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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
21 June 2018 (21.06.2018)

W!P O | PCT

(10) International Publication Number
WO 2018/111835 Al

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12P 19/34* (2006.01)
C12N 15/11 (2006.01) *C12Q 1/68* (2018.01)

(21) International Application Number:
PCT/US2017/065747

(22) International Filing Date:
12 December 2017 (12.12.2017)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/433,071	12 December 2016 (12.12.2016)	US
62/502,128	05 May 2017 (05.05.2017)	US
62/522,856	21 June 2017 (21.06.2017)	US

(71) Applicant: **DANA-FARBER CANCER INSTITUTE, INC.** [US/US]; 450 Brookline Avenue, Boston, MA 02215 (US).

(72) Inventor: **MAKRIGIORGOS, Gerassimos;** 63 Payson Road, Chestnut Hill, MA 02467 (US).

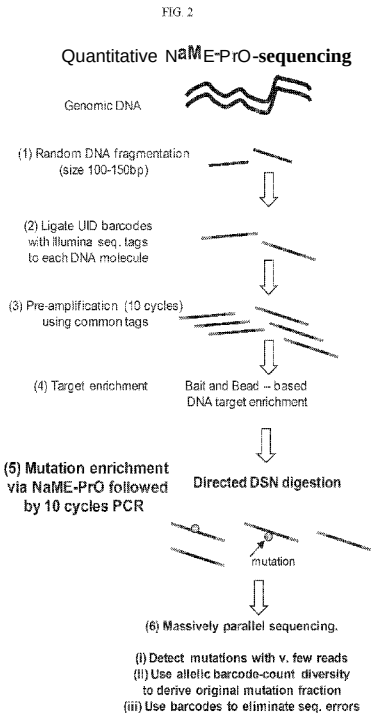
(74) Agent: **GATES, Edward, R.;** Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*

(54) Title: COMPOSITIONS AND METHODS FOR MOLECULAR BARCODING OF DNA MOLECULES PRIOR TO MUTATION ENRICHMENT AND/OR MUTATION DETECTION



(57) Abstract: Provided herein are methods to determine the original abundance of mutant alleles of one or more barcoded target sequences following mutation enrichment and sequencing.

WO 2018/111835 A1

COMPOSITIONS AND METHODS FOR MOLECULAR BARCODING OF DNA MOLECULES PRIOR TO MUTATION ENRICHMENT AND/OR MUTATION DETECTION

RELATED APPLICATIONS

This application claims the benefit of U.S. provisional application No. 62/433,071, filed on December 12, 2016, U.S. provisional application No. 62/502,128, filed on May 5, 2017, and U.S. provisional application No. 62/522856, filed on June 21, 2017, the entire disclosures of which are incorporated by reference herein.

BACKGROUND OF THE INVENTION

Low-level tumor somatic DNA mutations can have profound implications for development of metastasis, prognosis, choice of treatment, follow-up or early cancer detection. Unless effectively detected, these low-level mutations can misinform patient management decisions or become missed opportunities for personalized medicine. Next generation sequencing (NGS) technologies reveal prevalent somatic mutations, yet they 'lose steam' when it comes to detecting low-level DNA mutations in tumors with clonal heterogeneity, or in bodily fluids during 'liquid biopsy', and their integration with clinical practice is not straightforward. For mutations at an abundance of ~2-5% or less, both NGS and the methods required to prepare DNA for NGS generate false positives ('noise') independent of sequencing depth and hinder personalized clinical decisions based on mutational profiling. Recent enhancements employing single molecule barcoding enable NGS to reduce noise and detect 'ultra-rare mutations'. However, these approaches invariably diminish NGS throughput capability and increase expense as they require numerous reads per sequence (e.g. > 5,000 - 20,000x Mean Coverage).

SUMMARY OF THE INVENTION

The present disclosure relates to novel methods for retaining both throughput breadth and sequencing depth. Accordingly, some aspects of the disclosure provide a method of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA. In some embodiments, the method comprises: providing a sample of double-stranded genomic DNA fragments anticipated to comprise mutant alleles of target nucleic acid and wild-type

alleles of the target nucleic acid, wherein each terminus of the genomic DNA fragments is attached to a unique double stranded barcode and a double stranded common sequence tag, wherein the common sequence tag is located upstream of the unique barcode; amplifying a portion of the double-stranded genomic DNA fragments using primers that are complementary to the common sequence tag; enriching in the amplified portion the mutant alleles of the target nucleic acids relative to the wild-type alleles of the target nucleic acids; obtaining a measure of total number of unique barcodes associated with mutant target nucleic acid sequence in the enriched portion; and, obtaining a measure of total number of unique barcodes associated with the sample. The original abundance of the mutant target nucleic acid in the sample of genomic DNA is represented by the ratio of the total number of unique barcodes associated with mutant target nucleic acid sequence to the total number of unique barcodes associated with the sample.

In some embodiments, the total number of unique barcodes associated with the sample is the total number of unique barcodes associated with mutant and wild-type target nucleic acid sequence in the enriched portion

In some embodiments when the mutant alleles are enriched relative to the wild-type alleles, wild-type alleles become difficult to detect. In some embodiments, to provide additional information towards quantification of the original mutation abundance, in addition to applying mutation enrichment, a control sample is treated without mutation enrichment, as an un-enriched control sample in parallel. The combined information from both samples (mutation enriched sample and un-enriched control sample) provide additional information as compared to only the mutation enriched sample alone. For example, the unenriched control sample may provide a more accurate estimate of uniquely-barcoded wild-type alleles, while the mutation enriched sample may provide the most accurate estimate of uniquely-barcoded mutant alleles. By combining the information from both enriched sample and un-enriched control sample, it may be possible to more accurately quantify original mutation abundance than by screening either the mutant enriched sample or the un-enriched control sample alone.

Accordingly, a method of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA further comprises obtaining a measure of total number of unique barcodes associated with wild-type target nucleic acid sequence in an un-enriched control sample, wherein the total number of unique barcodes associated with the sample is the sum of the total number of unique barcodes associated with mutant target nucleic acid sequence

in the enriched portion and the total number of unique barcodes associated with wild-type target nucleic acid sequence in an un-enriched control sample. In some embodiments, a method of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA further comprises amplifying the double-stranded genomic DNA fragments in the un-enriched control sample using primers that are complementary to common sequence tags prior to obtaining a measure of total number of unique barcodes associated with wild-type target nucleic acid sequence in the un-enriched control sample. Therefore, in some embodiments, an un-enriched control sample is an aliquot of the sample that is processed in parallel to the sample that is enriched for mutant alleles of the target nucleic acids relative to the wild-type alleles of the target nucleic acids. Accordingly, in some embodiments, double-stranded genomic DNA fragments in a sample and unenriched control sample are attached to unique double-stranded barcodes in separate reactions. In some embodiments, double-stranded genomic DNA fragments in a sample and unenriched control sample are amplified using primers that are complementary to a common sequence tag in separate reactions.

In some embodiments, an un-enriched control sample is an aliquot of an amplified portion.

In some embodiments, the number of sequencing reads used to obtain the total number of unique barcodes associated with mutant target nucleic acid sequence in the enriched portion is less than or equal to 1 million. In some embodiments, the number of sequencing reads used to obtain the total number of unique barcodes associated with wild-type target nucleic acid sequence in the un-enriched control sample is less than or equal to 1 million. In some embodiments, the number of sequencing reads used to obtain the total number of unique barcodes associated with mutant target nucleic acid sequence in the enriched portion is equal to the number of sequencing reads used to obtain the total number of unique barcodes associated with wild-type target nucleic acid sequence in the un-enriched control sample. In some embodiments, the enriched portion and the un-enriched control sample are sequenced in the same sequencing reaction. Sample indices can be used to identify the origin of DNA fragments when the enriched portion and the amplified control sample are sequenced in the same sequencing reaction. Accordingly, in some embodiments, before sequencing, the DNA fragments of the enriched portion are attached to a first sample-index and the DNA fragments of the un-enriched amplified control sample are attached to a second sample-index, wherein the first sample-index

is different from the second sample-index. In some embodiments, the sample is divided prior to amplification and a first sample index is attached to a first portion of the sample which will be later amplified and enriched and a second sample index is provided to a second portion of the sample will is the control sample, prior to sequencing (with or without amplification).

