min and at 70° C. for 10 min. Ligations with adaptors (Table 2) were performed by combining:

- [0265] 60  $\mu$ L end-repaired cfDNA
- [0266] 12  $\mu$ L adaptor duplex (10  $\mu$ M)
- [0267] 10  $\mu$ L 10x ligase buffer (NEB)
- [0268] 15  $\mu$ L 50% PEG<sub>8000</sub>
- [0269] 3  $\mu$ L HC T4 DNA ligase

TABLE 9

<u>Samples and codes used for four CRC plasma samples</u>				
Progenex ID	Sample #	Pre-treatment	Adapter	SEQ ID NO:
23149	1	none	NNNNNTTTGTGTGTGTG	275
23407	2	none	NNNNNACTACACACACAC	276
23406	3	none	NNNNNCTCGTGTGTGTGT	277
23185	4	none	NNNNNGAACACACACACACA	278
23149	5	frag	NNNNNCATGTGTGTGTGT	279
23407	6	frag	NNNNNNGTCACACACACACA	280
23406	7	frag	NNNNNATAACACACACACAC	281
23185	8	frag	NNNNNTACTGTGTGTGTG	282

[0270] Reactions were incubated at 22° C. for one hour and 65° C. for 10 min. Ligation products were purified by the addition of 100  $\mu$ L beads, washing, and elution in 40  $\mu$ L TEzero. All 40  $\mu$ L of ligation product was amplified by PCR with primer ACA2 (SEQ ID NO: 283) and the samples were combined in equal mass for targeted capture.

[0271] Targeted sequence capture and sequencing: The four unfragmented and four fragmented colon plasma samples (FIG. 9C) were hybridized with our high-density, 40 nucleotide probe set that targets TP53, ALK, among others. The capture complexes were processed as described above in Example 2.

## Results

[0272] Library appearance: A false-color picture of a 2% agarose gel loaded with 50 ng of each library is shown in FIG. 9A. The average fragment size was in a tight range of 260 $\pm$ 20 bp. These data indicated that the clonable fraction of cfDNA is present predominantly as nucleosomal fragments. In addition, the size of the cfDNA libraries had the same basic superficial appearance as cfDNA in kidney dialysis patients (Atamaniuk et al., *Clinical Chemistry* 52(3):pp. 523-26 (2006)) except that the cfDNA libraries were shifted to higher mass by the addition of adaptor sequences (FIG. 9B). In contrast, the cfDNA libraries differed dramatically from sonicated gDNA libraries, which appear as broad smears.

[0273] Four additional sets of cfDNA libraries were constructed from the two ovarian cancer patient plasma samples and two plasma samples from healthy volunteers. 38  $\mu$ L aliquots of cfDNA were end-repaired in 50  $\mu$ L total volume. Ligations included 40  $\mu$ L of end-repaired fragment, 16  $\mu$ L of adaptor (10  $\mu$ M), 8  $\mu$ L of 10x ligase buffer, 16  $\mu$ L of 50% PEG and 4  $\mu$ L of HC T4 DNA ligase in a total volume of 80  $\mu$ L. The ligation reaction was incubated at 20° C. for 1 hour and 65° C. for 10 min. For purification, 20  $\mu$ L of TEzero and 150  $\mu$ L of beads were added. The purified ligation products were resuspended in 40  $\mu$ L, all of which was used in a

subsequent 200  $\mu$ L library amplification by PCR. The resulting amplified libraries are shown in FIG. 9C.

[0274] Sequencing data analysis: The average unique read count observed in each of the eight libraries ranges from ~700 unique reads to >3000 unique reads, defining a range of sensitivities from ~0.15% to ~0.03%. FIG. 10. A rare mutant read will likely be observed more than once, meaning minimum sensitivities are less than those calculated above. In preferred embodiments, unique reads provide the lower bound on statistically defensible observation frequencies.

[0275] cfDNA cloning efficiency: Sample 23407 was used as a benchmark. 10 ng/mL of cfDNA was recovered from the plasma sample and 20 ng of the isolated cfDNA was used in each of two library construction efforts. The unique read counts indicated that we recovered an average of 700 unique reads (genome equivalents) from unfragmented DNA ("23407" in FIG. 10). Given that each genome contains 0.003 ng of gDNA, 2.1 ng of input DNA in this library (10% cloning efficiency) was recovered.

[0276] Fragmentation prior to library construction with this sample increased the library yield by more than two-fold ("23407 frag" in FIG. 10). This indicates that much of the DNA present in the 23407 sample was high molecular weight DNA that required fragmentation in order to be clonable. Thus, the library cloning efficiency was likely far greater than 10% and was likely in the range of 20% for input cfDNA. This cloning efficiency is comparable to highly purified genomic DNA and indicates that cfDNA was not likely modified in any way that is deleterious to downstream cloning efforts.

[0277] Library coverage: The cfDNA libraries resembled a set of discrete bands with random coverage of target regions. FIG. 11 shows a random sampling of sequence data. A random set of reads from sample 23407 that was not fragmented prior to cloning (see FIG. 10), and that were captured by the TP53 probe "chr17:7579351:region\_3: 280nt:41:80:r" (SEQ ID NO: 201) were aligned using BLAT. Given the way that the sample was prepared, these are likely a reflection of cfDNA fragments in general because the left hand portion of these reads (the read start sites) are randomly distributed across the target region. This random distribution indicates the random breakage of genomic DNA, and it demonstrates that despite the band-like appearance of cfDNA libraries, the sequencing output was a random coverage of the target region. The random distribution is important for effective genetic analysis using technology contemplated herein.

[0278] FIG. 12 provides a more high resolution overview of TP53 coding region sequencing for a typical cfDNA library. The elements of targeted sequencing—coverage across all target regions and uniform depth at each sequenced base—are readily apparent. At this depth of >4000 unique reads per base, and with a requirement that legitimate candidate base changes must be encountered at least twice, it is possible to estimate that the mutation detection sensitivity for this particular library was about 1 mutation in 2000 sequences, or 0.05%. This level of sensitivity represents a surprising and unexpected outstanding technical achievement.

## Conclusion

[0279] cfDNA was isolated and cloned from plasma clones with an efficiency comparable to highly purified

gDNA isolated from cell lines (the gold standard). The cfDNA libraries resembled circulating nucleosomal-sized DNA fragments+adaptors and the ends possessed sufficiently random character, which enabled efficient genetic analysis. In addition, the highly uniform size characteristic of plasma cfDNA libraries allows designing capture strategies and underlying probe sequences to maximize reliable coverage of targets as far as 120 bp (=160–40) from the ends of probes.

**Example 4: Measurement of Genome Equivalents in Circulating DNA Libraries**

**Purpose and Background**

[0280] One of the major challenges in the analysis of circulating, cell-free DNA is achieving sufficient assay sensitivity. If sufficient sensitivity is not achieved then analysis of the cfDNA libraries is confounded: if a sample is sequenced and no mutational events are detected, that result could be interpreted to mean that no mutations are present, or that significant events were missed because the sampling depth was too small. The sensitivity of an assay is defined in statistical terms as the false negative rate. In the context of sequencing circulating, cell-free DNA, a significant obstacle is the detection of a rare sequence that is blended in a large excess of reference sequence.

[0281] One method for determining assay sensitivity is to measure the occurrences of mutant sequence in a set of samples where mutant sequence is progressively diluted into non-mutant, reference sequence. The dilution at which mutant sequences are no longer detected defines assay sensitivity. This method is adequate if both the identity of the mutant sequence and the extent of dilution are known. Unfortunately, clinical samples do not generally provide either parameter. Often the identity of the mutant sequence is not known, and the extent of dilution varies from sample to sample. In this context, assay sensitivity is established on a sample-by-sample basis.

[0282] To assign a sensitivity value on a sample-by-sample basis, the numbers of different and distinct sequences that are present in each sample are measured, by measuring the number of genome equivalents that are present in a sequencing library. By way of a non-limiting example, if a DNA sequencing library is known to contain 3 ng (3000 pg) of human genomic DNA and each human genome has a mass of 3 pg, then the library possesses  $3000 \div 3 = 1000$  genome equivalents of DNA. If a mutant DNA sequence must be detected twice to be statistically significant, then an estimate of the best possible sensitivity of detection for this particular library is  $2 \text{ mutant sequences} \div 1000 \text{ total sequences} = 0.002 = 0.2\%$ . To establish sensitivity, the number of genome equivalents must be measured for each sample library.

**Summary**

[0283] Two methods were used to measure genome equivalents. The first method is based on quantitative PCR (qPCR). A genomic library was constructed using ligation of adaptors to genomic fragments and a pair of PCR primers, one that is specific to a common genomic sequence (e.g., Alu I repeat) and one that is specific to the adaptor. The abundance of ligated adaptor: fragment sequences of these cfDNA libraries was measured. A standard library of known

concentration was used to construct a standard curve and the measurements were fit to the resulting standard curve and a value for genome equivalents was derived from the fit.

[0284] The second method to measure genome equivalents used bioinformatics counting after sequencing was performed. Each unique sequence in a library was identified by its random sequence label and the starting nucleotide of the genomic sequence. Moreover, each unique sequence must be derived from an independent genome. Therefore, the sum of unique sequences present in sequence data established a precise quantitative measurement of the number of genome equivalents present in a sample.

**Methods and Results**

[0285] qPCR assay development: The first version of a qPCR-based genome equivalence assay used the ACA2 primer (Table 10), but this assay chronically under-reports the number of genome equivalents that are present in a cfDNA library (FIG. 13).

TABLE 10

PCR primers used in the development of the genome equivalent qPCR assay		
Name	SEQ ID NO:	Sequence
ACA2	283	TGCAGGACCAGAGAATTCTGAATACA
ACA2_FLFP	284	AATGATAACGGCGACCACCGAGATCTACACGTCTACAGGACCAGAGAATTCTGAATACA
Alu_F1	285	CGGTGGCTCACGCCGTGA
Alu_R1	286	GCCTCGGCTCCCAAAGT
Alu_F2	287	GAGGCTGAGGCAGGAGAACG
Alu_R2	288	GTCGCCAGGCTGGAGTG

[0286] The improved version of the assay was based on endogenous repeats (e.g., Alu repeats) that are found at very high frequency throughout the human genome. By coupling an Alu-specific primer with an adaptor-specific primer, the frequency with which adaptors are joined to genomic fragments was reliably measured. Standard curves using libraries of known genome equivalents were generated, and the number of genome equivalents in cloned libraries was measured by fitting to the curve.

[0287] The PCR primers used to develop an Alu+adaptor-based qPCR assay are shown in Table 10. The PCR primers for Alu amplification were designed from consensus a consensus human Alu sequence (Batzer & Deininger, *Nat Rev Genet.* 3(5):370-9 (2002)) using PRIMER3 (Alu\_F1 & Alu\_R1, SEQ ID NOS: 285 and 286, respectively). The remaining two Alu primers (Alu\_F2 and Alu\_R2, SEQ ID NOS: 287 and 288, respectively) were reported in the literature (Marullo et al., *Genome Biology* 11:R9 (2010)).

[0288] A schematic of the assay design is provided in FIG. 14. Because a single PCR primer can be used to amplify the genomic DNA libraries (FIG. 14A), a primer that recognizes the adaptor sequence but that cannot amplify genomic clones was used. The 58 nucleotide ACA2-FLFP primer (henceforth abbreviated AF, SEQ ID NO: 284) fills these criteria because its length induces strong stem-loop PCR suppression (FIG. 14B). Additionally, a functional pair of

Alu primers were used (FIG. 14C). Moreover, a primer pair consisting of one Alu primer and the long ACA2 primer that did not amplify genomic DNA used (FIG. 14D). These same primers also amplified genomic library clones (FIG. 14E). [0289] All of the required elements for a functional Alu-based assay were validated. FIG. 15. Specifically, the long primer alone was inert, both sets of Alu primer pairs recognized human genomic DNA, and any combination of one Alu primer and the long ACA2 primer amplified

genome equivalents in cfDNA libraries is shown in FIG. 17. There is a substantial amount of molecular engineering within the 45 nucleotide ligation strand, which is the strand that becomes attached to end repaired cfDNA fragments. The adaptors comprise at least five elements.

[0294] Element 1 is a PCR primer binding site for the single-primer library amplification primer ACA2 (Table 12).

TABLE 12

Element 1 Summary			
Element	Function	Number of sequences	SEQ ID NO:
Element 1 PCR primer binding site		1	TGCAGGACCAGAGAATTCTGAATACA 289

genomic library clones (FIG. 15A). Finally, the ability of Alu primer plus long ACA2 primer pair to discriminate between genomic DNA and genomic library clones is shown in FIG. 15B. The combination of Alu\_R1 and AF primers were used for measuring genome equivalents in the newly constructed libraries.

[0290] A direct comparison between the ACA2-based and the Alu-based qPCR assays is shown in FIG. 16. An 8-fold difference in genome equivalents was found. In addition the Alu-based assays provided a more consistent performance library-to-library and a better alignment between qPCR derived equivalents and bioinformatically counted tag equivalents in sequencing runs (Table 11).

TABLE 11

qPCR vs counted sequencing tags		
Sample	Alu-based qPCR	counted tags
Run_68_50tol	6962	3459
Run_68_1000tol	10937	4641

[0291] High-sensitivity library adaptors for sequence-based countingg of enome equivalents: As discussed above, the reality of disease surveillance using cfDNA is that mutant sequences may be rare constituents in an otherwise vast excess of “normal” (meaning germline) DNA sequences. Thus, highly sensitive and quantifiable sequencing assays are needed. Assay sensitivity could be made by counting the number of unique sequences present in a sequencing library. However, such counting would lead to a false underestimate of sensitivity because cfDNA fragments are rather short (~165 bp) and may lead to identical reads that were actually derived from independent cloning events. One solution to this problem is to mark each independent sequencing clone during library construction by including, for example, a set of DNA tags in the adaptors used to construct libraries.

[0292] A set of such library construction adaptors was specifically designed to measure the number of genome equivalents present in cfDNA libraries, and, by extension, the sensitivity of sequencing assays used to monitor mutant sequences.