In some embodiments, the enriched portion and the un-enriched control sample are sequenced in separate sequencing reactions.

In some embodiments, a method of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA involves target enrichment. Accordingly, some embodiments of any one of the methods of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA further comprises enriching both the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample. In some embodiments, enriching the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample comprises contacting the sample with one or more probes that specifically bind target nucleic acids. In some embodiments, one or more probes are bound to a solid substrate. In some embodiments, the solid substrate comprises beads.

In some embodiments, enrichment of mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample is performed by multiplex PCR. The present invention describes novel methods for performing multi-stage multiplex PCR optionally incorporating molecular barcodes at one of the stages, and optionally performing mutation enrichment, and producing a sequencing-ready sample.

In some embodiments, the enriching the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample involves amplifying the double-stranded genomic DNA fragments in the sample using a first and second primer. In some embodiments, the first primer is complementary to the double stranded common sequence tag, and the second primer is a nested gene-specific primer comprising a sequence tag that is different from the double-stranded common sequence tag. In some embodiments, the sequence tag on a second primer that is a nested gene-specific primer is common for the same target, but different for different targets. In some embodiments, the enriching the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids is performed after the amplifying of the

double-stranded genomic DNA fragments using primers that are complementary to the common sequence tag. In some embodiments, the amplifying the double-stranded genomic DNA fragments in the sample using a first and second primer is performed for 5-15 cycles. In some embodiments, the first primer comprises a sequencing adapter (e.g., an Illumina adapter).

Some embodiments of any one of the methods of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA further comprises fragmenting genomic DNA to form the double-stranded genomic DNA fragments. In some embodiments, the genomic DNA is fragmented using Shearase or any other DNA-ase type of enzyme that digests DNA randomly. In some embodiments, enzymatic end-repairing of the DNA fragments to obtain blunt ended 5' containing double-stranded DNA is done after fragmentation (e.g., by an enzyme).

In some embodiments, any one of the methods disclosed herein does not include obtaining a measure of total number of barcodes that are associated with non-target mutant nucleic acids or non-target wild-type nucleic acids in the enriched portion.

In some embodiments, the method further comprises obtaining sequences of the mutant target nucleic acids and wild-type target nucleic acids in the enriched sample.

In some embodiments of any one of the methods of determining original abundance of mutant alleles of target nucleic acid as disclosed herein, the double-stranded genomic DNA fragments are 40-200bp in length. In some embodiments, the unique barcode is 8-14bp in length. In some embodiments, the common tag is 16-40 bp in length.

Some embodiments of any one of the methods of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA further comprises attaching the double-stranded genomic DNA fragments to the unique double-stranded barcodes. In some embodiments, attaching of double-stranded fragments of genomic DNA to the unique barcodes is performed at a ratio of 10^7-10^9 unique barcodes to lOOng double-stranded genomic DNA fragments present in the sample. In some embodiments, wherein the attaching comprises starting from a single-stranded barcode, synthesizing the opposite strand of the single-stranded barcode using an extension reaction to form a double stranded barcode, and then ligating to an end of the double-stranded genomic DNA fragment. In some embodiments, the attaching of barcodes comprises PCR using oligonucleotide primers, wherein each oligonucleotide primer comprises a

common tag portion, a unique barcode portion, and a target-specific portion. In some embodiments, 1-6 cycles of PCR are used to attach barcodes using PCR.

In some embodiments of any one of the method of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA, the barcoded DNA fragments are amplified using PCR or COLD-PCR. In some embodiments, 3-30 cycles of PCR or COLD-PCR are performed for amplifying the barcoded DNA fragments.

Any one of the methods disclosed herein can be performed on a solid substrate. Accordingly, in some embodiments of any one of the methods disclosed herein, the common tag is biotinylated. In some embodiments, the double-stranded genomic DNA fragments attached to barcodes and common tags are bound to streptavidin-beads.

In some embodiments of any one of the methods disclosed herein, the genomic DNA is obtained from a biological sample. In some embodiments, the biological sample is selected from tissue, blood, plasma, serum, urine, saliva, and cerebrospinal fluid. In some embodiments, the tissue sample is a tumor biopsy. In some embodiments, the biological sample comprises circulating cell-free DNA.

In some embodiments of any one of the methods of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA, following incorporation of molecular barcodes to DNA molecules, the mutant target nucleic acid in the amplified portion is enriched by 2-300 fold relative to the wild-type target nucleic acid in the amplified portion. In some embodiments, the enriching the mutant alleles of the target nucleic acids in the amplified portion relative to the wild-type alleles of the target nucleic acids in the amplified portion comprises subjecting the amplified portion to one or more of the following: Nuclease-assisted Minor-allele Enrichment using Probe Overlap (NaME-PrO), Coamplification at Lower Denaturation temperature-PCR (COLD-PCR), Improved and Complete Enrichment COLD-PCR (ice-COLD-PCR), Temperature-Tolerant-ice-COLD-PCR (TT-ice-COLD-PCR), toehold PCR, blocker-PCR using locked nucleic acids (LNA), peptide nucleic acids (PNA or XNA), or blockers with other modified nucleotides, CRISP-R-mediated mutation enrichment by selecting mutated alleles or removing wild-type alleles and Differential Strand Separation at Critical Temperature (DiSSECT).

In some embodiments, the NaMe-PrO comprises

(a) preparing an amplification reaction mixture comprising the double-stranded mutant and wild-type target nucleic acids, a thermostable double strand-specific nuclease (DSN), PCR amplification components, and a pair of oligonucleotide probes, one of which is complementary to the wild-type nucleic acid top strand and the other is complementary to the wild-type nucleic acid bottom strand, wherein the probes overlap each other by 10-15 probes such that the overlap coincides with the target region;

(b) subjecting the reaction mixture to a denaturing temperature to permit denaturation of the wild-type nucleic acid and the mutant target nucleic acid;

(c) reducing the temperature to permit hybridization of the probes to their corresponding sequences on the wild-type and mutant target nucleic acids thereby forming complementary wild-type-probe duplexes, wherein the DSN cleaves the complementary wild-type-probe duplexes but not the partially complementary target mutant -probe duplexes; and

(d) subjecting the reaction mixture to an amplification condition thereby enriching the uncleaved mutant target nucleic acid relative to the cleaved wild-type nucleic acid.

In some embodiments of any of the methods that utilize NaMe-Pro, the target region that coincides with the overlap of the probes comprises more than one mutation. In some embodiments, the probes have a 3'-terminal polymerase block. In some embodiments, to avoid identifying SNPs in the genome around the mutation site, the NaMe-Pro probes are designed so that they are complementary to SNPs near target mutations.

In some embodiments, the COLD-PCR comprises

(a) denaturing the double-stranded mutant and wild-type target nucleic acids by subjecting the double-stranded target mutant and wild-type nucleic acids to a first denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(b) forming a target mutant/wild-type strand duplex;

(c) denaturing said mutant/wild-type strand duplex by subjecting the nucleic acid amplified portion to a critical temperature (Tc) that is below the Tm of the wild-type nucleic acids;

(d) annealing a primer pair to the mutant and wild-type target nucleic acid strands; and

(e) extending said primer pair so as to enrich said mutant target sequence relative to said wild-type strand.

In some embodiments of methods disclosed herein that utilize COLD-PCR, (b) and (c) of COLD-PCR (as described above) are repeated 1-19 times before performing (d). In some embodiments, (a) to (e) are repeated 1-29 times.

In some embodiments, ice-COLD-PCR comprises

(a) exposing the mutant and wild-type target nucleic acids to a reference sequence that is complementary the target sequence;

(b) denaturing the double-stranded target mutant and wild-type nucleic acids by subjecting the double-stranded mutant and wild-type target nucleic acids to a first denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(c) forming a target mutant/reference strand and target wild-type/reference strand duplexes;

(d) denaturing said mutant/reference strand duplex by subjecting the nucleic acid sample to a critical temperature (Tc) that is below the Tm of the wild-type/reference duplex;

(d) annealing a primer pair to the mutant and wild-type target nucleic acid strands; and

(e) extending said primer pair so as to enrich said mutant target sequence relative to said wild-type target nucleic acid.

In some embodiments, TT-ice-COLD-PCR comprises

(a) exposing the mutant and wild-type target nucleic acids to a reference sequence that is complementary the target sequence;

(b) denaturing the double-stranded target mutant and wild-type target nucleic acids by subjecting the double-stranded target mutant and wild-type nucleic acids to a denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(c) forming a target mutant/reference strand and target wild-type/reference strand duplexes;

(d) denaturing said mutant/reference strand duplex by subjecting the nucleic acid sample to a first critical temperature (Tc) that is below the Tm of the wild-type/reference duplex;

(e) annealing a primer pair to the mutant and wild-type target nucleic acid strands;

(f) extending said primer pair so as to enrich said mutant target nucleic acid relative to said wild-type target nucleic acid; and

(f) repeating steps (d) to (f) at least once at a second Tc which is above the first Tc.

In some embodiments, DiSSECT comprises:

(a) allowing mutant and wildtype target nucleic acids to bind to complementary probes which are immobilized to beads, wherein the probes resemble the wild-type nucleic acids;

(b) denaturing the target mutant/probe duplex by subjecting the nucleic acid sample to a critical temperature such that the wild-type/probe duplex does not denature;

(c) collecting the eluate from the beads; and

(d) repeating at least once (a)-(c) using beads on which the probes are unbound to any nucleic acid.

Any of the methods disclosed herein can also be applied to samples in which the abundance of more than one mutation is determined. Accordingly, any one of the methods of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA further comprises determining the original abundance of one or more additional mutant alleles of target nucleic acids at different loci on the genomic DNA. In some embodiments of methods to determine abundance of one or more additional mutant alleles of target nucleic acids, the mutant alleles of the target nucleic acids are enriched relative to wild-type alleles of the target nucleic acids by subjecting the amplified portion to multiplexed NaME-PrO. In some embodiments, the multiplexed NaME-PrO uses overlapping probe pairs in parallel reactions, wherein each parallel reaction uses a distinct pair of overlapping probes, such that the totality of overlapping probe pairs span a contiguous region of genomic DNA of interest. In some embodiments, the multiplexed NaME-PrO uses continuous and non-overlapping probe pairs in parallel reactions, wherein each parallel reaction uses a distinct pair of non-overlapping probes, such that the totality of non-overlapping probe pairs span a contiguous region of genomic DNA of interest.

In some aspects, provided herein is a method of determining disease diagnosis or prognosis in a subject, the method comprising determining original abundance of a mutant target nucleic acid in a sample of genomic DNA according to any one of the methods disclosed herein. In some embodiments, the determining is repeated in two or more genomic DNA samples obtained from a subject over a period of time, and wherein an increase in original abundance of the target mutant nucleic acids in the sample over time is indicative of a worsening diagnosis or prognosis in the subject.

In some aspects, provided herein is a method of determining efficacy of a therapeutic treatment in a subject, the method comprising determining original abundance of a mutant target nucleic acid in a sample of genomic DNA according to any one of the methods disclosed herein. In some embodiments, the determining is repeated in genomic DNA samples obtained from a subject before and after treatment, and wherein a decrease in original abundance of the target mutant nucleic acids in the sample after treatment is indicative of therapeutic efficacy in the subject. In some embodiments, the determining is repeated in 2-10 genomic DNA samples obtained over a period of time. In some embodiments, the determining is repeated over a period of time of 2 months to 2 years.

In some aspects, provided herein is a method of obtaining a fingerprint of mutations associated with a disease, the method comprising determining original abundance of mutant alleles of target nucleic acids of a plurality of mutations according to any one of the methods disclosed herein. In some embodiments, the determining is performed on one or more samples of genomic DNA obtained from one or more subjects suffering from the disease.

BRIEF DESCRIPTION OF THE DRAWINGS

The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present disclosure, which can be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein. It is to be understood that the data illustrated in the drawings in no way limit the scope of the disclosure.

FIG. 1 describes Nuclease-assisted Minor-allele Enrichment using PRobe-Overlap (NaME-PrO). FIG. 1A shows partially overlapping oligonucleotide probes are included in excess molar ratio as compared to target DNA; and duplex specific nuclease (DSN). FIG. IB describes NaMe-PrO. Probes are designed to bind the WT sequence of top and bottom target DNA strands. The overlap region of the probes defines the DNA region targeted for mutation enrichment. FIG. 1C shows a NaME-PrO workflow; following fragmented genomic DNA denaturation, the excess probes bind their targets. Addition of DSN generates preferential digestion at double stranded regions defined by the probes, but not at mutation-caused mismatches. Numerous targeted sites can be targeted simultaneously in this manner.

Subsequent amplification enhances DNA target strands that escape DSN digestion, thereby enriching for mutated DNA by several hundred-fold.

FIG. 2 describes qNAME-PRO-NGS using single-molecule barcoding to enable strictly quantitative assessment of original mutation abundance by implementing allelic barcode-count diversity (the 'abed' approach). qNAME-PRO-NGS (i) enriches mutations by >100fold and diminishes the number or reads required to call low-level mutations; (ii) derives original mutation abundance; and (iii) filters-out sequencing and PCR errors using the molecular barcodes. qNAME-PRO technology increases throughput by >100-fold and reduces cost for all existing NGS technologies and can be applied unchanged for future sequencing developments.

FIG. 3A describes derivation of original mutation frequency using allelic barcode-count diversity (abed), following qNaME-PrO-NGS. The abed approach de-couples the mutation enrichment from the original mutation abundance, by using the number of distinct molecular barcode families that are mutant versus WT for digital quantification of mutation abundance. Mutation-containing barcodes are enriched and appear more frequently than barcodes with WT sequences. However this does not affect the quantification since it is the diversity of barcodes, not their frequency that is used.

FIG. 3B shows data obtained from an experiment in which the workflow of FIG. 2 was applied to a set of 33 DNA targets with mutations in circulating DNA from a breast cancer patient without without using the abed approach (left panel) and using the abed approach to determine the original abundance of mutant target nucleic acid (right panel). The light grey bars represent data obtained with mutant target nucleic acid enrichment using NaME-PrO. The dark grey bars represent data obtained without enrichment of mutant target nucleic acid.

FIG. 4 describes attachment of molecular barcodes by using multiplexed-PCR instead of ligation. Following multiplexed PCR, the molecular barcodes are attached in the first few cycles of a second PCR, followed by several cycles of PCR using the universal 'tails'. NaME-PrO is then applied, followed by library preparation and sequencing. The mutation enrichment occurs during the NaME-PrO step, i.e. after incorporation of the barcodes and initial expansion by the first few cycles of the second PCR.

FIG. 5 describes use of molecular barcodes for quantification when using multiplexed COLD-PCR. Following multiplexed PCR, the molecular barcodes are attached in the first few cycles of a second PCR, followed by several cycles of COLD-PCR using the universal 'tails'.

The mutation enrichment occurs during the COLD-PCR cycles, i.e. after incorporation of the barcodes and initial expansion by the first few cycles of the second PCR. Mutation-enriched COLD-PCR products are then processed for library preparation and sequencing.

FIG. 6 describes use of molecular barcodes for quantification when using multiplexed COLD-PCR and NaMe-PrO. Following multiplexed PCR, the molecular barcodes are attached in the first few cycles of a second PCR, followed by several cycles of COLD-PCR using the universal 'tails'. Mutation enrichment occurs during the COLD-PCR cycles, i.e. after incorporation of the barcodes and initial expansion by the first few cycles of the second PCR. Mutation-enriched COLD-PCR products are then further enriched for mutations using NaMe-PrO, and then processed for library preparation and sequencing.