[0293] The architecture of high-sensitivity library adaptors that were configured to accommodate large numbers of

TABLE 13

Element 2 Summary		
Element	Function	Number of sequences
Element 2	Distinct sequence labels	249
	Sequences (5' -> 3')	SEQ ID NO:
	CGGGT	290
	CGGTG	291
	CGTGG	292
	GCGGT	293
	GCGTG	294
	GCTGG	295
	GGCGT	296
	GGCTG	297
	GGGCT	298
	TTAAA	299
	TTACC	300

TABLE 13-continued

Element 2 Summary	
TTATT	301
TTCAC	302
TTCCA	303
TTTAT	304
TTTTA	305
GCACG	306
GCAGC	307
GCCAG	308
GCCGA	309
GCGAC	310
GCGCA	311
GGAAA	312
GGACC	313
GGATT	314
GGCAC	315
GGCCA	316
GGTAT	317
GGTTA	318
GTAGT	319
GTATG	320
GTGAT	321
CCGTC	322
CCTCG	323
CCTGC	324
CGAAT	325
CGATA	326
CGCCT	327
CGCTC	328
CGTAA	329
CGTCC	330
CGTTT	331
CTAAG	332
CTAGA	333
CTCCG	334
CTCGC	335
CTGAA	336
CTGCC	337

TABLE 13-continued

Element 2 Summary	
TTGTC	338
TTTCG	339
TTTGC	340
AAAAA	341
AAACC	342
AAATT	343
AACAC	344
AACCA	345
AATAT	346
AATTA	347
ACAAC	348
ACACA	349
ACCAA	350
ACCCC	351
ACCTT	352
ACTCT	353
GGGTC	354
GGTCG	355
GGTGC	356
GTCGG	357
GTGCC	358
GTGGC	359
TGCGG	360
TGGCG	361
TGGGC	362
AAAAG	363
AAGAG	364
AAGGA	365
ACCGG	366
ACCGC	367
ACGGC	368
AGAAG	369
GTGTA	370
GTTAG	371
GTTGA	372
TAGGT	373
TAGTG	374
TATGG	375

TABLE 13-continued

Element 2 Summary	
TGAGT	376
TGATG	377
TGGAT	378
TGGTA	379
TGTAG	380
TGTGA	381
TTAGG	382
TTGAG	383
TTGGA	384
AACGT	385
CTGTT	386
CTTGT	387
CTTTG	388
GAACT	389
GAATC	390
GACAT	391
GACTA	392
GATAc	393
GATCA	394
GCAAT	395
GCATA	396
GCCCT	397
GCCTC	398
GCTAA	399
GCTCC	400
GCTTT	401
ACTTC	402
ATAAT	403
ATATA	404
ATCCT	405
ATCTC	406
ATTAA	407
ATTCC	408
ATTTT	409
CAAAC	410
CAACA	411
CACAA	412

TABLE 13-continued

Element 2 Summary	
CACCC	413
CACTT	414
CATCT	415
CATTC	416
CCAAA	417
AGAGA	418
AGCCG	419
AGCGC	420
AGGAA	421
AGGCC	422
AGGTT	423
AGTGT	424
AGTTG	425
ATGGT	426
ATGTG	427
ATTGG	428
CACGG	429
CAGCG	430
CAGGC	431
CCAGG	432
CCGAG	433
AACTG	434
AAGCT	435
AAGTC	436
AATCG	437
AATGC	438
ACAGT	439
ACATG	440
ACGAT	441
ACGTA	442
ACTAG	443
ACTGA	444
AGACT	445
AGATC	446
AGCAT	447
AGCTA	448
AGTAC	449
GTAAC	450

TABLE 13-continued

Element 2 Summary	
GTACA	451
GTCAA	452
GTCCC	453
GTCTT	454
GTTCT	455
GTTTC	456
TAACG	457
TAAGC	458
TACAG	459
TACGA	460
TAGAC	461
TAGCA	462
TCAAG	463
TCAGA	464
TCCCC	465
CCACC	466
CCATT	467
CCCAC	468
CCCCA	469
CCTAT	470
CCTTA	471
CTACT	472
CTATC	473
CTCAT	474
CTCTA	475
CTTAC	476
CTTCA	477
TAAAT	478
TAATA	479
TACCT	480
TACTC	481
CCGGA	482
CGACG	483
CGAGC	484
CGCAG	485
CGCGA	486
CGGAC	487

TABLE 13-continued

Element 2 Summary	
CGGCA	488
GAAAG	489
GAAGA	490
GACCG	491
GACGC	492
GAGAA	493
GAGCC	494
GAGTT	495
GATGT	496
GATTG	497
AGTCA	498
ATACG	499
ATAGC	500
ATCAG	501
ATCGA	502
ATGAC	503
ATGCA	504
CAAAT	505
CAATG	506
CAGAT	507
CAGTA	508
CATAG	509
CATGA	510
CCCGT	511
CCCTG	512
CCGCT	513
TCCGC	514
TCGAA	515
TCGCC	516
TCGTT	517
TCTGT	518
TCTTG	519
TGAAC	520
TGACA	521
TGCAA	522
TGCC	523
TGCTT	524
TGTCT	525

TABLE 13-continued

Element 2 Summary	
TGTTC	526
TTCGT	527
TTCTG	528
TTGCT	529
TATAA	530
TATCC	531
TATTT	532
TCACT	533

TABLE 13-continued

Element 2 Summary	
TCATC	534
TCCAT	535
TCCTA	536
TCTAC	537
TCTCA	538

[0296] Element 3 is a 3 nucleotide sample code that differ by at least two base changes. This element was used to identify different samples and enabled sample multiplexing within a sequencing run. Table 14.

TABLE 14

Element 3 Summary		Number of sequences (5' -> 3')	SEQ ID NO:
Element	Function		
Element 3	Distinct sample labels; sample multiplexing	16	AAG 539
		CTC	540
		GGT	541
		TCA	542
		ACT	543
		CGA	544
		GTG	545
		TAC	546
		AGC	547
		CCG	548
		GAA	549
		TTT	550
		ATA	551
		CAT	552
		GCC	553
		TGG	554

**[0297]** Element 4 is a 12 nucleotide anchor sequence with three important characteristics with respect to library construction and downstream sequencing. Table 15. These are A) each 12 base extension is part of a family of four 12 base extensions that collectively represent each of the four possible DNA bases at each site within extension. This feature, balanced base representation, is required by the Illumina sequencing instrument in order to calibrate proper base calling in sequencing reads. B) Each extension is composed of only two of four possible bases, and these are specifically chosen to be either 6 A's+6 C's or 6 G's+6 T's. This extension formed from only two bases greatly reduces the possibility that the extension sequence will participate in secondary structure formation that would preclude proper adaptor function. C) Because each extension is composed of equal numbers of A+C or G+T, each extension shares roughly the same melting temperature and duplex stability as every other extension in a set of four.

TABLE 15

Element 4 Summary				
Element	Function	Number of sequences	Sequences (5' -> 3')	SEQ ID NO:
Element 4	12 nucleotide extension that provides a duplexing site for the partner oligonucleotide	4	ACCCACACCAAA	555
			CAAACACAAACCC	556
			GTGTGGGTTGTT	557
			TGTGTTGGTGG	558

**[0298]** Element 5 is the two base sequence found at the 3' end of Element 4. The particular two base extensions were chosen based on empirical data that shows that these two base sequences are efficient substrates for ligation. Table 15.

**[0299]** The adaptor module is hybridized to a partner oligonucleotide. Table 16. The hybridization takes place between the sequence within Element 4 and the partner oligonucleotide. The double-stranded adaptor was ligated to end-repaired cfDNA.

TABLE 16

Element 4 Summary				
Element	Function	Number of Sequences	SEQ ID NO:	
Element 4	12 nucleotide extension that provides a duplexing site for the partner oligonucleotide	4	TTGGTGTGGT	559
			GGGTTGTGTTG	560
			AACAACCCACAC	561
			CCACCAAAACACA	562

**[0300]** To convert a set of 256 independently synthesized and pooled ligation strands (each of which shares a common sample code and therefore constitutes a single sample adaptor set) to duplexes suitable for ligation, the 45 nucleotide ligation strand was combined with the appropriate complementary 12 nucleotide partner strand, heated to 95° C., cooled to 65° C. for 5 min, then cooled to room temperature.

This duplex formed a blunt end ligation substrate as shown in FIG. 17B. Following ligation and DNA purification, a DNA polymerase-mediated step that occurs prior to PCR amplification displaced the partner strand and copied the ligation strand to form a double-strand adaptor that was suitable for exponential amplification by single-primer PCR.

**[0301]** The quantitative analysis of genome equivalents derived from targeted sequencing data was then performed. Each unique read was considered a unique ligation event and the sum of unique reads was considered equivalent to the number of genome equivalents analyzed.

**[0302]** A rough, “back-of-the-envelope”, “rule-of-thumb” calculation was performed to determine the number of genome equivalents that could be analyzed. Each cfDNA clone was approximately 150 base pairs, 50 base pairs of which were required for binding to capture probes. This left roughly 100 possible sequence start sites within any captured cfDNA clone. The attachment of 249 random codes to each of the 100 possible start sites created a total repertoire of ~249,000 possible unique clones. As the number of unique clones approaches the total number of possible sequence combinations, probability dictates that the same code and start site combinations will be created by independent events and that these independent events will be inappropriately grouped within single families. The net result will be an underestimate of genome equivalents analyzed, and rare mutant reads may be discarded as sequencing errors because they overlap with wild-type reads bearing the same identifiers. To avoid this, efforts were made using the qPCR assay to constrain genomic inputs to one tenth or less the number of possible unique clones. For example, a single adaptor has 24,900 possible clones and thus, has a reliable capacity to provide accurate analysis for libraries consisting of 2500 or fewer genome equivalents.

**[0303]** The procedure that is outlined is provided as an example and the methods contemplated herein are not meant to be bound by this example. In some cases, the number of genome equivalents to be analyzed may well exceed the 2500 limit illustrated in the preceding paragraph. To expand the depth of genome equivalents, two solutions to this problem are readily available. The first solution is to use more than one adaptor set per sample. By combining adaptors, it is possible to expand the total number of possible clones and therefore, expand the comfortable limits of genomic input. As a non-limiting example, the combination of four adaptor sets used for one sample would expand the analysis to  $24,900 \times 4 = 99,600$  possible sequences and ~10,000 reasonably analyzed genome equivalents. The second solution is to expand the code in Element 2 of FIG. 17A to 6, 7, or more bases. The number of possible codes that differ by at least 2 bases from every other code scales as  $4^{(n-1)}$  where n is the number of bases within Element 2. Thus, in the non-limiting example presented here, n=5 and  $4^{(5-1)} = 256$ ; therefore, the inclusion of additional bases expands the available repertoire by a factor of four for each additional base.

### Conclusion

**[0304]** The results from this example showed that two independent methods for the determination of genome equivalents have utility in sample processing workflow. The first method, qPCR, was implemented during the library construction phase of cfDNA analysis and it was used as a quality control step to ensure that adequate numbers of genome equivalents are moved through library amplification, targeted sequence capture, and DNA sequencing. The other method uses explicit counting of unique reads as a more direct measure of the actual number of genome equivalents that fell under informatics consideration.

### Example 5: Quantitative Genetic Analysis

#### Purpose

**[0305]** The purpose of this example was to apply quantitative genetic analysis to normal DNA admixed cancer genomes and to uncharacterized cfDNA isolated from the plasma of cancer patients.

#### Background

**[0306]** Three types of genomic events are prevalent in human cancers. These are somatic mutations that alter the function of the affected gene and its expressed protein product(s); genomic rearrangements that create chimeric gene fusions and therefore expressed fusion proteins with novel biological properties; and changes in gene copy number that lead to gene loss and under expression of gene product(s), or, conversely, amplification of genes and over-representation of the corresponding gene product(s). In the circulating DNA of a cancer patient, these aberrant loci, many of which have critical significance in guiding patient care, are admixed (blended) with the patient’s normal, germline DNA.

#### Summary

**[0307]** In the previous examples, technology has been described that was configured for the analysis of circulating, cell-free DNA (cfDNA), with an aim toward cancer surveillance. However, the technology is widely applicable to any analytical, diagnostic and monitoring paradigm including, but not limited to r genetic diseases; fetal testing; mendelian disorders; pathogen screening; and organ transplant monitoring in which circulating DNA is a potential analyte. In this example, the technical features highlighted in previous examples are applied to the analysis of admixed cancer samples. In the first phase of this validation, cancer-derived cell lines were admixed with normal human DNA at defined dilutions, and quantitative genetic analysis was performed. In the second phase of this study, uncharacterized cfDNA was isolated from the plasma of cancer patients and subsequently examined using quantitative genetic analysis.

#### Methods

**[0308]** Admixtures of cell line genomic DNA with normal human DNA: The following DNA samples were used:

**[0309]** NA06994—normal human genomic DNA (Coriell repository);

**[0310]** NCI-H2228—non-small cell lung cancer cell line (ATCC), harbors mutation in TP53 (Q331\*) and EML4-ALK gene fusion (breakpoint unknown); and

**[0311]** NCI-H69—small cell lung cancer cell line (ATCC), harbors amplification of the MYCN gene (~100 copies).

**[0312]** Library preparation: Genomic DNA isolated from cell lines (all three above) is high molecular weight material that is dissimilar to the small size of cfDNA. To mimic cfDNA in these validation experiments, genomic DNAs were first fragmented on the “150 bp” setting using a Covaris Acoustic Sonicator. The sonication generally produces a broad smear, and the DNA was further processed using “two sided” bead selection. A dilute solution of DNA purification beads were added to the sample and the higher molecular mass fragments that adhere to the beads were discarded (the size of purified DNA is proportional to the

amount of beads added). An additional aliquot of beads are added to the remaining supernatant and in this second round, DNA that adheres to the beads that (in a higher overall concentration of binding buffer) are purified. This “two-sided” purification produces a narrow size distribution that is a reasonable proxy for cfDNA (FIG. 18).

[0313] Fragmented genomic DNA was end repaired, quantified, and mixed in the various ratios shown in Table 17 and described in the results section below.

TABLE 17

<u>Samples and admixtures used for validation studies</u>	
Sample, Admixture	genome equivalents (qPCR)
Pure H2228	2248
NA06994:H2228 4:1	2616
NA06994:H2228 10:1	2600
NA06994:H2228 20:1	2968
NA06994:H2228 50:1	5000
NA06994:H2228 1000:1	10000
Pure H69	2472
NA06994:H69 4:1	2768
NA06994:H69 10:1	3088
NA06994:H69 20:1	2944
NA06994:H69 40:1	1616
NA06994:H69 100:1	1920
NA06994:H69 200:1	2920
NA06994:H69 500:1	17520

[0314] cfDNA libraries may have limited DNA inputs. The amount of cfDNA obtained per mL of patient plasma is widely variable, but the lower limits (e.g., Example 3) are generally ~10 ng/mL, which is equivalent to 3300 human genomes. To guard against limited cfDNA quantities, the admixture experiments were modeled to reflect the lower limits of cfDNA that were routinely collected from patients. This constraint was applied to all but the most extreme admixtures. In these latter admixtures, libraries were made to mimic inputs from 4 mLs (NA06994:H2228 1000:1) or 8 mLs (NA06994:H69 500:1) of low yield patient cfDNA. Admixed samples were then ligated to the adaptor sets described in Example 4. Measurement of the genome equivalents in each purified library using qPCR (Example 4) is also shown in Table 17. Libraries were amplified, quantified, and equivalent masses of each library were pooled (500 ng of each). The pooled sample was hybridized with the proof-of-concept, high-density 40 mer capture probes listed in Table 6 of Example 2. The resulting complexes were captured on streptavidin-coated beads, washed, processed, amplified, purified and size-selected as described in previous examples. The resulting library was analyzed using an Illumina 150 bp-V3 MiSeq sequencing kit on the Illumina MiSeq instrument.

[0315] For bioinformatics analysis, a rare somatic variant caller was used to detect mutations, a split read aligner was used to detect fusion genes, and in-house analysis that quantifies and statistically fits tags was used to call copy number variation (CNV).

[0316] The detection of an admixed point mutation in the TP53 gene is shown in FIG. 19. The “expected” frequency deviates from the admix ratio because it is known that TP53 is hemizygous in the NCI-H2228 cell line. Automated software was able to call the mutant allele in the 50:1 admixture. Manual curation was required to call the mutant event at 1000:1. With respect to specificity, the tag filtering

described in Example 1 was applied to the analysis, and no other mutation calls in TP53 were detected after applying this tag filter.

[0317] Cell line NCI-H2228 is known to harbor a fusion gene between EML4 and ALK; the cell line serves as a positive control in both fluorescence in situ hybridization assays and in detection of fusion gene transcripts using RT-PCR. There are no published reports of the exact location of the gene fusion junction. Using dense probe coverage of the intron 19 region of ALK, sequence analysis revealed precise location and sequence of the junction formed when the two genes fused (FIG. 20). The frequency of normal reads versus junction reads in the NCI-H2228 cell line (378 vs 249, respectively) suggests that the fusion gene is heterozygous with a normal copy of ALK.

[0318] Detection of junction reads as a function of admixture is shown in FIG. 21. As with point mutation detection, the expected values were adjusted to reflect the fact that the mutant allele is found in one copy per diploid genome. No fusion reads were detected in the 1000:1 admixed sample.

[0319] FIG. 22 shows the results of CNV determination for the MYCN gene as a function of admixture. The NCI-H69 cell line harbors a highly amplified MYCN gene. MYCN is normally found as a single copy gene, two per diploid genome, and thus the expected result for progressively more dilute admixtures is that the tag-calculated CNV should asymptotically approach 2 copies (the asymptote is highlighted in the figure). The validation experiment shown here indicated that the assay system described in this invention is robustly sensitive to highly amplified genes.

[0320] Variant discovery in cfDNA from cancer patients: The most rigorous validation of the technology contemplated herein is to apply it to cfDNA samples in which the spectrum of mutations is unknown. An analysis was performed by sequencing matched cfDNA, tumor and normal adjacent tissue (NAT) samples from two ovarian cancer patients. In addition, two cfDNA samples from colorectal cancer (CRC) patients and two cfDNA samples from healthy volunteers were analyzed. In no case, were mutations, fusions or abnormal CNV detected in the healthy volunteer samples.

[0321] Libraries of cfDNA from the four cancer patients were initially screened using the targeted probes described in Table 6 of Example 2. These probes were primarily configured to detect point mutations in the TP53 gene, gene fusions with ALK and amplifications of the MYCN. The results of this initial sequencing screen are shown in FIG. 23. A point mutation occurring at the same base position was found in the cfDNA, tumor, and NAT of one ovarian patient. No TP53 mutations were found in the other set of ovarian cancer patient matched samples. Point mutations were also detected in the two CRC cfDNA libraries for which matching tissues were unavailable. All of these point mutations have been previously identified in tumors and all are known to be causal drivers of tumorigenesis. The mutant sequence detection in cfDNA library CRC406 of 0.9% was well within the scope of assay sensitivity. Sensitivity is defined by the presence of multiple families of tagged reads, all of which possess the mutant sequence. These data highlight the clinical utility of the system contemplated herein.

[0322] To further explore the detection of cancer-related changes in cfDNA libraries and associated tissues, the same libraries were hybridized to a set of 679 probes that are directed to a total of 20 different cancer-related genes (Table

18). In this probe set, all of the coding regions of 14 genes were targeted while select loci were targeted in the remaining 6 genes.

TABLE 18

Cancer genes targeted	
Genes: coding region targeted	Genes: select regions targeted
BRCA1	ALK
BRCA2	DPYD
BRAF	EPHX1
CDH1	MYC
ERBB2	TNFRSF14
JAK2	ALDH4A1
NF2	
PIK3CA	
RB1	
CDKN2A	
KRAS	
MYCN	
PTEN	
TP53	

[0323] As shown in FIG. 24, the OVA1 sample, which lacked any detectable changes in TP53, carried a mutation in KRAS that was found in both cfDNA and in the corresponding tumor. This observation highlights a significant feature of the assay system described here. Libraries created from cfDNA can be interrogated with hundreds (as in this example), and even thousands, of targeting probes. The sequencing of the resulting targeted libraries revealed somatic mutations that reside within the tumor and not within the germline of the affected individual. These tumor-associated somatic markers can also be used to quantify the amount of circulating DNA that is shed from the tumor (as opposed to cfDNA that has germline sequence). Thus, the discovery of mutations, regardless of their biological significance, also provides an estimate of tumor content in admixed cfDNA.

[0324] Many targeted therapies are most successful in the presence of normal genes (e.g., EGFR inhibitors work only in the presence of wild-type KRAS). A quantitative assessment of circulating tumor DNA levels becomes especially significant in these cases where mutations in genes are not found. In other words, the demonstrable presence of circulating tumor DNA coupled with a wild-type sequencing result at a particular target gene suggests that the target gene is normal within the tumor, and such results can have a significant influence in guiding the choice of therapies. Such is the case with the OVA1 sample highlighted in FIG. 24. The presence of the KRAS mutation in the cfDNA library suggested that this patient's tumor harbors a wild-type TP53 gene(s).

[0325] Another example of aberrant gene discovery is shown in FIG. 25. The targeted quantitative genetic analysis system revealed the presence of a significant amplification in the ERBB2 gene, otherwise referred to as HER-2/neu. While this type of amplification has been much publicized in breast cancer, it is occasionally identified in colorectal carcinomas as well.

#### Conclusion

[0326] Validation experiments with cell line DNA revealed the thresholds of detection of three types of genetic variation that are central to driving neoplastic growth in cancers. Characterization of cfDNA derived from cancer patients revealed tumor-associated genetic changes that were well above the thresholds set by reconstruction experiments in all four samples analyzed. These data indicated that the quantitative analysis system contemplated herein may have significant clinical utility, especially in settings where liquid biopsies are most appropriate.

[0327] In general, in the following claims, the terms used should not be construed to limit the claims to the specific embodiments disclosed in the specification and the claims, but should be construed to include all possible embodiments along with the full scope of equivalents to which such claims are entitled. Accordingly, the claims are not limited by the disclosure.

---

#### SEQUENCE LISTING

```

Sequence total quantity: 562
SEQ ID NO: 1      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 1
taaacattgg aaaggtttct aattaaccag gagatccaaa agaaagcggt tcaagtagca 60

SEQ ID NO: 2      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 2
gatctcagtt ttttggta actatgtatt ttggatatg aagcttctgg gtttgcaca 60

SEQ ID NO: 3      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 3
gacagataag catacatatt aacatggata tatatgtcaa tttcattcaa atggcttc 60

```

---

-continued

---

```

SEQ ID NO: 4      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 4
agctcttagc ctttgggggg atgacactct tgagcggacg tggggacgcc tcgtcttta 60

SEQ ID NO: 5      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 5
aagccccccac cgccgcctct ttccaaaata aacaccagcc agccgcggag cccggagtgc 60

SEQ ID NO: 6      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 6
gcctcccttc cccctccccc cccgacageg gccgctcggg ccccggtctt cggttataag 60

SEQ ID NO: 7      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 7
ggtgtggcag ccaggggggc gcactctgt ctggctgggc cccttctccc atgtttctt 60

SEQ ID NO: 8      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 8
ttacacaacc tttgggcttg gacaacactt tgggtccaa agaacctaag agtctttctg 60

SEQ ID NO: 9      moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 9
tgatgaaact tgggctggat ggggcacagg tagggtgctt gttgcttca gtcagatgaa 60

SEQ ID NO: 10     moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 10
aatgaaagaa aaggaggcca gattgctact cctggccctt gccacacact agtacccta 60

SEQ ID NO: 11     moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 11
attgacaata cctacataaa actcttcca gaatgttgtt aagtcttagt cattagggag 60

SEQ ID NO: 12     moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 12
ggatattccac caacactgta ttcatgtacc cattttctc ttaacctaac tttattggtc 60

SEQ ID NO: 13     moltype = DNA length = 60
FEATURE          Location/Qualifiers
source           1..60
mol_type = other DNA

```

---

---

-continued

---

```

SEQUENCE: 13          organism = synthetic construct
caaaggggga aaaccatcg gacattattt aacaacgaa atatctaact gaaaggcaa  60

SEQ ID NO: 14          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 14          organism = synthetic construct
caggcagacc aaccaaagtc tttgttccac cttttaaac taaatcacat tttcacagag  60

SEQ ID NO: 15          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 15          organism = synthetic construct
ccccagccag cggtccgcaa cccttgccgc atccacgaaa ctttgcccat agcagcggc  60

SEQ ID NO: 16          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 16          organism = synthetic construct
cgactcatct cagcattaaa gtgataaaaa aataaattaa aaggcaagtg gacttcggtg  60

SEQ ID NO: 17          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 17          organism = synthetic construct
ctgtggcgcg cactgcccgc tgccgcaggt ttccgcacca agacccttt aactcaagac  60

SEQ ID NO: 18          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 18          organism = synthetic construct
ttctactgcg acgaggaggaa gaacttctac cagcagcagc agcagagcga gctgcagccc  60

SEQ ID NO: 19          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 19          organism = synthetic construct
accgagctgc tgggaggaga catggtgaac cagagttca tctgcgaccc ggacgacag  60

SEQ ID NO: 20          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 20          organism = synthetic construct
ggccgcgcct cagagtgcatt cgaccgcctcg gtggtcttc cctaccctct caacgacagc  60

SEQ ID NO: 21          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 21          organism = synthetic construct
ggcggctagg ggacaggggc ggggtggca gcagctcgaa tttctccag atatcctcgc  60

SEQ ID NO: 22          moltype = DNA  length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 22          organism = synthetic construct
agacgagctt ggccggggcc gagaagccgc tccacataca gtcctggatg atgatgttt  60

```

---

-continued

---

```

SEQ ID NO: 23      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 23
aggagagcag agaatccgag gacggagaga aggcgctgga gtcttgcgag gcgcaggact 60

SEQ ID NO: 24      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 24
taagagtggc ccgttaata agctgcaat gaaaatggga aaggtatcca gcccggact 60

SEQ ID NO: 25      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 25
ttgtatttgt acagcattaa tctggtaatt gattattta atgtaacctt gctaaaggag 60

SEQ ID NO: 26      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 26
gaggccacag caaacctcct cacagccac tggcctcaa gaggtgcac gtcacac 60

SEQ ID NO: 27      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 27
agaggaggaa cgagctaaa cggagcttt ttgccctgcg tgaccagatc ccggagttgg 60

SEQ ID NO: 28      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 28
tccaacttga ccctttggc agcaggatag tccttccag tggaggagg cgctgcgtt 60

SEQ ID NO: 29      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 29
gcttggacgg acaggatgtt tgctgtggc ttttaagga taactacctt gggggccctt 60

SEQ ID NO: 30      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 30
gcatttgcatttga aacaagttca taggtgattt ctcaggatcat ttctgttaga 60

SEQ ID NO: 31      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 31
cgccccgcgc cctcccaagcc gggtccagcc ggagccatgg ggccggagcc gcagtggca 60

SEQ ID NO: 32      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA

```

---

---

-continued

---

```

SEQUENCE: 32          organism = synthetic construct
ctctggcccc gccggcccgcg ggacctcgcc ggggcattcca cagggcaggg tcccggcgct 60

SEQ ID NO: 33          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 33
ggcatgactt ggagttagtt tggatgggtt ggccaggctt gagaaggctt cccggcaactg 60

SEQ ID NO: 34          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 34
gcagggcacc ttcttcgtcc acccacctgt aaacagaggg ctcaagccag ctggaggcag 60

SEQ ID NO: 35          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 35
cccaagatct ccaagtactg gggaaacccca gggaggccct ggggggtggc agtgttctta 60

SEQ ID NO: 36          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 36
ctaatgcaca caaagcctcc ccctggtagt cagtgccctt ggtcagctt gaataaccaa 60

SEQ ID NO: 37          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 37
ctgctccctt tttagaaggc aggaggggccca caagggaagc agaagggttac agaaggggaa 60

SEQ ID NO: 38          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 38
tggggcactg gcgggcaggc actgggttgt aagttggag tttgcggctg gggtcaggct 60

SEQ ID NO: 39          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 39
tctgctgttgc tttgtgcctt tctctgttac taacccgtcc tctcgctttt agacatctt 60

SEQ ID NO: 40          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 40
cccacccctt ccatgtcacc ttttatgacac ctgcattcca cccggcccca gcccctccct 60

SEQ ID NO: 41          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 41
tggggcagggt agtctcccta gaagggtatg ctgtatgggg tctgggtcccc agggcgccac 60

```

---

-continued

---

```

SEQ ID NO: 42      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 42
ggtgccacc cttgcattc tgggggttag agcacattgg gcacaaagca gaggcacata 60

SEQ ID NO: 43      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 43
caccctgccc ggtactgcc tattgccc ggcacaccag ggcaaaacag cacagtggaa 60

SEQ ID NO: 44      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 44
ccatttacag aaacaaacct cccccccaaa atgagaaaac tgtgttttc cctggcactc 60

SEQ ID NO: 45      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 45
ttatttttct tgtgcctggg cacggtaatg ctgctcatgg tggtgacgga agggccagg 60

SEQ ID NO: 46      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 46
gaaggatagg acaggggtggg ctggccagg ctgcattgcg agagggacag gaactgcagc 60

SEQ ID NO: 47      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 47
ggcccgac cctgatgctc atgtggctgt tgacctgtcc cggtatgaag gctgagacgg 60

SEQ ID NO: 48      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 48
tctgtctccatccccca agagatgctg ccacatctgg atcctcagga ctctgtctgc 60

SEQ ID NO: 49      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 49
tcacgtccca gggcagttttt cttccctgaa gaaagttgga tggcatgatc tgtctccca 60

SEQ ID NO: 50      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 50
gtgttgagaa cagactactg acttctaata gcagcgactt cttaacccata ataaaccaca 60

SEQ ID NO: 51      moltype = DNA length = 60
FEATURE           Location/Qualifiers
source            1..60
mol_type = other DNA