FIG. 7 describes comprehensive sequencing of cfDNA from adult patients with metastatic cancer. FIG. 7A shows an example workflow. FIG. 7B shows analytical benchmarking. FIG. 7C shows somatic mutations in cfDNA and tumor biopsies. FIG. 7D shows somatic copy number alterations in cfDNA and tumor biopsies..

FIG. 8 describes multiplexed mutation enrichment via NaME-PrO applied to clinical samples (Song et al 2016). FIG. 8A shows 10-plex NaME-PrO applied directly to genomic DNA containing mutations in 10 actionable genes: mutations without NaME-PrO ('No DSN') vs. after NaME-PrO are compared, 50-100-fold enrichment is evident on all targets. FIG. 8B shows ddPCR screening of cfDNA from colon CA patients: improvement via NaME-PrO is evident for the 2 positive samples. FIG. 8C shows Sanger sequencing without vs. with NaMe-PrO: low level mutations are revealed. FIG. 8D shows MiSeq sequencing of 0.5% allelic abundance mutations with vs without NaME-PrO: the mutation becomes clearly distinguished from the sequencing noise after NaMe-PrO. This mutation was detectable with just 10 MiSeq sequencing reads following enrichment from 0.5% to 52% allelic abundance via NaME-PrO[8].

FIG. 9 describes NaME-PrO reveals multiple low-level mutations in cfDNA with a defined tumor signature. Whole-exome sequencing of both tumor tissue and corresponding cfDNA from a breast cancer patient revealed revealed 50 somatic mutations present in the tumor but absent in cfDNA. NaME-PrO was applied to these same 50 targets using 30ng cfDNA as starting material. Out of 50 targets, 31 (62%) revealed rare mutations following NaME-PrO-MiSeq NGS. Therefore the majority of the 'missing' tumor fingerprint was recovered in plasma

following NaME-PrO enrichment. In general, low-level mutations were detected with 1-2 orders of magnitude fewer reads as compared to standard NGS.

FIG. 10 describes NaME-PrO application for enriching multiple mutations along tumor suppressor genes. FIG. 10A shows a technique using overlapping probes: Mutation enrichment at all positions on TP53 exons 6-9 using concurrent NaME-PrO 'tiling' reactions. Genomic DNA mix of cell lines containing mutations at several TP53 positions in exons 6-9 was treated via NaME-PrO in seven parallel reactions using distinct groups of overlapping probes per reaction, such that NaME-PrO overlapping probes 'tile' the entire size of each exon. Mutation enrichment at all mutated sequences was observed following Sanger sequencing. FIG. 10B shows a technique using contiguous, non-overlapping probes: Using contiguous, non-overlapping probes that cover 25+25=50bp per reaction it is possible to cover an average cfDNA fragment size of ~150bp using 3-4 NaME-PrO parallel reactions, while still obtaining good mutation enrichment at all positions of all exons targeted. An example for multiple mutations along TP53 exon 8 is shown. Mutation enrichment 30-40-fold was derived by performing digital droplet PCR before & after NaME-PrO as described[8]. Significantly lung CA gene fusions (e.g., EML4-ALK fusionl) that result to breakpoints within the mapped exons will also be enriched by NaME-PrO, since probes corresponding to the WT cfDNA fragment will not bind adequately the fusion breakpoint and DSN will not digest those fragments, resulting to enrichment.

FIG. 11 shows sequencing cost (Illumina MiSeq) as a function of number of unique DNA mutation targets interrogated. Assumptions: 100% pass filter unique reads, uniform amplicon generation, one hotspot per amplified DNA fragment, $1,924 per lane (Broad SSF pricing). Sample preparation costs not included in this calculation. NaME-PrO application itself costs $17.3 for a 1000-target reaction including 2,000 unlabeled probes plus enzyme.

FIG. 12 shows attachment of molecular barcodes using ligation at the first step, followed by target enrichment using nested gene-specific primers on one end of each DNA target, plus the generic linker ligated in the first step.

DETAILED DESCRIPTION

Provided herein is a transformative technology that retains both high throughput (breadth) and sequencing depth for mutation analysis. The disclosed technology combines mutation enrichment methods (e.g., NaME-PrO or COLD-PCR) with a novel use of molecular

barcoding, to provide strict enumeration of original mutation abundance for all mutant sequences following their enrichment. Provided methods permit converting rare mutations to high abundance mutations, boosting confidence in their detection and circumventing the need for repeated and wasteful sequence reads during NGS. This enables NGS to reliably identify low-level alterations while retaining low cost and high throughput capabilities. Provided methods allow the quantification of the original mutation frequency in a sample (i.e., abundance of mutation frequency in a sample before it is processed for mutation enrichment), which was not previously possible.

There are several advantages of the disclosed methods over current state of the art. Exemplary advantages of the disclosed methods are discussed herein. For example, mutations down to 0.1% mutation abundance are clinically significant in heterogeneous tumors, circulating DNA and other clinical samples. Clinically actionable mutations in tumors or circulating DNA often exist at levels <0.1% which falls well below the ability of existing targeted re-sequencing technology. Provided methods allow quantification of mutation abundance after mutation enrichment so that detection is possible. Furthermore, methods disclosed herein are compatible with sequencing methods. The disclosed methods allow identification and quantification of mutations with low abundances with the current sequencing technologies available and in the current sample preparation framework.

In some embodiments, provided mutation enrichment methods (e.g., NaME-PrO) enable single tube enrichment of mutations that enable NGS-based detection down to extremely low abundances without requiring a substantial increase in the number of sequence reads as compared to high-level mutations. This may be achieved by pre-enriching the sample for mutations over all targets, prior to entering the NGS stage, thus converting low-level mutations to clonal mutations which are detectable with few sequence reads, and thereby reducing time and cost. FIG. 11 demonstrates an exemplary major reduction in cost as a function of the sequenced target size using the Illumina MiSeq as an example. The method is applicable to all platforms. However, the modest cost of MiSeq instrument makes it accessible to many groups or labs that do not have access to a high-end facility.

Reliability of mutant identification is another advantage of the methods disclosed herein. Even though molecular barcoding technology and sophisticated filtering can distinguish the few genuine sequencing from noise, at high noise levels there is doubt calling 2-3 mutation reads

amongst 100,000 WT reads. By increasing the allele fraction of mutations above the noise floor of NGS, errors associated with NGS can be overcome and reliability increased.

'Wild type target sequence' or 'wild type nucleic acid', used interchangeably herein, refers to a nucleic acid that is more prevalent in a nucleic acid sample than a corresponding target sequence (e.g., same region of gene but different nucleic acid sequence). The wild type sequence makes-up over 50% of the total wild type sequence + mutant target sequence in a sample. The wild type sequence can be expressed at the RNA and/or DNA level 10X, 15X, 20X, 25X, 30X, 35X, 40X, 45X, 50X, 60X, 70X, 80X, 90X 100X, 150X, 200X or more than the target sequence. For example, a sample (e.g., blood sample, urine sample, tissue sample) may contain numerous normal cells and few cancerous cells. The normal cells contain wild-type alleles (non-mutant) sequences, while the small number of cancerous cells contain target sequences. As used herein, a 'wild type strand' refers to a single nucleic acid strand of a wild type sequence. The term 'wild-type' typically refers to the most common polynucleotide sequence or allele for a certain gene in a population. Generally, the wild-type allele will be obtained from normal cells.

The wild type sequence is about 13-2000 nucleotides long. In some embodiments, the wild type sequence is 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or more nucleotides long. Wild type sequences will share at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology to the corresponding target sequence, but will differ by at least one nucleotide from the target sequence.