```

---

-continued

---

```

SEQUENCE: 51          organism = synthetic construct
aaaaaaaaagga tgggttccat atgggtggtg tcaagtgccc acctccttagc aagtcagcag 60

SEQ ID NO: 52          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 52          moltype = DNA length = 60
ccttcacaag gtcaaagcta tacatcagct cctgtgacat tgactcatcc cccagacatt 60

SEQ ID NO: 53          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 53          moltype = DNA length = 60
aaccaccccgaa gatctgcaaa ctttgcagga tgcaccagat gtctttagc catgggtcaa 60

SEQ ID NO: 54          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 54          moltype = DNA length = 60
tgccctccctc aggtgcctct gcacaaaacc agattgcctc cctctaagag tatggtagt 60

SEQ ID NO: 55          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 55          moltype = DNA length = 60
gttttacttt gcctttagct gtggtcttcc aaaccacccat ccctccttat cttcctctgc 60

SEQ ID NO: 56          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 56          moltype = DNA length = 60
ctctgcaatt tgttttccca tattaaagaa ctgaagagct cagtgtggta ggctggcaag 60

SEQ ID NO: 57          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 57          moltype = DNA length = 60
tttaaatga tgttttaaag aattgaaact aacatactgt tctgctttct cccccgggtt 60

SEQ ID NO: 58          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 58          moltype = DNA length = 60
cctgcccacc acttctccct aaactgaaggccacatting gagcagtcatttctt 60

SEQ ID NO: 59          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 59          moltype = DNA length = 60
ggttgcgttc caatcatgtt acataaccta cgccaaggta tcgacaggat catactcctg 60

SEQ ID NO: 60          moltype = DNA length = 60
FEATURE
source
1..60
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 60          moltype = DNA length = 60
gcacagttac atttgcgttggcaacatcc ttaaaaatta ataactgata ggtcacggac 60

```

---

-continued

---

SEQ ID NO: 61	moltype = DNA length = 60
FEATURE	Location/Qualifiers
source	1..60
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 61	
cgtccccggc cgccacttgg cctgcctccg tcccgccgca ccacttcgccc tgccctccgtc	60
SEQ ID NO: 62	moltype = DNA length = 60
FEATURE	Location/Qualifiers
source	1..60
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 62	
ctgttaaggcg aggaggacga tgcgtcccct ccctcgcagg attgaggtta ggactaaacg	60
SEQ ID NO: 63	moltype = DNA length = 35
FEATURE	Location/Qualifiers
source	1..35
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 63	
gtgaaaaaccg ggtcaactc ccgtgccagt cacat	35
SEQ ID NO: 64	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 64	
nnnnnnnnnc tggccgcagg	20
SEQ ID NO: 65	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 65	
nnnnnnnnnat cttagtgca	20
SEQ ID NO: 66	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 66	
nnnnnnnnnncg gaactcggag	20
SEQ ID NO: 67	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 67	
nnnnnnnnnnga ctccgatccc	20
SEQ ID NO: 68	moltype = DNA length = 95
FEATURE	Location/Qualifiers
source	1..95
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 68	
atgtgactgg cacgggagtt gatctgggtt ttacacgggtt tgagtgccat gagctaccta	60
ctggatgtgc ctgactgttt ccccttcttc ttccc	95
SEQ ID NO: 69	moltype = DNA length = 95
FEATURE	Location/Qualifiers
source	1..95
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 69	
atgtgactgg cacgggagtt gatctgggtt ttacacctata tccaggatgg agagaggaa	60
aaaaaaagatg ggtctgtgtg ggagggcagg tactt	95
SEQ ID NO: 70	moltype = DNA length = 95
FEATURE	Location/Qualifiers

-continued

---

```

source          1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 70
atgtgactgg cacgggagtt gatctgggtt ttcacgaaag aaggcaggtc ttcaattaat 60
aagattccct ggtctcgttt gtctacctgt taatg                           95

SEQ ID NO: 71      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 71
atgtgactgg cacgggagtt gatctgggtt ttcaccagac tcgcgccaa tttccccca 60
ccccttgta ttgccacaaa atcctgagga tgata                           95

SEQ ID NO: 72      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 72
atgtgactgg cacgggagtt gatctgggtt ttcacaagca cctagccca ttctgtcga 60
gcaggaggtg gcaggtaccc cagactggga ggtaa                           95

SEQ ID NO: 73      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 73
atgtgactgg cacgggagtt gatctgggtt ttcacagtgc gtggggccag gatgaggccc 60
agtctgttca cacatggctg ctgcctctca gctct                           95

SEQ ID NO: 74      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 74
atgtgactgg cacgggagtt gatctgggtt ttcaccctgg ccctcagcca gtacagaaaag 60
tcatttgtca aggccttcag ttgcagacg tgctc                           95

SEQ ID NO: 75      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 75
atgtgactgg cacgggagtt gatctgggtt ttcacagaat tcattgccag ctataaatct 60
gtggaaacgc tgccacacaa tcttagcaca caaga                           95

SEQ ID NO: 76      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 76
atgtgactgg cacgggagtt gatctgggtt ttcacttact tccctccagt ttttgtgctt 60
gcaaaaacaac agaatcttct ctccatgaaa tcatg                           95

SEQ ID NO: 77      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 77
atgtgactgg cacgggagtt gatctgggtt ttcaccagggt gtatcttattt tccccattt 60
ctcacaaagg aaaccaagat aaaaggtaa aatgg                           95

SEQ ID NO: 78      moltype = DNA  length = 95
FEATURE           Location/Qualifiers
source            1..95
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 78

```

---

-continued

---

```

atgtgactgg cacggagtt gatcctggtt ttcactgtta cctttaaaag acatctgctt 60
tctgcacaaa ttaatgtgct gaacttaaac ttacc 95

SEQ ID NO: 79      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 79
atgtgactgg cacggagtt gatcctggtt ttcacttccc agtaaattac tcttaccaat 60
gcaacagact ttaaagaagt tgtgtttac aatgc 95

SEQ ID NO: 80      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 80
atgtgactgg cacggagtt gatcctggtt ttcactaaat gacataacag ttatgattt 60
gcagaaaaca gatctgtatt tatttcagtg ttact 95

SEQ ID NO: 81      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 81
atgtgactgg cacggagtt gatcctggtt ttcacgacag gtttgaaag atatttgtt 60
tactaatgac tgtgctataa ctttttttc ttcc 95

SEQ ID NO: 82      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 82
atgtgactgg cacggagtt gatcctggtt ttcacctgtg gcgcgactg cgcgctgcgc 60
cagggttccg caccaagacc cctttaactc aagac 95

SEQ ID NO: 83      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 83
atgtgactgg cacggagtt gatcctggtt ttcacggcgg ctaggggaca gggggggggt 60
gggcagcagc tcgaatttct tccagatatac ctgc 95

SEQ ID NO: 84      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 84
atgtgactgg cacggagtt gatcctggtt ttcacaccga gctgctggga ggagacatgg 60
tgaaccagag tttcatctgc gaccggacg acgag 95

SEQ ID NO: 85      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 85
atgtgactgg cacggagtt gatcctggtt ttcacaggag agcagagaat ccgaggacgg 60
agagaaggcg ctggagtctt gcgaggcgcga ggact 95

SEQ ID NO: 86      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 86
atgtgactgg cacggagtt gatcctggtt ttcacctgtt agttatcgta aaaaggagca 60
tcttaggtttagc caatgttacc cgatt 95

SEQ ID NO: 87      moltype = DNA length = 95

```

-continued

FEATURE source Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 87 atgtgactgg cacgggagtt gatcctgggtt ttccaaatgg ccattttcc aggaggcaca 60  
gaaattacag gccatgcaca gagagaata cccga 95

SEQ ID NO: 88 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 88 atgtgactgg cacgggagtt gatcctgggtt ttccacctgt tcgttccttg tactgagacc 60  
ctagtctgcc actgaggatt tggttttgc ccttc 95

SEQ ID NO: 89 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 89 atgtgactgg cacgggagtt gatcctgggtt ttccacatcaa gactcatcg taccatcaa 60  
agttgatgaa acatgtgtaa agttcaaca gaaat 95

SEQ ID NO: 90 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 90 atgtgactgg cacgggagtt gatcctgggtt ttcaactgtgt ccagctgtga aactcagaga 60  
tgttaactgtc gacatccccc ctatttgca tctca 95

SEQ ID NO: 91 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 91 atgtgactgg cacgggagtt gatcctgggtt ttccacattt aaacaatttt atcatgaatg 60  
ccatgacca agtattcttc tgtatcttct ttctt 95

SEQ ID NO: 92 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 92 atgtgactgg cacgggagtt gatcctgggtt ttcaactgtatg ggtgggtcc cgaaggggcc 60  
tccccagac ttgcgaaggcccactctctt gggcg 95

SEQ ID NO: 93 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 93 atgtgactgg cacgggagtt gatcctgggtt ttccaccagggt tgccggggca tccaggctgc 60  
ccaagcggag gctggggccgg ctgtgctggc ctctt 95

SEQ ID NO: 94 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

SEQUENCE: 94 atgtgactgg cacgggagtt gatcctgggtt ttcaactttt aaatgtgggt ttgttgcatt 60  
gaaacgtgttcaagcatag ttttgacaga taacg 95

SEQ ID NO: 95 moltype = DNA length = 95  
FEATURE Location/Qualifiers  
source 1..95 mol\_type = other DNA  
organism = synthetic construct

---

-continued

---

```

SEQUENCE: 95
atgtgactgg cacgggagtt gatcctgggtt ttcaactgccc taaaagtgtta tgtataacat 60
ccctgatgtc tgcatggc ctttgactgg tggtt 95

SEQ ID NO: 96          moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 96
atgtgactgg cacgggagtt gatcctgggtt ttcaacaacc ctcgagggtc agaccttgg 60
agcaggagtgc tgattctggc caaccaccc ctctg 95

SEQ ID NO: 97          moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 97
atgtgactgg cacgggagtt gatcctgggtt ttaccataaa atatgtgtgc tagtcctgtt 60
agacccaagt gtcggccaa ggcagcgccc tgctc 95

SEQ ID NO: 98          moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 98
atgtgactgg cacgggagtt gatcctgggtt ttcaactactt gttaattaaa aattcaagag 60
tttttttttc ttattctgag gttatcttt tacca 95

SEQ ID NO: 99          moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 99
atgtgactgg cacgggagtt gatcctgggtt ttcaacccaaa atctgttttc caataaattc 60
tcagatccag gaagaggaaa ggaaaaacat caaaa 95

SEQ ID NO: 100         moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 100
atgtgactgg cacgggagtt gatcctgggtt ttcacatact ccatactcccg taaaaatagt 60
gagacttgag taatgtttga tgtcaacttgt cttc 95

SEQ ID NO: 101         moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 101
atgtgactgg cacgggagtt gatcctgggtt ttcaaccgtc accactatata tattcttagt 60
atccccagaaa agttaaagtca aatctgaaa catat 95

SEQ ID NO: 102         moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 102
atgtgactgg cacgggagtt gatcctgggtt ttcaaccgcc cgcgtccgac ccgcggatcc 60
cgcggcgtcc ggcccccgtg gtctggatcg cggag 95

SEQ ID NO: 103         moltype = DNA  length = 95
FEATURE
source
1..95
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 103
atgtgactgg cacgggagtt gatcctgggtt ttcaaccata cgggcagcac gacgcgcgga 60
ctgcgattgc agaagatgac ctggggaggc tcgcg 95

```

---

-continued

---

```

SEQ ID NO: 104      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 104
atgtgactgg cacgggagtt gatcctgggtt ttcaactagag gggcttcaga ccgtgctatc 60
gtccctgtc ggtcgccgcct aagegccggg cccgt 95

SEQ ID NO: 105      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 105
atgtgactgg cacgggagtt gatcctgggtt ttacacggcgc cgaggaggag atggaggccg 60
ggcgccgcgc gcccgtgtc actcg 95

SEQ ID NO: 106      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 106
atgtgactgg cacgggagtt gatcctgggtt ttacacgggtt gggccaccgt gcccagccac 60
cggtgtggct cttaacaac ctttgcttgc cccga 95

SEQ ID NO: 107      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 107
atgtgactgg cacgggagtt gatcctgggtt ttacacaagtg gtctatccctg tacttaccac 60
aacaaccta tctttttaaa aagtaaaacg tcagt 95

SEQ ID NO: 108      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 108
atgtgactgg cacgggagtt gatcctgggtt ttacacccgt tggttccttg tactgagacc 60
ctagtctgcc actgaggatt tggttttgc ccttc 95

SEQ ID NO: 109      moltype = DNA length = 95
FEATURE           Location/Qualifiers
source            1..95
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 109
atgtgactgg cacgggagtt gatcctgggtt ttcacatcaa gactcatcag taccatcaa 60
agctgagatg aaacagtgtt agttcaaca gaaat 95

SEQ ID NO: 110      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 110
atgtgactgg cacgggagtt gatcctgggtt ttacacaccta ctggatgtgc ctgactgttt 60
cccccttcc ttccc 75

SEQ ID NO: 111      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 111
atgtgactgg cacgggagtt gatcctgggtt ttacacggaa aaaaaagatg ggtctgtgt 60
ggagggcagg tactt 75

SEQ ID NO: 112      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA

```

---

---

-continued

---

```

SEQUENCE: 112          organism = synthetic construct
atgtgactgg cacggagtt gatcctgggtt ttcacttaat aagattccct ggtctcggtt 60
gtctacctgt taatg                75

SEQ ID NO: 113          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 113
atgtgactgg cacggagtt gatcctgggtt ttcaccccca ccccttgta ttgccacaaa 60
atcctgagga tgatc                75

SEQ ID NO: 114          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 114
atgtgactgg cacggagtt gatcctgggtt ttcacgtga gcaggagggtg gcaggtaccc 60
cagactggaa ggtaa                75

SEQ ID NO: 115          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 115
atgtgactgg cacggagtt gatcctgggtt ttcacggccc agtctgttca cacatggctg 60
ctgcctctca gctct                75

SEQ ID NO: 116          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 116
atgtgactgg cacggagtt gatcctgggtt ttcacgaaag tcatttgta aggcattcag 60
ttggcagacg tgctc                75

SEQ ID NO: 117          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 117
atgtgactgg cacggagtt gatcctgggtt ttcacaatct gtggaaacgc tgccacacaa 60
tcttagcaca caaga                75

SEQ ID NO: 118          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 118
atgtgactgg cacggagtt gatcctgggtt ttcactgctt gcaaaacaac agaatcttct 60
ctccatgaaa tcatg                75

SEQ ID NO: 119          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 119
atgtgactgg cacggagtt gatcctgggtt ttcacatttt ctcacaaagg aaaccaagat 60
aaaagggtta aatgg                75

SEQ ID NO: 120          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 120
atgtgactgg cacggagtt gatcctgggtt ttcactgctt tctgccaaaa ttaatgtgct 60
gaacttaaac ttacc                75

```

---

-continued

---

```

SEQ ID NO: 121      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 121
atgtgactgg cacgggagtt gatcctggtt ttcacccaat gcaacagact ttaaaagaagt 60
tgtgttttac aatgc 75

SEQ ID NO: 122      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 122
atgtgactgg cacgggagtt gatcctggtt ttcacatttt gcagaaaaca gatctgtatt 60
tatttcagtg ttact 75

SEQ ID NO: 123      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 123
atgtgactgg cacgggagtt gatcctggtt ttcactgtgt tactaatgac tgtgtataaa 60
cttttttc ttcc 75

SEQ ID NO: 124      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 124
atgtgactgg cacgggagtt gatcctggtt ttcactgegc caggttccg caccagacc 60
cctttaactc aagac 75

SEQ ID NO: 125      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 125
atgtgactgg cacgggagtt gatcctggtt ttcacgggtt gggcagcagc tcgaatttct 60
tccagatatac ctcgc 75

SEQ ID NO: 126      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 126
atgtgactgg cacgggagtt gatcctggtt ttcaccatgg tgaaccagag tttcatctgc 60
gaccggacg acgag 75

SEQ ID NO: 127      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 127
atgtgactgg cacgggagtt gatcctggtt ttcacgacgg agagaaggcg ctggagttct 60
gcgaggcgca ggact 75

SEQ ID NO: 128      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 128
atgtgactgg cacgggagtt gatcctggtt ttcacgagca tcttaggtagg tctttgtac 60
caatgttacc cgatt 75

SEQ ID NO: 129      moltype = DNA  length = 75
FEATURE
source
1..75

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct

SEQUENCE: 129
atgtgactgg cacgggagtt gatcctgggtt ttcacgcaca gaaattacag gccatgcaca 60
gagagaaaata cccga 75

SEQ ID NO: 130      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 130
atgtgactgg cacgggagtt gatcctgggtt ttcacagacc ctagtctgcc actgaggatt 60
tggttttgc ctttc 75

SEQ ID NO: 131      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 131
atgtgactgg cacgggagtt gatcctgggtt ttcactcaaa agctgagatg aaacagtgtta 60
agtttcaaca gaaat 75

SEQ ID NO: 132      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 132
atgtgactgg cacgggagtt gatcctgggtt ttcacagaga tgttaactgct gacatccccc 60
ctattttgca tctca 75

SEQ ID NO: 133      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 133
atgtgactgg cacgggagtt gatcctgggtt ttcacgaatg ccatgaccaa agtatttttc 60
tgttatcttct ttctt 75

SEQ ID NO: 134      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 134
atgtgactgg cacgggagtt gatcctgggtt ttcacgggcc tcccgacac ttgcgaagtt 60
cccaactctct gggcg 75

SEQ ID NO: 135      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 135
atgtgactgg cacgggagtt gatcctgggtt ttcacgctgc ccaagcggag gctggccgg 60
ctgtgctggc ctctt 75

SEQ ID NO: 136      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 136
atgtgactgg cacgggagtt gatcctgggtt ttcacgcat gaaacgttt tcaagcatag 60
ttttgacaga taacg 75

SEQ ID NO: 137      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 137
atgtgactgg cacgggagtt gatcctgggtt ttcacaacat ccctgtatgc tgcatggc 60

```

-continued

---

```

ctttgactgg tgttt                                75
SEQ ID NO: 138      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 138
atgtgactgg cacgggagtt gatcctgggtt ttcaacttgg agcaggagtg tgattctggc 60
caaccaccct ctctg                                75

SEQ ID NO: 139      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 139
atgtgactgg cacgggagtt gatcctgggtt ttcaacctgtt agacccaagt gctccccaaag 60
ggcagcgcgc tgctc                                75

SEQ ID NO: 140      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 140
atgtgactgg cacgggagtt gatcctgggtt ttcaacaagag ttttttttc ttattcttag 60
gttatctttt tacca                                75

SEQ ID NO: 141      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 141
atgtgactgg cacgggagtt gatcctgggtt ttcaacaattc tcagatccag gaagaggaaa 60
ggaaaaaacat caaaa                                75

SEQ ID NO: 142      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 142
atgtgactgg cacgggagtt gatcctgggtt ttcacatagt gagacttgag taatgtttga 60
tgtcaacttgtt ctttc                                75

SEQ ID NO: 143      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 143
atgtgactgg cacgggagtt gatcctgggtt ttcaacttaggt atcccagaaa agttaaagtc 60
aaatctgaaa catat                                75

SEQ ID NO: 144      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 144
atgtgactgg cacgggagtt gatcctgggtt ttcaacgatcc cgccggcgatcc ggccgggtg 60
gtctggatcg cggag                                75

SEQ ID NO: 145      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 145
atgtgactgg cacgggagtt gatcctgggtt ttcaacgatcc ctgcgattgc agaagatgac 60
ctggggagggc tcgcg                                75

SEQ ID NO: 146      moltype = DNA  length = 75
FEATURE

```

---

-continued

---

```

source          1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 146
atgtgactgg cacgggagtt gatcctgggtt ttcacccatac gtccctgcgtg ggtcgccct 60
aagcgccggg cccgt                               75

SEQ ID NO: 147      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 147
atgtgactgg cacgggagtt gatcctgggtt ttcacggccg ggccggcccg gcccgtgtg 60
cgctcggtga actcg                               75

SEQ ID NO: 148      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 148
atgtgactgg cacgggagtt gatcctgggtt ttcacgccac cggtgtggct cttaacaac 60
cttgcttgt cccga                               75

SEQ ID NO: 149      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 149
atgtgactgg cacgggagtt gatcctgggtt ttcacaccac aacaaccta tcttttaaa 60
aagtaaaacg tcagt                               75

SEQ ID NO: 150      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 150
atgtgactgg cacgggagtt gatcctgggtt ttcacagacc ctatgtgtgcc actgaggatt 60
tgggttttgc ctttc                               75

SEQ ID NO: 151      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 151
atgtgactgg cacgggagtt gatcctgggtt ttcactcaa agctgagatg aaacagtgt 60
agttcaaca gaaat                               75

SEQ ID NO: 152      moltype = DNA  length = 58
FEATURE           Location/Qualifiers
source            1..58
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 152
aatgatacgg cgaccaccga gatctacacg tcatgcgagg ccagagaatt cgaataca 58

SEQ ID NO: 153      moltype = DNA  length = 57
FEATURE           Location/Qualifiers
source            1..57
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 153
caagcagaag acggcatacg agatgtgact ggcacgggg ttgatcctgg ttttcac 57

SEQ ID NO: 154      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 154
atgtgactgg cacgggagtt gatcctgggtt ttcaccgaat gaggggtatg ttttcccg 60
gcacccctt caggat                               75

```

---

-continued

---

```

SEQ ID NO: 155      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 155
atgtgactgg cacgggagtt gatcctggtt ttcacgttg agtcggcat gatggtcag 60
gtgcggagct tgctc          75

SEQ ID NO: 156      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 156
atgtgactgg cacgggagtt gatcctggtt ttcacgcaga tcctggtgc tccggcgta 60
cactgcaggat gggtg          75

SEQ ID NO: 157      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 157
atgtgactgg cacgggagtt gatcctggtt ttcacctaca caggccactt cctacaggaa 60
gcctccctgg atetc          75

SEQ ID NO: 158      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 158
atgtgactgg cacgggagtt gatcctggtt ttcacgaaat actaataaaa tgattaaga 60
aggtgtgtct ttaat          75

SEQ ID NO: 159      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 159
atgtgactgg cacgggagtt gatcctggtt ttcactataat ggaaaataat tatttgtatt 60
atataaggca gagtc          75

SEQ ID NO: 160      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 160
atgtgactgg cacgggagtt gatcctggtt ttcacattag acccaatatg gtctgcagat 60
ttttagagaa gaaat          75

SEQ ID NO: 161      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 161
atgtgactgg cacgggagtt gatcctggtt ttcacgtgaa ccagcagact gtgttgcag 60
tataacccca cgtga          75

SEQ ID NO: 162      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 162
atgtgactgg cacgggagtt gatcctggtt ttcacgcat ggagcctaag gaagttcag 60
caaggcccta agggg          75

SEQ ID NO: 163      moltype = DNA  length = 75
FEATURE
source
1..75

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct

SEQUENCE: 163
atgtgactgg cacgggagtt gatcctggtt ttcaccccg gaattggcct gccttagtat 60
ttctgctgtg ctca 75

SEQ ID NO: 164      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 164
atgtgactgg cacgggagtt gatcctggtt ttcacttga gggatcgttgg 60
tcagttgtgt ttcc 75

SEQ ID NO: 165      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 165
atgtgactgg cacgggagtt gatcctgggtt ttcaccat catgaaaaga tctctgaatt 60
ggtgtctggg gatct 75

SEQ ID NO: 166      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 166
atgtgactgg cacgggagtt gatcctggtt ttcactgagg accaggtcac aggaccttt 60
tggactgcag ttcc 75

SEQ ID NO: 167      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 167
atgtgactgg cacgggagtt gatcctggtt ttcactaacc actgccactc cccaccctct 60
agggttgtaa atgaa 75

SEQ ID NO: 168      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 168
atgtgactgg cacgggagtt gatcctggtt ttcacgagct ctaccaatgt gagtgaccat 60
tatcactctt acatg 75

SEQ ID NO: 169      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 169
atgtgactgg cacgggagtt gatcctggtt ttcacaaaat tggatcgttggtagatt 60
ctgtgtgtaa agccc 75

SEQ ID NO: 170      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 170
atgtgactgg cacgggagtt gatcctggtt ttcactatgt gtcgtttcc ctctctatg 60
caatggaccg accgt 75

SEQ ID NO: 171      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 171
atgtgactgg cacgggagtt gatcctggtt ttcacgtgtaa aattgccgag cacgtagtaa 60

```

---

-continued

---

ccatgcaaca agtgt	75
SEQ ID NO: 172	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 172	
atgtgactgg cacgggagtt gatcctgggtt ttcaactgggg acacagtgtg tgctgccatc	60
tcccttctac cggca	75
SEQ ID NO: 173	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 173	
atgtgactgg cacgggagtt gatcctgggtt ttcaacaagag cctttccctc tgccctttc	60
aagcctctgc ccatac	75
SEQ ID NO: 174	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 174	
atgtgactgg cacgggagtt gatcctgggtt ttcacgacca cactgagttc tctgtgacct	60
gcagggtcagc tcacc	75
SEQ ID NO: 175	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 175	
atgtgactgg cacgggagtt gatcctgggtt ttcaactttcc tatctctctg cctggagggt	60
ggttggaggggc tggtt	75
SEQ ID NO: 176	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 176	
atgtgactgg cacgggagtt gatcctgggtt ttcacaaaca ggagctgcgc cggtggaagc	60
atgtgggagc tagaa	75
SEQ ID NO: 177	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 177	
atgtgactgg cacgggagtt gatcctgggtt ttcaacggaca ctgaaggagc tccccacccc	60
ctgtatcagcc aggag	75
SEQ ID NO: 178	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 178	
atgtgactgg cacgggagtt gatcctgggtt ttcaacggaa ctgcagctgc tctgggggg	60
ggaagggttg gagct	75
SEQ ID NO: 179	moltype = DNA length = 75
FEATURE	Location/Qualifiers
source	1..75
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 179	
atgtgactgg cacgggagtt gatcctgggtt ttcaacaccca attccaggga ctgcataaac	60
gaagtgcacac cttgg	75
SEQ ID NO: 180	moltype = DNA length = 75
FEATURE	Location/Qualifiers

---

-continued

---

```

source          1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 180
atgtgactgg cacgggagtt gatctgggtt ttcaccctgc ccccttggga gtccctgggg 60
ctctgtgcac tcacc                               75

SEQ ID NO: 181      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 181
atgtgactgg cacgggagtt gatctgggtt ttcacggaag caccccccgtt attaaaacga 60
acggggcggaa aagaa                               75

SEQ ID NO: 182      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 182
atgtgactgg cacgggagtt gatctgggtt ttcacctaacc aaaggggacg cgaccgggg 60
tccagtgtccc caggg                               75

SEQ ID NO: 183      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 183
atgtgactgg cacgggagtt gatctgggtt ttcaccctgg ggggactggg tggcctcacc 60
cccaaccggc tcatac                               75

SEQ ID NO: 184      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 184
atgtgactgg cacgggagtt gatctgggtt ttcaccgcgc tccagtttctt cgccggcgga 60
gaagccgcgtc catat                               75

SEQ ID NO: 185      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 185
atgtgactgg cacgggagtt gatctgggtt ttcaccccac ccggccgcgg agtgcgttgg 60
tcccgccgtt gtctt                               75

SEQ ID NO: 186      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 186
atgtgactgg cacgggagtt gatctgggtt ttcacggca cgggcgttgg ctgcgcgtt 60
ttcacggaa aaaaaa                               75

SEQ ID NO: 187      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 187
atgtgactgg cacgggagtt gatctgggtt ttcacaacat ggatatatat gtgaatttca 60
ttcaaatggt tctca                               75

SEQ ID NO: 188      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 188

```

---

-continued

---

```

atgtgactgg cacggagtt gatcctgggtt ttcaactaaac caacattctt aatgtcaaca 60
caatgttgtt ttaaa 75

SEQ ID NO: 189      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 189
atgtgactgg cacggagtt gatcctgggtt ttcaccccta cgtggagagt gaggatgcac 60
ccccacagaa gaaga 75

SEQ ID NO: 190      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 190
atgtgactgg cacggagtt gatcctgggtt ttcacatgac actcttgagc ggacgtgggg 60
acgcctcgct ctta 75

SEQ ID NO: 191      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 191
atgtgactgg cacggagtt gatcctgggtt ttcaactctca cgctcaggga ccacgtgccg 60
gagttggtaa agaat 75

SEQ ID NO: 192      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 192
atgtgactgg cacggagtt gatcctgggtt ttcaccagtgc gccttttca aaatgaccac 60
cttggcgccc ttctc 75

SEQ ID NO: 193      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 193
atgtgactgg cacggagtt gatcctgggtt ttcacgctag ggggctgggg ttgggggtggg 60
ggttggggcc ctgcc 75

SEQ ID NO: 194      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 194
atgtgactgg cacggagtt gatcctgggtt ttcaccagtgc tccataggc tgaaaatgtt 60
tcctgactca gaggg 75

SEQ ID NO: 195      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 195
atgtgactgg cacggagtt gatcctgggtt ttcaccgtgc atggaggagc cgcaatcaga 60
tcctgactca gagcc 75

SEQ ID NO: 196      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 196
atgtgactgg cacggagtt gatcctgggtt ttcaactcatg ctggatcccc actttcctc 60
ttgcagcagc cagac 75

SEQ ID NO: 197      moltype = DNA length = 75