A 'target nucleic acid' or 'target sequence', used interchangeably herein, refers to a nucleic acid of interest (e.g., an allele). In some embodiments, a target nucleic acid has mutant alleles and wild-type alleles. In some embodiments, the target sequence makes-up less than 50% of the total amount of DNA (e.g., target sequence+non-target sequence) in a sample. Preferably the target sequence is expressed at the RNA and/or DNA level 1:10, 1:15, 1:20, l:25x, 1:30, 1:35, 1:40, 1:45, 1:50, 1:60, 1:70, 1:80, 1:90, 1:100, 1:150, l:200x or less than the non-target sequence. In some embodiments, the target sequence is a mutant allele. In some embodiments, the target sequence is a wild-type allele. For example, a sample (e.g., blood sample, urine sample, tissue sample) may contain numerous normal cells and few cancerous cells. The normal cells contain wild-type (i.e., non-mutant) sequences, while the small number of cancerous cells contain target mutant sequences. In some embodiments, the target sequence is repeat sequences

that occur at large numbers in human genome (including but not limited to ALU elements, LINE elements, SINE elements, di-nucleotide repeats, tri-nucleotide repeats). As used herein, a "target strand" refers to a single nucleic acid strand of a target sequence. A target sequence may comprise coding regions, non-coding regions, both coding and non-coding regions, a whole gene or a portion or a gene.

In some embodiments, the target sequence is 13-2000 nucleotides long. In some embodiments, the target sequence is 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or more nucleotides long. Target sequences share at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology to the corresponding wild type sequence, but differs by at least one nucleotide from the wild type sequence.

In some embodiments, target nucleic acids are enriched relative to total nucleic acids. In some embodiments, mutant target nucleic acids and wild-type target nucleic acids are enriched relative to total nucleic acids in a sample of double-stranded genomic DNA. Various methods of enriching target nucleic acids relative to total nucleic acids are known in the art and are contemplated herein. For example, target nucleic acids in a sample can be enriched by contacting the sample with one or more probes that specifically bind target nucleic acids. Probes that bind to target nucleic acids may be immobilized on a solid substrate (e.g., a bead) for easier separation of the target nucleic acids from the non-target nucleic acids. In some embodiments, probes bind to target nucleic acids by hybridization. In some embodiments, hybridization between probes and target nucleic acids is complete. In some embodiments, hybridization between probes and target nucleic acids is partial. For example, there may be 1 or more mismatches between a probe and target sequence. In some embodiments, only a part of a probe hybridizes with a target sequence.

Another method of target enrichment is shown in FIG. 12. Accordingly, in some embodiments, target enrichment is achieved using steps comprising amplification of a double-stranded DNA fragment using primers, at least one of which is a nested gene-specific primer. A 'nested gene-specific primer' is a primer than anneals to the product of amplification made using another set of primers to amplify a target region. In FIG. 12 for example, the nested gene-specific primer in step (4) nests such that it anneals to the product of a preceding pre-amplification step that uses primers that are complementary to the common sequence tags.

Using the example in FIG. 12, the nested gene-specific primer of step (4) is nested relative to, or nests within, the primers complementary to the common sequence tags used in step (3). Nested primers may produce and amplification product of a shorter length than the product of the primers within which they are nested. Accordingly, a nested gene-specific primer may nest within a set of multiplexed primers such that it anneals to the product of the multiplexed amplification. Multiplexed primers within which a target-specific primer nests may be random primers, gene specific primers, or primers complementary to a common sequence tag. In some embodiments, an entire gene-specific primer anneals to a target sequence. In some embodiments, a portion of a gene-specific primer anneals to a target sequence. In some embodiments, a gene-specific primer anneals to a target sequence that is genomic DNA of a coding region, or genomic DNA of a non-coding region. In some embodiments, a gene-specific primer anneals to a target sequence that comprises a mutation or is adjacent (e.g., up to 30, 50, 100 or more nucleotides upstream or downstream) to a mutation. In some embodiments, a nested gene-specific primer comprises a sequence tag, that can be used for subsequent amplification using primers complementary to this sequence tag. A sequence tag on a gene-specific primer is different from double-stranded common sequence tag used in prior steps of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA, such that primers designed to anneal to a common sequence tag will not anneal to the sequence tag on a nested gene-specific primer, or primers designed to anneal to a sequence tag on a nested gene-specific primer will not anneal to a double-stranded common sequence tag. In some embodiments, a sequence tag on a gene-specific primer is different from sequence tags on gene-specific primers that anneal to other targets, such that primers designed to anneal to a sequence tag on a gene-specific primer targeting one target will not anneal to a sequence tag on a gene-specific primer targeting another target. In some embodiments of any one of the methods for target enrichment, a primer may comprise a sequencing adaptor that can be used during sequencing. In some embodiments, target enrichment achieved using amplification of a double-stranded DNA fragment using primers, at least one of which is a nested gene-specific primer, is performed for 2 to 20 cycles (e.g., 2 to 5, 4 to 18, 5 to 15 or 8 to 15 cycles).

In some embodiments, target nucleic acids (i.e., wild-type and mutant alleles of a target) are enriched 2 to 1000 times (e.g., 2 to 5, 2 to 10, 2 to 50, 2 to 100, 10 to 100, 50 to 200, 100 to 500, 200 to 1000 or 400 to 1000 times) relative to an unenriched sample.

'Target mutant sequence' or 'mutant target sequence' refers to a nucleic acid that is less prevalent in a nucleic acid sample than a corresponding wild type sequence. The target mutant sequence typically makes-up less than 50% of the total amount of wild type sequence + mutant sequence in a sample. The target mutant sequence may be expressed at the RNA and/or DNA level 1 :10, 1 :15, 1 :20, 1 :25X, 1 :30, 1 :35, 1 :40, 1 :45, 1 :50, 1 :60, 1 :70, 1 :80, 1 :90, 1 :100, 1 :150, 1 :200X or less than the wild type sequence. For example, a sample (e.g., blood sample) may contain numerous normal cells and few cancerous cells. The normal cells contain wild-type (non-mutant) alleles, while the small number of cancerous cells contain target mutant sequences. In some embodiments, the invention is directed to detecting fetal DNA in a nucleic acid sample obtained from a mother. In this embodiment, the target mutant sequence is the fetal DNA while the more prevalent mother DNA is the wild type sequence. As used herein, a target mutant sequence is meant to include fetal DNA obtained from a pregnant mother. A mutation may be in a coding region or a non-coding region of a genome. A mutation may be in a known location or site. In some embodiments, a mutation is unknown. In some embodiments, the exact location or a mutation is unknown. In some embodiments, a mutation and its site is unknown.

The target mutant sequence is about 13-2000 nucleotides long. In some embodiments, the target mutant sequence is 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or more nucleotides long. Target mutant sequences share at least 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology to the corresponding wild type sequence, but differs by at least one nucleotide from the wild type sequence.

The term 'mutant' refers to a nucleotide change (i.e., a single or multiple nucleotide substitution, deletion, insertion, or methylation, or alteration in the number of poly-nucleotide repeats) in a nucleic acid sequence. A nucleic acid which bears a mutation has a nucleic acid sequence (mutant allele) that is different in sequence from that of the corresponding wild- type sequence. Herein, the term "mutant target nucleic acid" is used interchangeably with "mutant alleles of target nucleic acid." Similarly, the term "wild-type target nucleic acid" is used interchangeably with "wild-type alleles of target nucleic acid." The mutant alleles can contain between 1 and 500 nucleotide sequence changes. A mutant allele may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400 or 500 nucleotide sequence changes compared to a corresponding wild-type allele.

Typically, a mutant allele will contain between 1 and 10 nucleotide sequence changes, and more typically between 1 and 5 nucleotide sequence changes. The mutant allele will have 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology to the wild-type allele. Generally, the mutant allele will be obtained from diseased tissues or cells and is associated with a disease state.

As used herein, a 'region of interest' is a sequence that will be interrogated for variations such as clinically relevant mutations.

As used herein, the term "relative abundance" refers to the amount of the nucleic acid of the same type as compared to the amount of the total nucleic acid in a sample. For example, the relative abundance of a mutant nucleic acid in a sample may be defined as the percentage of the abundance of the mutant nucleic acid to the total number of nucleic acid corresponding to the target sequence in which the mutations lies, which includes both mutant and wild-type nucleic acid for that target sequence.