```

---

-continued

---

```

FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 197
atgtgactgg cacgggagtt gatcctggtt ttcacagccc cccagccctc caggccccca 60
gcctccagg tcccc 75

SEQ ID NO: 198      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 198
atgtgactgg cacgggagtt gatcctggtt ttcacgcaga gacctgtggg aagcgaaaat 60
tccatggac tgact 75

SEQ ID NO: 199      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 199
atgtgactgg cacgggagtt gatcctggtt ttcacgcagg gggatacggc caggcatgta 60
agtctcatgg aagcc 75

SEQ ID NO: 200      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 200
atgtgactgg cacgggagtt gatcctggtt ttcacccgtg caagtcacag acttggctgt 60
cccagaatgc aagaa 75

SEQ ID NO: 201      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 201
atgtgactgg cacgggagtt gatcctggtt ttcacccaga aaacctacca gggcagctac 60
ggtttccgtc tgggc 75

SEQ ID NO: 202      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 202
atgtgactgg cacgggagtt gatcctggtt ttcacgaagg gacagaagat gacaggggcc 60
aggagggggc tggtg 75

SEQ ID NO: 203      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 203
atgtgactgg cacgggagtt gatcctggtt ttcacgtggc ccctgcacca gcagctcccta 60
caccggcggc ccctg 75

SEQ ID NO: 204      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 204
atgtgactgg cacgggagtt gatcctggtt ttcacggggg gagcagccctc tggcattctg 60
ggagcttcat ctgga 75

SEQ ID NO: 205      moltype = DNA  length = 75
FEATURE          Location/Qualifiers
source           1..75
                  mol_type = other DNA
                  organism = synthetic construct

```

---

-continued

---

```

SEQUENCE: 205
atgtgactgg cacgggagtt gatcctgggtt ttcacccccc gacgatattg aacaatggtt 60
cactgaagac ccagg 75

SEQ ID NO: 206      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 206
atgtgactgg cacgggagtt gatcctgggtt ttcacccccc gggctggggg gctgaggacc 60
tggtcctctg actgc 75

SEQ ID NO: 207      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 207
atgtgactgg cacgggagtt gatcctgggtt ttcacccccc gcaaccagcc ctgtcgcttc 60
tccagccccca gctgc 75

SEQ ID NO: 208      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 208
atgtgactgg cacgggagtt gatcctgggtt ttcacccatc gctatctgag cagcgctcat 60
ggtgccggca ggcgc 75

SEQ ID NO: 209      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 209
atgtgactgg cacgggagtt gatcctgggtt ttcacccatc ctacaaggcag tcacagcaca 60
tgacggaggt tgtga 75

SEQ ID NO: 210      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 210
atgtgactgg cacgggagtt gatcctgggtt ttcacccatc cgccggacgcg ggtgccggc 60
gggggtgtgg aatca 75

SEQ ID NO: 211      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 211
atgtgactgg cacgggagtt gatcctgggtt ttcacccatc caactggcca agacctgccc 60
tgtgcagctg tgggt 75

SEQ ID NO: 212      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 212
atgtgactgg cacgggagtt gatcctgggtt ttcacccatc tatctgttca cttgtccct 60
gacttcaac tctgt 75

SEQ ID NO: 213      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 213
atgtgactgg cacgggagtt gatcctgggtt ttcacccatc ccactgacaa ccaccctaa 60
ccctccctcc cagag 75

```

---

-continued

---

```

SEQ ID NO: 214      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 214
atgtgactgg cacgggagtt gatcctggtt ttcaccctca ggccgctcat agggcaccac 60
cacactatgt cgaaa 75

SEQ ID NO: 215      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 215
atgtgactgg cacgggagtt gatcctggtt ttcacaggaa atttgcgtgt ggagtatttg 60
gatgacagaaa acact 75

SEQ ID NO: 216      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 216
atgtgactgg cacgggagtt gatcctggtt ttcacccagg gtccccaggc ctctgattcc 60
tcactgattg ctctt 75

SEQ ID NO: 217      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 217
atgtgactgg cacgggagtt gatcctggtt ttcacgaggc aagcagaggc tggggcacag 60
caggccatgt tgca 75

SEQ ID NO: 218      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 218
atgtgactgg cacgggagtt gatcctggtt ttcacccctgg agtcttccag tgtgtatgtg 60
gtgaggatgg gcetc 75

SEQ ID NO: 219      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 219
atgtgactgg cacgggagtt gatcctggtt ttcacactac atgtgtaca gttcctgcat 60
gggcggcatg aaccg 75

SEQ ID NO: 220      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 220
atgtgactgg cacgggagtt gatcctggtt ttcaccttg c acaggcttc cccaaggcgc 60
actggcctca tcttg 75

SEQ ID NO: 221      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA
                  organism = synthetic construct
SEQUENCE: 221
atgtgactgg cacgggagtt gatcctggtt ttcacctgca cccttggct cctccaccgc 60
ttcttgctt gttt 75

SEQ ID NO: 222      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
                  mol_type = other DNA

```

---

-continued

---

```

SEQUENCE: 222          organism = synthetic construct
atgtgactgg cacggagtt gatcctgggtt ttcaccctcg ctttagtgctc cctggggca 60
gctcgtggta aggtc                                         75

SEQ ID NO: 223          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 223
atgtgactgg cacggagtt gatcctgggtt ttcacgaccg ggcacacagag gaagagaatc 60
tccgcaagaa agggg                                         75

SEQ ID NO: 224          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 224
atgtgactgg cacggagtt gatcctgggtt ttcactctcc caggacacaggc acaaacaacgc 60
acctcaaaggc tgttc                                         75

SEQ ID NO: 225          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 225
atgtgactgg cacggagtt gatcctgggtt ttcactctct tttcctatcc tgagtatgg 60
taatctactg ggacg                                         75

SEQ ID NO: 226          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 226
atgtgactgg cacggagtt gatcctgggtt ttcacggaca ggtaggacct gattcccta 60
ctgcctcttg ctctc                                         75

SEQ ID NO: 227          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 227
atgtgactgg cacggagtt gatcctgggtt ttcacggcat tttagtggtt agactggaaa 60
cttccactt gataa                                         75

SEQ ID NO: 228          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 228
atgtgactgg cacggagtt gatcctgggtt ttcaccctga agggtgaaat attctccatc 60
cagtggtttc ttctt                                         75

SEQ ID NO: 229          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 229
atgtgactgg cacggagtt gatcctgggtt ttcaccctag cactgccaa caacaccagc 60
tcctctcccc agcca                                         75

SEQ ID NO: 230          moltype = DNA length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 230
atgtgactgg cacggagtt gatcctgggtt ttcactgcct cagattcaact tttatcacct 60
ttcccttgcct ctttc                                         75

```

---

-continued

---

```

SEQ ID NO: 231      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 231
atgtgactgg cacgggagtt gatcctggtt ttcacatggc tttccaaacct aggaaggcag 60
gggagtaggg ccagg 75

SEQ ID NO: 232      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 232
atgtgactgg cacgggagtt gatcctggtt ttcacccctgg agtgagccct gctccccct 60
ggctccctcc cagcc 75

SEQ ID NO: 233      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 233
atgtgactgg cacgggagtt gatcctggtt ttcactccga gagctgaatg aggcccttga 60
actcaaggat gcca 75

SEQ ID NO: 234      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 234
atgtgactgg cacgggagtt gatcctggtt ttcaccatct tttaactcag gtactgtgta 60
tatacttact tctcc 75

SEQ ID NO: 235      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 235
atgtgactgg cacgggagtt gatcctggtt ttcacggcag gggagggaga gatgggggtg 60
ggaggctgtc agtgg 75

SEQ ID NO: 236      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 236
atgtgactgg cacgggagtt gatcctggtt ttcacgtcag tctgagtca gccccttgt 60
cttgaacatg agttt 75

SEQ ID NO: 237      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 237
atgtgactgg cacgggagtt gatcctggtt ttcacccctga agtccaaaaa gggtcagtct 60
acctcccgcc ataaa 75

SEQ ID NO: 238      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 238
atgtgactgg cacgggagtt gatcctggtt ttcacggcac agaccctctc actcatgtga 60
tgtcatctct cctcc 75

SEQ ID NO: 239      moltype = DNA  length = 75
FEATURE
source
1..75

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct

SEQUENCE: 239
atgtgactgg cacgggagtt gatcctgggtt ttcaccaggg gcttatgtgt ctcccttgatg 60
acctgcggcg acgtc 75

SEQ ID NO: 240      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 240
atgtgactgg cacgggagtt gatcctgggtt ttcaccatc atctccccc ttcccccttct 60
gcccaggctg ttgca 75

SEQ ID NO: 241      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 241
atgtgactgg cacgggagtt gatcctgggtt ttcacaattc tggcttctcc ctgctcacac 60
tttcttccat tgcat 75

SEQ ID NO: 242      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 242
atgtgactgg cacgggagtt gatcctgggtt ttcacagtca gtcgtgttgg caacatacca 60
tttcaacct ctgca 75

SEQ ID NO: 243      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 243
atgtgactgg cacgggagtt gatcctgggtt ttcactaattt tcttggcccc tcttcggtaa 60
ccctgagcca aatgt 75

SEQ ID NO: 244      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 244
atgtgactgg cacgggagtt gatcctgggtt ttcacgggtaa aataaaggaa gatacttagtt 60
ttgctgaaaa tgaca 75

SEQ ID NO: 245      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 245
atgtgactgg cacgggagtt gatcctgggtt ttcacactca tttgtatctg aagtggaaacc 60
aaatgataact gatcc 75

SEQ ID NO: 246      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 246
atgtgactgg cacgggagtt gatcctgggtt ttcacagttt agaccattca caggccaaag 60
acggtacaac ttcct 75

SEQ ID NO: 247      moltype = DNA length = 75
FEATURE           Location/Qualifiers
source            1..75
mol_type = other DNA
organism = synthetic construct

SEQUENCE: 247
atgtgactgg cacgggagtt gatcctgggtt ttcacgaaaa tgaatgctct gagctttgga 60

```

---

-continued

---

```

agctctcagg gtaca                                75
SEQ ID NO: 248      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 248
atgtgactgg cacgggagtt gatcctgggtt ttcacgggcc atcgcgtatgt cgcacggtag 60
ctgcgcgcgg ctgcg 75

SEQ ID NO: 249      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 249
atgtgactgg cacgggagtt gatcctgggtt ttcaccccca tccagcttca aaagctttc 60
gaatcattga tgtgc 75

SEQ ID NO: 250      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 250
atgtgactgg cacgggagtt gatcctgggtt ttcaactgcca agcctgaact acccctttt 60
tacactccta ttgat 75

SEQ ID NO: 251      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 251
atgtgactgg cacgggagtt gatcctgggtt ttcacccacc ctgactgtgc tctgtcccc 60
cagggctgga catcc 75

SEQ ID NO: 252      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 252
atgtgactgg cacgggagtt gatcctgggtt ttcacagtca ggagtggat gatctataaa 60
aactcgtaga aagag 75

SEQ ID NO: 253      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 253
atgtgactgg cacgggagtt gatcctgggtt ttcacacctcg gggagacaac gacggcggtg 60
gcgggagttt ctcca 75

SEQ ID NO: 254      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 254
atgtgactgg cacgggagtt gatcctgggtt ttcaccatac agtcctggat gatgtatgtt 60
ttgtatgaagg tctcg 75

SEQ ID NO: 255      moltype = DNA  length = 75
FEATURE
source
1..75
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 255
atgtgactgg cacgggagtt gatcctgggtt ttcactttta ctgttcttcc tcagacatcc 60
aaacgtgttt tgatc 75

SEQ ID NO: 256      moltype = DNA  length = 75
FEATURE

```

---

-continued

---

```

source          1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 256
atgtgactgg cacgggagtt gatcctgggtt ttcacgtgga agcatactgc aaaatatttg 60
tttcagtct ctgca                                         75

SEQ ID NO: 257      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 257
atgtgactgg cacgggagtt gatcctgggtt ttcacacgta cccctctcg cccctccct 60
tggactccag ccatg                                         75

SEQ ID NO: 258      moltype = DNA  length = 75
FEATURE           Location/Qualifiers
source            1..75
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 258
atgtgactgg cacgggagtt gatcctgggtt ttcacgtggc gtaagcgccg cacgccccgc 60
agtggteccc gtcct                                         75

SEQ ID NO: 259      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 259
nnnnnaagat ctttagtgca c                                         21

SEQ ID NO: 260      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 260
nnnnncgaca gaactattgc c                                         21

SEQ ID NO: 261      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 261
nnnnnactat ctttagtgca c                                         21

SEQ ID NO: 262      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 262
nnnnnctcca gaactattgc c                                         21

SEQ ID NO: 263      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 263
nnnnnagcat ctttagtgca c                                         21

SEQ ID NO: 264      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21
               mol_type = other DNA
               organism = synthetic construct
SEQUENCE: 264
nnnnncatca gaactattgc c                                         21

SEQ ID NO: 265      moltype = DNA  length = 21
FEATURE           Location/Qualifiers
source            1..21

```

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 265
nnnnnataat cttagtggca c 21

SEQ ID NO: 266      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 266
nnnnnaagaa ggttagaccct c 21

SEQ ID NO: 267      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 267
nnnnnnttct ctactcgtga c 21

SEQ ID NO: 268      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 268
nnnnnactaa ggttagaccct c 21

SEQ ID NO: 269      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 269
nnnnngaagc tacgagtatc c 21

SEQ ID NO: 270      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 270
nnnnnagcaa ggttagaccct c 21

SEQ ID NO: 271      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 271
nnnnncattg acgtctagag c 21

SEQ ID NO: 272      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 272
nnnnntcact ctactcgtga c 21

SEQ ID NO: 273      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 273
nnnnnataaa ggttagaccct c 21

SEQ ID NO: 274      moltype = DNA length = 21
FEATURE          Location/Qualifiers
source           1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 274
nnnnntacct ctactcgtga c 21

```

---

-continued

---

SEQ ID NO: 275	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 275	
nnnnnntttg tgtgtgtgt	20
SEQ ID NO: 276	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 276	
nnnnnactac acacacacac	20
SEQ ID NO: 277	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 277	
nnnnnnctcg gtgtgtgtgt	20
SEQ ID NO: 278	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 278	
nnnnnngaaca cacacacaca	20
SEQ ID NO: 279	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 279	
nnnnncatgt gtgtgtgtgt	20
SEQ ID NO: 280	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 280	
nnnnngtgc a c a c a c a c a	20
SEQ ID NO: 281	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 281	
nnnnnataac a c a c a c a c a c	20
SEQ ID NO: 282	moltype = DNA length = 20
FEATURE	Location/Qualifiers
source	1..20
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 282	
nnnnnntactg tgtgtgtgt	20
SEQ ID NO: 283	moltype = DNA length = 25
FEATURE	Location/Qualifiers
source	1..25
	mol_type = other DNA
	organism = synthetic construct
SEQUENCE: 283	
tgcaggacca gagaattcga ataca	25
SEQ ID NO: 284	moltype = DNA length = 58
FEATURE	Location/Qualifiers
source	1..58

---

-continued

---

```

mol_type = other DNA
organism = synthetic construct
SEQUENCE: 284
aatgatacgg cgaccaccga gatctacacg tcatgcagga ccagagaatt cgaataaca      58

SEQ ID NO: 285      moltype = DNA  length = 18
FEATURE
source          Location/Qualifiers
1..18
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 285
cggtggctca cgcctgtca                                         18

SEQ ID NO: 286      moltype = DNA  length = 18
FEATURE
source          Location/Qualifiers
1..18
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 286
gcctcggcct cccaaagt                                         18

SEQ ID NO: 287      moltype = DNA  length = 21
FEATURE
source          Location/Qualifiers
1..21
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 287
gaggctgagg caggagaatc g                                         21

SEQ ID NO: 288      moltype = DNA  length = 18
FEATURE
source          Location/Qualifiers
1..18
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 288
gtcgcccagg ctggagtg                                         18

SEQ ID NO: 289      moltype = DNA  length = 25
FEATURE
source          Location/Qualifiers
1..25
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 289
tgcaggacca gagaattcga ataca                                         25

SEQ ID NO: 290      moltype =   length =
SEQUENCE: 290
000

SEQ ID NO: 291      moltype =   length =
SEQUENCE: 291
000

SEQ ID NO: 292      moltype =   length =
SEQUENCE: 292
000

SEQ ID NO: 293      moltype =   length =
SEQUENCE: 293
000

SEQ ID NO: 294      moltype =   length =
SEQUENCE: 294
000

SEQ ID NO: 295      moltype =   length =
SEQUENCE: 295
000

SEQ ID NO: 296      moltype =   length =
SEQUENCE: 296
000

SEQ ID NO: 297      moltype =   length =
SEQUENCE: 297
000

```

---

-continued

---

SEQ ID NO: 298 moltype = length =  
SEQUENCE: 298  
000

SEQ ID NO: 299 moltype = length =  
SEQUENCE: 299  
000

SEQ ID NO: 300 moltype = length =  
SEQUENCE: 300  
000

SEQ ID NO: 301 moltype = length =  
SEQUENCE: 301  
000

SEQ ID NO: 302 moltype = length =  
SEQUENCE: 302  
000

SEQ ID NO: 303 moltype = length =  
SEQUENCE: 303  
000

SEQ ID NO: 304 moltype = length =  
SEQUENCE: 304  
000

SEQ ID NO: 305 moltype = length =  
SEQUENCE: 305  
000

SEQ ID NO: 306 moltype = length =  
SEQUENCE: 306  
000

SEQ ID NO: 307 moltype = length =  
SEQUENCE: 307  
000

SEQ ID NO: 308 moltype = length =  
SEQUENCE: 308  
000

SEQ ID NO: 309 moltype = length =  
SEQUENCE: 309  
000

SEQ ID NO: 310 moltype = length =  
SEQUENCE: 310  
000

SEQ ID NO: 311 moltype = length =  
SEQUENCE: 311  
000

SEQ ID NO: 312 moltype = length =  
SEQUENCE: 312  
000

SEQ ID NO: 313 moltype = length =  
SEQUENCE: 313  
000

SEQ ID NO: 314 moltype = length =  
SEQUENCE: 314  
000

SEQ ID NO: 315 moltype = length =  
SEQUENCE: 315  
000

SEQ ID NO: 316 moltype = length =  
SEQUENCE: 316  
000

---

-continued

---

```
SEQ ID NO: 317      moltype = length =
SEQUENCE: 317
000

SEQ ID NO: 318      moltype = length =
SEQUENCE: 318
000

SEQ ID NO: 319      moltype = length =
SEQUENCE: 319
000

SEQ ID NO: 320      moltype = length =
SEQUENCE: 320
000

SEQ ID NO: 321      moltype = length =
SEQUENCE: 321
000

SEQ ID NO: 322      moltype = length =
SEQUENCE: 322
000

SEQ ID NO: 323      moltype = length =
SEQUENCE: 323
000

SEQ ID NO: 324      moltype = length =
SEQUENCE: 324
000

SEQ ID NO: 325      moltype = length =
SEQUENCE: 325
000

SEQ ID NO: 326      moltype = length =
SEQUENCE: 326
000

SEQ ID NO: 327      moltype = length =
SEQUENCE: 327
000

SEQ ID NO: 328      moltype = length =
SEQUENCE: 328
000

SEQ ID NO: 329      moltype = length =
SEQUENCE: 329
000

SEQ ID NO: 330      moltype = length =
SEQUENCE: 330
000

SEQ ID NO: 331      moltype = length =
SEQUENCE: 331
000

SEQ ID NO: 332      moltype = length =
SEQUENCE: 332
000

SEQ ID NO: 333      moltype = length =
SEQUENCE: 333
000

SEQ ID NO: 334      moltype = length =
SEQUENCE: 334
000

SEQ ID NO: 335      moltype = length =
SEQUENCE: 335
000
```

---

-continued

---

SEQ ID NO: 336 moltype = length =  
SEQUENCE: 336  
000

SEQ ID NO: 337 moltype = length =  
SEQUENCE: 337  
000

SEQ ID NO: 338 moltype = length =  
SEQUENCE: 338  
000

SEQ ID NO: 339 moltype = length =  
SEQUENCE: 339  
000

SEQ ID NO: 340 moltype = length =  
SEQUENCE: 340  
000

SEQ ID NO: 341 moltype = length =  
SEQUENCE: 341  
000

SEQ ID NO: 342 moltype = length =  
SEQUENCE: 342  
000

SEQ ID NO: 343 moltype = length =  
SEQUENCE: 343  
000

SEQ ID NO: 344 moltype = length =  
SEQUENCE: 344  
000

SEQ ID NO: 345 moltype = length =  
SEQUENCE: 345  
000

SEQ ID NO: 346 moltype = length =  
SEQUENCE: 346  
000

SEQ ID NO: 347 moltype = length =  
SEQUENCE: 347  
000

SEQ ID NO: 348 moltype = length =  
SEQUENCE: 348  
000

SEQ ID NO: 349 moltype = length =  
SEQUENCE: 349  
000

SEQ ID NO: 350 moltype = length =  
SEQUENCE: 350  
000

SEQ ID NO: 351 moltype = length =  
SEQUENCE: 351  
000

SEQ ID NO: 352 moltype = length =  
SEQUENCE: 352  
000

SEQ ID NO: 353 moltype = length =  
SEQUENCE: 353  
000

SEQ ID NO: 354 moltype = length =  
SEQUENCE: 354  
000

---

-continued

---

```
SEQ ID NO: 355      moltype = length =
SEQUENCE: 355
000

SEQ ID NO: 356      moltype = length =
SEQUENCE: 356
000

SEQ ID NO: 357      moltype = length =
SEQUENCE: 357
000

SEQ ID NO: 358      moltype = length =
SEQUENCE: 358
000

SEQ ID NO: 359      moltype = length =
SEQUENCE: 359
000

SEQ ID NO: 360      moltype = length =
SEQUENCE: 360
000

SEQ ID NO: 361      moltype = length =
SEQUENCE: 361
000

SEQ ID NO: 362      moltype = length =
SEQUENCE: 362
000

SEQ ID NO: 363      moltype = length =
SEQUENCE: 363
000

SEQ ID NO: 364      moltype = length =
SEQUENCE: 364
000

SEQ ID NO: 365      moltype = length =
SEQUENCE: 365
000

SEQ ID NO: 366      moltype = length =
SEQUENCE: 366
000

SEQ ID NO: 367      moltype = length =
SEQUENCE: 367
000

SEQ ID NO: 368      moltype = length =
SEQUENCE: 368
000

SEQ ID NO: 369      moltype = length =
SEQUENCE: 369
000

SEQ ID NO: 370      moltype = length =
SEQUENCE: 370
000

SEQ ID NO: 371      moltype = length =
SEQUENCE: 371
000

SEQ ID NO: 372      moltype = length =
SEQUENCE: 372
000

SEQ ID NO: 373      moltype = length =
SEQUENCE: 373
000
```

---

-continued

---

SEQ ID NO: 374 moltype = length =  
SEQUENCE: 374  
000

SEQ ID NO: 375 moltype = length =  
SEQUENCE: 375  
000

SEQ ID NO: 376 moltype = length =  
SEQUENCE: 376  
000

SEQ ID NO: 377 moltype = length =  
SEQUENCE: 377  
000

SEQ ID NO: 378 moltype = length =  
SEQUENCE: 378  
000

SEQ ID NO: 379 moltype = length =  
SEQUENCE: 379  
000

SEQ ID NO: 380 moltype = length =  
SEQUENCE: 380  
000

SEQ ID NO: 381 moltype = length =  
SEQUENCE: 381  
000

SEQ ID NO: 382 moltype = length =  
SEQUENCE: 382  
000

SEQ ID NO: 383 moltype = length =  
SEQUENCE: 383  
000

SEQ ID NO: 384 moltype = length =  
SEQUENCE: 384  
000

SEQ ID NO: 385 moltype = length =  
SEQUENCE: 385  
000

SEQ ID NO: 386 moltype = length =  
SEQUENCE: 386  
000

SEQ ID NO: 387 moltype = length =  
SEQUENCE: 387  
000

SEQ ID NO: 388 moltype = length =  
SEQUENCE: 388  
000

SEQ ID NO: 389 moltype = length =  
SEQUENCE: 389  
000

SEQ ID NO: 390 moltype = length =  
SEQUENCE: 390  
000

SEQ ID NO: 391 moltype = length =  
SEQUENCE: 391  
000

SEQ ID NO: 392 moltype = length =  
SEQUENCE: 392  
000

---

-continued

---

```
SEQ ID NO: 393      moltype = length =
SEQUENCE: 393
000

SEQ ID NO: 394      moltype = length =
SEQUENCE: 394
000

SEQ ID NO: 395      moltype = length =
SEQUENCE: 395
000

SEQ ID NO: 396      moltype = length =
SEQUENCE: 396
000

SEQ ID NO: 397      moltype = length =
SEQUENCE: 397
000

SEQ ID NO: 398      moltype = length =
SEQUENCE: 398
000

SEQ ID NO: 399      moltype = length =
SEQUENCE: 399
000

SEQ ID NO: 400      moltype = length =
SEQUENCE: 400
000

SEQ ID NO: 401      moltype = length =
SEQUENCE: 401
000

SEQ ID NO: 402      moltype = length =
SEQUENCE: 402
000

SEQ ID NO: 403      moltype = length =
SEQUENCE: 403
000

SEQ ID NO: 404      moltype = length =
SEQUENCE: 404
000

SEQ ID NO: 405      moltype = length =
SEQUENCE: 405
000

SEQ ID NO: 406      moltype = length =
SEQUENCE: 406
000

SEQ ID NO: 407      moltype = length =
SEQUENCE: 407
000

SEQ ID NO: 408      moltype = length =
SEQUENCE: 408
000

SEQ ID NO: 409      moltype = length =
SEQUENCE: 409
000

SEQ ID NO: 410      moltype = length =
SEQUENCE: 410
000

SEQ ID NO: 411      moltype = length =
SEQUENCE: 411
000
```

---

-continued

---

```
SEQ ID NO: 412      moltype = length =
SEQUENCE: 412
000

SEQ ID NO: 413      moltype = length =
SEQUENCE: 413
000

SEQ ID NO: 414      moltype = length =
SEQUENCE: 414
000

SEQ ID NO: 415      moltype = length =
SEQUENCE: 415
000

SEQ ID NO: 416      moltype = length =
SEQUENCE: 416
000

SEQ ID NO: 417      moltype = length =
SEQUENCE: 417
000

SEQ ID NO: 418      moltype = length =
SEQUENCE: 418
000

SEQ ID NO: 419      moltype = length =
SEQUENCE: 419
000

SEQ ID NO: 420      moltype = length =
SEQUENCE: 420
000

SEQ ID NO: 421      moltype = length =
SEQUENCE: 421
000

SEQ ID NO: 422      moltype = length =
SEQUENCE: 422
000

SEQ ID NO: 423      moltype = length =
SEQUENCE: 423
000

SEQ ID NO: 424      moltype = length =
SEQUENCE: 424
000

SEQ ID NO: 425      moltype = length =
SEQUENCE: 425
000

SEQ ID NO: 426      moltype = length =
SEQUENCE: 426
000

SEQ ID NO: 427      moltype = length =
SEQUENCE: 427
000

SEQ ID NO: 428      moltype = length =
SEQUENCE: 428
000

SEQ ID NO: 429      moltype = length =
SEQUENCE: 429
000

SEQ ID NO: 430      moltype = length =
SEQUENCE: 430
000
```

---

-continued

---

```
SEQ ID NO: 431      moltype = length =
SEQUENCE: 431
000

SEQ ID NO: 432      moltype = length =
SEQUENCE: 432
000

SEQ ID NO: 433      moltype = length =
SEQUENCE: 433
000

SEQ ID NO: 434      moltype = length =
SEQUENCE: 434
000

SEQ ID NO: 435      moltype = length =
SEQUENCE: 435
000

SEQ ID NO: 436      moltype = length =
SEQUENCE: 436
000

SEQ ID NO: 437      moltype = length =
SEQUENCE: 437
000

SEQ ID NO: 438      moltype = length =
SEQUENCE: 438
000

SEQ ID NO: 439      moltype = length =
SEQUENCE: 439
000

SEQ ID NO: 440      moltype = length =
SEQUENCE: 440
000

SEQ ID NO: 441      moltype = length =
SEQUENCE: 441
000

SEQ ID NO: 442      moltype = length =
SEQUENCE: 442
000

SEQ ID NO: 443      moltype = length =
SEQUENCE: 443
000

SEQ ID NO: 444      moltype = length =
SEQUENCE: 444
000

SEQ ID NO: 445      moltype = length =
SEQUENCE: 445
000

SEQ ID NO: 446      moltype = length =
SEQUENCE: 446
000

SEQ ID NO: 447      moltype = length =
SEQUENCE: 447
000

SEQ ID NO: 448      moltype = length =
SEQUENCE: 448
000

SEQ ID NO: 449      moltype = length =
SEQUENCE: 449
000
```

---

-continued

---