'Enriching a mutant target sequence' refers to increasing the amount of a mutant target sequence and/or increasing the ratio of mutant target sequence relative to the corresponding wild type sequence in a sample. For example, where the ratio of mutant sequence to wild type sequence is initially 5% to 95% in a sample, the mutant sequence may be preferentially amplified in an amplification reaction so as to produce a ratio of 70% mutant sequence to 30% wild type sequence. Thus, there is a 14 fold enrichment of the mutant sequence relative to the wild type sequence in this hypothetical example. Generally, enrichment of a mutant target sequence results in a 2X to 200X increase in the mutant target sequence relative to the wild type sequence prior to enrichment. The enrichment of the mutant target sequence is at least a 2X, 3X, 4X, 5X, 6X, 7X, 8X, 9X, 10X, 15X, 20X, 25X, 30X, 35X, 40X, 45X, 50X, 60X, 70X, 80X, 90X 100X, 150X, 200X or more fold enrichment. Enrichment of a mutant target sequence results in a sample having 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 80%, 90%, 95% or more, mutant target sequence compared to wild type sequence (e.g., 10% mutant target sequence : 90% wild type sequence to 95% mutant target sequence : 5% wild type sequence).

Several methods of enriching mutant target sequence relative to wild-type target sequence are known in the art. Non-limiting examples of mutation enrichment methods include Nuclease-assisted Minor-allele Enrichment using Probe Overlap (NaME-PrO), Coamplification

at Lower Denaturation temperature-PCR (COLD-PCR), Improved and Complete Enrichment COLD-PCR (ice-COLD-PCR), Temperature-Tolerant ice-COLD-PCR (TT-ice-COLD-PCR), toehold PCR blocker-PCR using locked nucleic acids (LNA), peptide nucleic acids (PNA or XNA), or blockers with other modified nucleotides, CRISP-R-mediated mutation enrichment by selecting mutated alleles or removing wild-type alleles and Differential Strand Separation at Critical Temperature (DiSSECT).

NaMe and NaMe-PrO methods are described in PCT/US2016/039167, which is incorporated herein by reference in its entirety. A non-limiting example of a NaMe-PrO protocol includes:

(a) preparing an amplification reaction mixture comprising the double-stranded mutant and wild-type target nucleic acids, a thermostable double strand-specific nuclease (DSN), PCR amplification components, and a pair of oligonucleotide probes, one of which is complementary to the wild-type nucleic acid top strand and the other is complementary to the wild-type nucleic acid bottom strand, wherein the probes may overlap each other by 10-15 probes such that the overlap coincides with the target region or be non-overlapping and contiguous;

(b) subjecting the reaction mixture to a denaturing temperature to permit denaturation of the wild-type nucleic acid and the mutant target nucleic acid;

(c) reducing the temperature to permit hybridization of the probes to their corresponding sequences on the wild-type and mutant target nucleic acids thereby forming complementary wild-type-probe duplexes, wherein the DSN cleaves the complementary wild-type-probe duplexes but not the partially complementary target mutant -probe duplexes; and

(d) subjecting the reaction mixture to an amplification condition thereby enriching the uncleaved mutant target nucleic acid relative to the cleaved wild-type nucleic acid.

In some embodiments, an overlap of NaMe-PrO probes coincides with one or more mutations. In some embodiments, NaMe-PrO probes have a 3'-terminal polymerase block. In some embodiments, the probes are complementary to SNPs near target mutations.

In some embodiments, NaMe or NaMe-PrO is then followed by amplification of remaining mutant and wild-type target nucleic acids. In some embodiments, 1-50 cycles (e.g., 1-40, 2-30, 5-15, 5-10 cycles) of PCR are used to amplify mutant and wild-type target nucleic acids after NaMe or NaMe-PrO.

In some embodiments, NaMe-PrO with or without amplification results in mutation enrichment relative to wild-type target nucleic acids of 1-200-fold (e.g., 1-150-, 5-100- or 10-100-fold) compared to the unenriched sample. In some embodiments, NaMe-PrO with or without amplification results in mutation enrichment relative to wild-type target nucleic acids of more than 200-fold (e.g., 250-fold or 300-fold).

In some embodiments, a form of COLD-PCR (e.g., ice-COLD-PCR, TT-ice-COLD-PCR or oscillating COLD-PCR) is used to enrich mutant target nucleic acids relative to wild-type target nucleic acids. Methods of COLD-PCR and oscillating COLD-PCR are described in WO 2009/017784, which is incorporated by reference herein in its entirety. A non-limiting example of a COLD-PCR protocol includes:

(a) denaturing the double-stranded mutant and wild-type target nucleic acids by subjecting the double-stranded target mutant and wild-type nucleic acids to a first denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(b) forming a target mutant/wild-type strand duplex;

(c) denaturing said mutant/wild-type strand duplex by subjecting the nucleic acid amplified portion to a critical temperature (Tc) that is below the Tm of the wild-type nucleic acids;

(d) annealing a primer pair to the mutant and wild-type target nucleic acid strands; and

(e) extending said primer pair so as to enrich said mutant target sequence relative to said wild-type strand.

In some embodiments, COLD-PCR is performed for 1-50 cycles (e.g., 1-40, 2-30, 5-25, 8-20 or 5-10 cycles) to enrich mutant target nucleic acid relative to wild-type target nucleic acid.

If the above example of COLD-PCR were to be adapted for oscillating COLD-PCR, steps (b) and (c) would be repeated. In some embodiments of oscillating COLD-PCR, forming a target mutant/wild-type strand duplex and denaturing said mutant/wild-type strand duplex, is repeated 1-29 times (e.g., 1-19 or 2-9 times).

Methods of ice-COLD-PCR and TT-COLD-PCR are described in WO 2012/135664, which is incorporated by reference herein in its entirety. A non-limiting example of an ice-COLD-PCR protocol includes:

(a) exposing the mutant and wild-type target nucleic acids to a reference sequence that is complementary the target sequence;

(b) denaturing the double-stranded target mutant and wild-type nucleic acids by subjecting the double-stranded mutant and wild-type target nucleic acids to a first denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(c) forming a target mutant/reference strand and target wild-type/reference strand duplexes;

(d) denaturing said mutant/reference strand duplex by subjecting the nucleic acid sample to a critical temperature (Tc) that is below the Tm of the wild-type/reference duplex;

(e) annealing a primer pair to the mutant and wild-type target nucleic acid strands; and

(f) extending said primer pair so as to enrich said mutant target sequence relative to said wild-type target nucleic acid.

A non-limiting example of a TT-ice-COLD-PCR protocol includes:

(a) exposing the mutant and wild-type target nucleic acids to a reference sequence that is complementary the target sequence;

(b) denaturing the double-stranded target mutant and wild-type target nucleic acids by subjecting the double-stranded target mutant and wild-type nucleic acids to a denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(c) forming a target mutant/reference strand and target wild-type/reference strand duplexes;

(d) denaturing said mutant/reference strand duplex by subjecting the nucleic acid sample to a first critical temperature (Tc) that is below the Tm of the wild-type/reference duplex;

(e) annealing a primer pair to the mutant and wild-type target nucleic acid strands;

(f) extending said primer pair so as to enrich said mutant target nucleic acid relative to said wild-type target nucleic acid; and

(f) repeating steps (d) to (f) at least once at a second Tc which is above the first Tc.

In some embodiments, any form of COLD-PCR (as described above) with or without amplification results in mutation enrichment relative to wild-type target nucleic acids of 1-200-fold (e.g., 1-150-, 5-100- or 10-100-fold) compared to the unenriched sample. In some embodiments, any form of COLD-PCR with or without amplification results in mutation enrichment relative to wild-type target nucleic acids of more than 200-fold (e.g., 250-fold or 300-fold) compared to the unenriched sample.

In some embodiments, DiSSECT is used to enrich mutant target nucleic acids relative to wild-type target nucleic acids. DiSSECT is a method that enriches unknown mutations of targeted DNA sequences purely based on thermal denaturation of DNA duplexes without the need for enzymatic reactions. Methods of DiSSECT are described in Guha et al. (Nucleic Acids Research, 2012, 1-9), which is incorporated herein by reference in its entirety. A non-limiting example of a DiSSECT protocol includes:

(a) allowing mutant and wildtype target nucleic acids to bind to complementary probes which are immobilized to beads, wherein the probes resemble the wild-type nucleic acids;

(b) denaturing the target mutant/probe duplex by subjecting the nucleic acid sample to a critical temperature such that the wild-type/probe duplex does not denature;

(c) collecting the eluate from the beads; and

(d) repeating at least once (a)-(c) using beads on which the probes are unbound to any nucleic acid.

In some embodiments, DiSSECT is performed for 1-20 cycles (e.g., 1-18, 2-6, 2-4, 2- 10 or 5-15 cycles). In some embodiments, DiSSECT results in mutation enrichment relative to wild-type target nucleic acids of 1-600-fold (e.g., 100-fold, 200-fold, 300-fold, 400-fold, 500-fold or 600-fold) compared to the unenriched sample.

'Allele' refers to alternative forms of a gene, portion thereof or non-coding region of DNA that occupy the same locus or position on homologous chromosomes that have at least one difference in the nucleotide sequence. The term allele can be used to describe DNA from any organism including but not limited to bacteria, viruses, fungi, protozoa, molds, yeasts, plants, humans, non-humans, animals, and archaebacteria. The alleles may be found in a single cell (e.g., two alleles, one inherited from the father and one from the mother) or within a population of cells (e.g., a wild-type allele from normal tissue and a somatic mutant allele from diseased tissue).

An allele can be 13-2000 nucleotides long. In one embodiment the allele is 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 or more nucleotides long. Alleles will generally share 50%, 60%, 70%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more homology to each other.

As used herein, a nucleic acid sample refers to any substance containing or presumed to contain a nucleic acid of interest (target and wild type sequences) or which is itself a nucleic acid containing or presumed to contain a target nucleic acid of interest. The term "nucleic acid sample" thus includes a sample of nucleic acid (genomic DNA, cDNA, RNA), cell, organism, tissue, or fluid, including but not limited to, for example, plasma, serum, spinal fluid, lymph fluid, synovial fluid, urine, tears, stool, external secretions of the skin, respiratory, intestinal and genitourinary tracts, saliva, blood cells, tumors, organs, tissue, samples of in vitro cell culture constituents, natural isolates (such as drinking water, seawater, solid materials), microbial specimens, and objects or specimens that have been "marked" with nucleic acid tracer molecules. The nucleic acid sample may be obtained from mammals, viruses, bacteria or plants. In some embodiments, the nucleic acid sample is DNA circulating in plasma, urine or other bodily fluids.

As used herein "oligonucleotide probes" refer to molecules comprising two or more deoxyribonucleotides or ribonucleotides. By "complimentary" it is meant that the probes hybridize without any mismatches to the sequences in the top and bottom stands of the wild type nucleic acid.

As used herein, "primers' refers to oligonucleotides that anneal to opposite strands of a mutant target and wild type sequence so as to form an amplification product during a PCR reaction.

The term "barcode" as used herein refers to a unique sequence of nucleotides that allows identification of the nucleic acid of which the barcode is a part. Barcoding a DNA fragment is a process by which the DNA fragment is uniquely tagged with one or more short identifying sequences. In some embodiments, it is desired for each DNA fragment in a sample to have a barcode that is unique from barcodes on any other DNA fragment in the sample. In some embodiments, each DNA fragment in a sample comprises one unique barcode. In some embodiments, each DNA fragment in as sample comprises two barcodes that are unique from each other and unique from any other barcode that is attached to any other DNA fragment in the sample. Such uniqueness of barcodes in a sample of DNA fragments can be accomplished, for example, by optimizing the length of each barcode (i.e., the number of nucleotides in each barcode) and/or the ratio of unique barcodes to DNA fragments during barcoding (i.e., attaching barcodes to DNA fragments).

Barcodes can be any appropriate length. In some embodiments, the length of each barcode used to barcode DNA fragments in a sample is 6-20 bp long (e.g., 8-18 bp, 8-14 bp, 10-16 bp or 12-14 bp). In some embodiments, the length of each barcode used to barcode DNA fragments in a sample is 14 bp long.

Barcodes can be attached to DNA fragments in any appropriate ratio. In some embodiments, the ratio of unique barcodes to DNA fragments during barcoding is 10^6-10^{10} (e.g., 10^6, 10^7, 10^8, 10^9, 10^{10}) unique barcodes to 100 ng of DNA (or 3xl0 ^{4} allelic copies).

Methods of attaching barcodes to nucleic acids are known in the art. Various publications provide descriptions of barcoding technology. For example, Wong and Moqtaderi (Curr Protoc Mol Biol. 2013;Chapter 7:Unit 7.11) describe a barcoding protocol for the preparation of up to 96 ChIP samples for multiplex sequencing in a single flow cell lane on the Illumina platform; and Stahlberg et al. (Nucleic Acids Res. 2016 Jun 20;44) describe a PCR-based barcoding method, both of which are incorporated herein by reference in their entirety. The following patents that also describe DNA barcoding methods are also incorporated herein by reference in their entirety: US8691509, US8268564 and US application 20120220494A1.

In some embodiments, double-stranded barcodes are attached to a double-stranded DNA fragment by ligation. In some embodiments, double-stranded DNA fragments are barcoded using PCR technique employing primers that comprise unique barcodes.

As used herein, the term, "common sequence tag" refers to a nucleotide sequence that is common to all the DNA fragments in a sample, e.g., a sample of genomic fragmented DNA. A common tag enables processing of all the DNA fragments in a sample. For example, primers complementary to a common tag in a sample may be used to amplify all DNA fragments in a sample, regardless of whether a DNA fragment contains target nucleic acid (e.g., mutant target nucleic acid or wild-type target nucleic acid) or non-target nucleic acid.

For any one of the methods disclosed herein, a sample of double-stranded genomic DNA may comprise double-stranded DNA fragments, wherein each terminus of the genomic DNA fragments is attached to a unique double-stranded barcode and a double-stranded common sequence tag, wherein the common sequence tag is located upstream of the unique barcode. By being located "upstream" of the unique tag, it is meant that the common tag is located 5' relative to the unique tag if the unique barcode, common tag and DNA fragment sequence are read from

5' to 3'. FIG. 3A depicts an example of how a common tag is located relative to a unique barcodes for any one of the methods described herein.

Common sequence tags can be any appropriate length. In some embodiments, a common tag is 16-40 bp long (e.g., 16-40, 18-36, 20-32, 22-30 or 24-28 bp long). In some embodiments, a common tag is 18 nucleotides long (i.e. an 18-mer). It is to be understood that a the terms "nucleotide" and 'base pair (bp)" are used interchangeably herein.

In some embodiments, a unique barcode and a common tag are attached to each end of a double-stranded DNA fragment at the same time using the same method. In some embodiments, a barcode and a common tag are attached to each end of a double-stranded DNA fragment by ligation.

In some embodiments, a unique barcode and common tag are attached to a terminus of a DNA fragment by starting from a single-stranded barcode, synthesizing the opposite strand of the single-stranded barcode using an extension reaction to form a double stranded barcode, and the ligating and end of the double-stranded genomic DNA fragment to the end of the barcode (FIG. 3A depicts such an embodiment).

In some embodiments, a barcode and a common tag are attached to each end of a double-stranded DNA fragment by using multiplexed-PCR. In such embodiments, PCR using oligonucleotide primers are used, wherein each oligonucleotide primer comprises a common tag portion, a unique barcode portion and a target-specific portion. The target-specific portion enables attachment of the oligonucleotide primer to anneal to DNA fragments. Step 3 of FIGs. 4, 5 and 6 depict such an embodiment.