```
SEQ ID NO: 450      moltype = length =
SEQUENCE: 450
000

SEQ ID NO: 451      moltype = length =
SEQUENCE: 451
000

SEQ ID NO: 452      moltype = length =
SEQUENCE: 452
000

SEQ ID NO: 453      moltype = length =
SEQUENCE: 453
000

SEQ ID NO: 454      moltype = length =
SEQUENCE: 454
000

SEQ ID NO: 455      moltype = length =
SEQUENCE: 455
000

SEQ ID NO: 456      moltype = length =
SEQUENCE: 456
000

SEQ ID NO: 457      moltype = length =
SEQUENCE: 457
000

SEQ ID NO: 458      moltype = length =
SEQUENCE: 458
000

SEQ ID NO: 459      moltype = length =
SEQUENCE: 459
000

SEQ ID NO: 460      moltype = length =
SEQUENCE: 460
000

SEQ ID NO: 461      moltype = length =
SEQUENCE: 461
000

SEQ ID NO: 462      moltype = length =
SEQUENCE: 462
000

SEQ ID NO: 463      moltype = length =
SEQUENCE: 463
000

SEQ ID NO: 464      moltype = length =
SEQUENCE: 464
000

SEQ ID NO: 465      moltype = length =
SEQUENCE: 465
000

SEQ ID NO: 466      moltype = length =
SEQUENCE: 466
000

SEQ ID NO: 467      moltype = length =
SEQUENCE: 467
000

SEQ ID NO: 468      moltype = length =
SEQUENCE: 468
000
```

---

-continued

---

```
SEQ ID NO: 469      moltype = length =
SEQUENCE: 469
000

SEQ ID NO: 470      moltype = length =
SEQUENCE: 470
000

SEQ ID NO: 471      moltype = length =
SEQUENCE: 471
000

SEQ ID NO: 472      moltype = length =
SEQUENCE: 472
000

SEQ ID NO: 473      moltype = length =
SEQUENCE: 473
000

SEQ ID NO: 474      moltype = length =
SEQUENCE: 474
000

SEQ ID NO: 475      moltype = length =
SEQUENCE: 475
000

SEQ ID NO: 476      moltype = length =
SEQUENCE: 476
000

SEQ ID NO: 477      moltype = length =
SEQUENCE: 477
000

SEQ ID NO: 478      moltype = length =
SEQUENCE: 478
000

SEQ ID NO: 479      moltype = length =
SEQUENCE: 479
000

SEQ ID NO: 480      moltype = length =
SEQUENCE: 480
000

SEQ ID NO: 481      moltype = length =
SEQUENCE: 481
000

SEQ ID NO: 482      moltype = length =
SEQUENCE: 482
000

SEQ ID NO: 483      moltype = length =
SEQUENCE: 483
000

SEQ ID NO: 484      moltype = length =
SEQUENCE: 484
000

SEQ ID NO: 485      moltype = length =
SEQUENCE: 485
000

SEQ ID NO: 486      moltype = length =
SEQUENCE: 486
000

SEQ ID NO: 487      moltype = length =
SEQUENCE: 487
000
```

---

-continued

---

SEQ ID NO: 488 moltype = length =  
SEQUENCE: 488  
000

SEQ ID NO: 489 moltype = length =  
SEQUENCE: 489  
000

SEQ ID NO: 490 moltype = length =  
SEQUENCE: 490  
000

SEQ ID NO: 491 moltype = length =  
SEQUENCE: 491  
000

SEQ ID NO: 492 moltype = length =  
SEQUENCE: 492  
000

SEQ ID NO: 493 moltype = length =  
SEQUENCE: 493  
000

SEQ ID NO: 494 moltype = length =  
SEQUENCE: 494  
000

SEQ ID NO: 495 moltype = length =  
SEQUENCE: 495  
000

SEQ ID NO: 496 moltype = length =  
SEQUENCE: 496  
000

SEQ ID NO: 497 moltype = length =  
SEQUENCE: 497  
000

SEQ ID NO: 498 moltype = length =  
SEQUENCE: 498  
000

SEQ ID NO: 499 moltype = length =  
SEQUENCE: 499  
000

SEQ ID NO: 500 moltype = length =  
SEQUENCE: 500  
000

SEQ ID NO: 501 moltype = length =  
SEQUENCE: 501  
000

SEQ ID NO: 502 moltype = length =  
SEQUENCE: 502  
000

SEQ ID NO: 503 moltype = length =  
SEQUENCE: 503  
000

SEQ ID NO: 504 moltype = length =  
SEQUENCE: 504  
000

SEQ ID NO: 505 moltype = length =  
SEQUENCE: 505  
000

SEQ ID NO: 506 moltype = length =  
SEQUENCE: 506  
000

---

-continued

---

```
SEQ ID NO: 507      moltype = length =
SEQUENCE: 507
000

SEQ ID NO: 508      moltype = length =
SEQUENCE: 508
000

SEQ ID NO: 509      moltype = length =
SEQUENCE: 509
000

SEQ ID NO: 510      moltype = length =
SEQUENCE: 510
000

SEQ ID NO: 511      moltype = length =
SEQUENCE: 511
000

SEQ ID NO: 512      moltype = length =
SEQUENCE: 512
000

SEQ ID NO: 513      moltype = length =
SEQUENCE: 513
000

SEQ ID NO: 514      moltype = length =
SEQUENCE: 514
000

SEQ ID NO: 515      moltype = length =
SEQUENCE: 515
000

SEQ ID NO: 516      moltype = length =
SEQUENCE: 516
000

SEQ ID NO: 517      moltype = length =
SEQUENCE: 517
000

SEQ ID NO: 518      moltype = length =
SEQUENCE: 518
000

SEQ ID NO: 519      moltype = length =
SEQUENCE: 519
000

SEQ ID NO: 520      moltype = length =
SEQUENCE: 520
000

SEQ ID NO: 521      moltype = length =
SEQUENCE: 521
000

SEQ ID NO: 522      moltype = length =
SEQUENCE: 522
000

SEQ ID NO: 523      moltype = length =
SEQUENCE: 523
000

SEQ ID NO: 524      moltype = length =
SEQUENCE: 524
000

SEQ ID NO: 525      moltype = length =
SEQUENCE: 525
000
```

---

-continued

---

```
SEQ ID NO: 526      moltype =  length =
SEQUENCE: 526
000

SEQ ID NO: 527      moltype =  length =
SEQUENCE: 527
000

SEQ ID NO: 528      moltype =  length =
SEQUENCE: 528
000

SEQ ID NO: 529      moltype =  length =
SEQUENCE: 529
000

SEQ ID NO: 530      moltype =  length =
SEQUENCE: 530
000

SEQ ID NO: 531      moltype =  length =
SEQUENCE: 531
000

SEQ ID NO: 532      moltype =  length =
SEQUENCE: 532
000

SEQ ID NO: 533      moltype =  length =
SEQUENCE: 533
000

SEQ ID NO: 534      moltype =  length =
SEQUENCE: 534
000

SEQ ID NO: 535      moltype =  length =
SEQUENCE: 535
000

SEQ ID NO: 536      moltype =  length =
SEQUENCE: 536
000

SEQ ID NO: 537      moltype =  length =
SEQUENCE: 537
000

SEQ ID NO: 538      moltype =  length =
SEQUENCE: 538
000

SEQ ID NO: 539      moltype =  length =
SEQUENCE: 539
000

SEQ ID NO: 540      moltype =  length =
SEQUENCE: 540
000

SEQ ID NO: 541      moltype =  length =
SEQUENCE: 541
000

SEQ ID NO: 542      moltype =  length =
SEQUENCE: 542
000

SEQ ID NO: 543      moltype =  length =
SEQUENCE: 543
000

SEQ ID NO: 544      moltype =  length =
SEQUENCE: 544
000
```

---

-continued

---

```

SEQ ID NO: 545      moltype = length =
SEQUENCE: 545
000

SEQ ID NO: 546      moltype = length =
SEQUENCE: 546
000

SEQ ID NO: 547      moltype = length =
SEQUENCE: 547
000

SEQ ID NO: 548      moltype = length =
SEQUENCE: 548
000

SEQ ID NO: 549      moltype = length =
SEQUENCE: 549
000

SEQ ID NO: 550      moltype = length =
SEQUENCE: 550
000

SEQ ID NO: 551      moltype = length =
SEQUENCE: 551
000

SEQ ID NO: 552      moltype = length =
SEQUENCE: 552
000

SEQ ID NO: 553      moltype = length =
SEQUENCE: 553
000

SEQ ID NO: 554      moltype = length =
SEQUENCE: 554
000

SEQ ID NO: 555      moltype = DNA length = 12
FEATURE
source
1..12
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 555
acccacacca aa
12

SEQ ID NO: 556      moltype = DNA length = 12
FEATURE
source
1..12
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 556
caaacacaac cc
12

SEQ ID NO: 557      moltype = DNA length = 12
FEATURE
source
1..12
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 557
gttgtgggtt tt
12

SEQ ID NO: 558      moltype = DNA length = 12
FEATURE
source
1..12
mol_type = other DNA
organism = synthetic construct
SEQUENCE: 558
tgtgttttgt gg
12

SEQ ID NO: 559      moltype = DNA length = 12
FEATURE
source
1..12

```

-continued

---

SEQUENCE: 559	mol_type = other DNA organism = synthetic construct	
tttggtgtgg gt		12
SEQ ID NO: 560	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
SEQUENCE: 560	mol_type = other DNA organism = synthetic construct	
gggttgtgtt tg		12
SEQ ID NO: 561	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
SEQUENCE: 561	mol_type = other DNA organism = synthetic construct	
aacaacccac ac		12
SEQ ID NO: 562	moltype = DNA length = 12	
FEATURE	Location/Qualifiers	
source	1..12	
SEQUENCE: 562	mol_type = other DNA organism = synthetic construct	
ccaccaaaca ca		12

---

**1-43.** (canceled)

**44.** A method for targeted genetic analysis of circulating cell-free DNA (cfDNA), said method comprising:

(a) hybridizing a plurality of capture probe modules to a target genetic locus in a tagged cfDNA library to form a plurality of tagged cfDNA-capture probe module complexes,

wherein the plurality of capture probe modules are not immobilized to a surface,

wherein each capture probe module comprises a capture probe sequence that hybridizes to a specific DNA target region,

wherein each of the plurality of capture probe modules hybridizes to the target genetic locus within about 200 bp of any other capture probe module, and

the plurality of capture probe modules comprises capture probe modules that hybridize to regions of redundant sequence or regions that exhibit extreme base composition bias in the target genetic locus;

(b) isolating the plurality of tagged cfDNA-capture probe module complexes from (a) to form a plurality of isolated tagged cfDNA-capture probe module complexes, each isolated tagged cfDNA-capture probe module complex comprising a tagged cfDNA molecule and a capture probe module; and

(c) for the plurality of isolated tagged cfDNA-capture probe module complexes from (b), performing 5' to 3' DNA polymerase extension of the capture probe module using the isolated tagged cfDNA molecule as a template.

**45.** The method of claim 44, wherein the targeted genetic analysis is performed on a plurality of target genetic loci.

**46.** The method of claim 44, wherein the capture probe sequences are about 25 nucleotides to about 45 nucleotides in length.

**47.** The method of claim 44, wherein the capture probe sequences are about 35 nucleotides to about 45 nucleotides in length.

**48.** The method of claim 44, wherein the capture probe sequences are 40 nucleotides in length.

**49.** The method of claim 44, further comprising determining a number of genome equivalents in the tagged cfDNA library.

**50.** The method of claim 44, further comprising one or more steps of:

amplifying the isolated target genetic locus;  
sequencing the amplified isolated target genetic locus; and  
performing bioinformatic analysis of the resulting sequence reads.

**51.** The method of claim 44, further comprising:  
amplifying the isolated target genetic locus;  
sequencing the amplified isolated target genetic locus; and  
performing bioinformatic analysis of the resulting sequence reads.

**52.** The method of claim 51, wherein the sequencing comprises:

obtaining one or more tagged cfDNA library clones, each clone comprising a first DNA sequence and a second DNA sequence, wherein the first DNA sequence comprises a sequence in the target genetic locus, and the second DNA sequence comprises a capture probe sequence;

performing a paired end sequencing reaction on the one or more tagged cfDNA library clones and obtaining one or more sequencing reads sufficient to identify both the first DNA sequence and the second DNA sequence; and  
ordering or clustering the one or more sequencing reads according to the capture probe sequences of the one or more sequencing reads.

**53.** The method of claim 51, wherein the bioinformatics analysis is used to detect one or more genetic lesions within the target genetic locus.

**54.** The method of claim **53**, wherein the one or more genetic lesions comprise a nucleotide transition or transversion, a nucleotide insertion or deletion, a genomic rearrangement, or a change in copy number.

**55.** The method of claim **53**, wherein the one or more genetic lesions comprise chromosomal rearrangements.

**56.** The method of claim **53**, wherein the one or more genetic lesions comprise a genomic rearrangement that fuses a 3' coding region of an ALK gene to another gene.

**57.** The method of claim **56**, wherein the 3' coding region of the ALK gene is fused to an EML4 gene.

**58.** The method of claim **44**, wherein each capture probe module comprises a capture probe sequence at a 3' end of the capture probe module and a tail sequence at a 5' end of the capture probe module.

**59.** The method of claim **58**, wherein:

the tail sequence is capable of hybridizing to a partner oligonucleotide comprising at its 3' end a first member of a binding pair to allow for isolation and/or purification of a tagged cfDNA-capture probe module complex; and

the isolating comprises binding the first member of the binding pair of a partner oligonucleotide hybridized to a tagged cfDNA-capture probe module complex to a second member of the binding pair.

**60.** The method of claim **59**, wherein the first member of the binding pair comprises biotin, and the second member of the binding pair comprises streptavidin.

**61.** The method of claim **59**, wherein the partner oligonucleotide is hybridized to the tail sequence prior to the isolating in (b).

**62.** The method of claim **60**, wherein the partner oligonucleotide is hybridized to the tail sequence prior to the isolating in (b).

**63.** The method of claim **44**, wherein a plurality of capture probe modules hybridize to the target genetic locus, and each of the plurality of capture probe modules hybridizes to the target genetic locus within about 100 bp of any other capture probe module.

\* \* \* \* \*