As used herein, the term, "sample-index" refers to a unique sequence of nucleotides that allows identification of the sample (e.g., an enriched or un-enriched sample) of DNA from which the DNA is sourced. For example, a sample of DNA fragments may be split so that one aliquot is enriched for mutant alleles, while another aliquot is not enriched. DNA from both samples can then be mixed and sequenced in the same sequencing reaction if sample indeces on the DNA allows identification of the source of the DNA, i.e., whether it was sourced from the enriched sample or the unenriched sample. For example, a sample-index on a DNA fragment from the enriched sample is the same as the sample-indices from all other DNA fragments from the enriched sample. Similarly, the sample-index on a DNA fragment from the un-enriched sample is the same as the sample-indices from all other DNA fragments from the un-enriched sample.

Sample-indexes, like barcodes, can be any appropriate length. In some embodiments, the length of each sample-index used to index each DNA sample is 6-20 bp long (e.g., 8-18 bp, 8-14 bp, 10-16 bp or 12-14 bp). In some embodiments, the length of each barcode used to barcode DNA fragments in a sample is 14 bp long.

In some embodiments, the methods disclosed herein require genomic DNA to be in fragmented form In some embodiments, genomic DNA in a sample collected from a subject is already fragmented. For example, a sample of cell-free genomic DNA or genomic DNA circulating in blood is fragmented when collected. In some embodiments, genomic DNA from samples of the urine of a subject is fragmented. In some embodiments, genomic DNA collected from bodily fluid or tissue sample of a subject is not fragmented and needs to be fragmented. In some embodiments, a sample of genomic DNA is fragmented but it is desired to fragment it further to make smaller fragments. Various techniques to fragment double-stranded DNA are known in the art. In some embodiments, DNA is sheared physically (e.g., using acoustic shearing using a Covaris instrument, sonication using a Bioruptor or hydrodynamic shearing using a Hydroshear instrument). In some methods, double-stranded DNA is sheared enzymatically using any DNAase type of enzyme that digests DNA randomly (e.g., a Shearase, DNAsel or a transposase). In some embodiments, double-stranded DNA is fragmented by chemical fragmentation (e.g., by exposing the DNA to be fragmented to heat and divalent metal cation. Depending on the method of DNA fragmentation, DNA fragments may be subjected to enzymatic end-repairing to obtain blunt ends.

In some embodiments of any one of the methods disclosed herein, a double-stranded DNA fragment is 20-400 bp long (e.g., 10-400, 40-200, 50-150 or 50-100 bp long).

In any of the methods disclosed herein, obtaining a measure of total unique barcodes in a sample may be accomplished using DNA sequencing methods. Several sequencing methods and protocols for sample preparation for these methods are well-established in the art. Indeed, one of the advantages of the methods disclosed herein is that they are compatible with established methods of sample preparation for sequencing methods used in the field. Examples of methods of sequencing include SANGER sequencing, MiSeq sequencing, massively parallel signature sequencing (MPSS), pology sequencing, 454 sequencing, Illumina (or Solexa) sequencing, SoLiD sequencing, Ion Torrent semiconductor sequencing, single molecule real time (SMRT) sequencing, and nanopore sequencing. The following publications describe various sequencing

options and are incorporated herein by reference in their entirely: Goodwin et al. (Coming of age: ten years of next-generation sequencing technologies, Nature Reviews Genetics 17, 333—351 (2016)), Heather and Chain (The sequence of sequencers: The history of sequencing DNA, Genomics, 107: 1-8 (2016)), and Moorthie et al. (Review of massively parallel DNA sequencing technologies, Hugo J. 2011 Dec; 5(1-4): 1-12).

Methods of determining the abundance of more than one mutation in a sample of genomic DNA

In some embodiments, any one of the methods disclosed here is applied to determine the abundance of more than one mutation in a sample of genomic DNA. In some embodiments, the more than one mutations are at different loci on genomic DNA. In some embodiments, mutations are present at known hotspots for a particular disease. For example, there are 292 known hotspots for mutations in lung cancer (Newman et al., Nat Med 2014, 20:548-54; Newman et al., Nat Biotechnol 2016, 34:547-55). In some embodiments, the more than one mutations are on the same exon. In some embodiments, the more than one mutations are present on different exons. In some embodiments, it might be desired to scan for mutations on a particular exon rather than testing particular hotspots.

In some embodiments, mutant alleles of a target nucleic acid are enriched relative to wild-type alleles of target nucleic acid by subjecting the sample to multiplexed NaME-Pro. In some embodiments, multiplexed NaME-PrO uses overlapping probe pairs in parallel reactions, wherein each parallel reaction uses a distinct pair of overlapping probes, such that the totality of overlapping probe pairs span a contiguous region of genomic DNA of interest. Such an embodiment is exemplified in FIG. 10A. In some embodiments, multiplexed NaME-PrO uses continuous and non-overlapping probe pairs in parallel reactions, wherein each parallel reaction uses a distinct pair of non-overlapping probes, such that the totality of non-overlapping probe pairs span a contiguous region of genomic DNA of interest. Such an embodiment is exemplified in FIG. 10B.

Provided herein are methods of diagnosis and or prognosis of a disease based on determining abundance of a mutant allele of a target nucleic acid or sequence in a sample of genomic DNA from a subject. Accordingly, provided herein is a method of determining disease diagnosis or prognosis in a subject, which comprises determining the abundance of a mutant allele of a target nucleic acid in a sample of genomic DNA of the subject according to any one of

the methods disclosed herein. An increase in the abundance of a mutant allele of a target nucleic acid in a sample of genomic DNA of the subject is indicative of a worsening diagnosis or prognosis of disease in the subject. In some embodiments, abundance of a mutant allele of a target nucleic acid in a sample of genomic DNA of the subject is determined over a period of time (e.g., over weeks, months or years).

In some embodiments, any one of the methods disclosed herein is used to determine efficacy of a therapeutic treatment or intervention in a subject. In some embodiments, efficacy of a therapeutic treatment or intervention is determined by determining the abundance of mutant alleles in a sample of genomic DNA from the subject after the therapeutic intervention or treatment and comparing it to the abundance of mutant alleles in a sample of genomic DNA from the subject before the subject was subjected to the therapeutic treatment or intervention. A decrease in abundance of mutant alleles in the sample after treatment is indicative of therapeutic efficacy in the subject. In some embodiments, abundance of mutant alleles in a sample of genomic DNA from the subject is determined more than once after a therapeutic treatment or intervention. In some embodiments, abundance of mutant alleles in a sample of genomic DNA from the subject is determined 2-5 times after the subject undergoes a therapeutic treatment or intervention (e.g., two times a week after a treatment, or once every 6 weeks repeated 5 times).

Any one of the methods described herein can be used to obtain a fingerprint of mutations that are associated with a particular disease or condition. In some embodiments, such a method comprises determining relative original abundance of mutant alleles of target nucleic acids of a plurality of mutations according to the methods described herein, wherein the abundance of mutant alleles of target nucleic acids of a one or a plurality of mutations is determined in genomic DNA samples from one or more than one subject with the disease or condition. In some embodiments, the disclosed fingerprinting method can be applied to one type of sample from subjects (e.g., plasma samples of lung cancer patients at different cancer stages). Such a fingerprint can enable non-invasive testing for diagnosing and prognosing a disease or condition in a subject.

Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present disclosure to its fullest extent. The following specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the

remainder of the disclosure in any way whatsoever. All publications cited herein are incorporated by reference for the purposes or subject matter referenced herein.

EXAMPLES

Example 1: comprehensive sequencing of cell-free DNA

This Example describes next generation sequencing of circulating cells, circulating DNA and bioinformatics Workflows for comprehensive sequencing of cell-free DNA were established and some of the most extensive benchmarking of liquid and conventional biopsies performed using both whole-exome and whole-genome sequencing in patients with metastatic cancer (FIG. 7 and [3]).

Example 2: qNaME-PrO for mutation enrichment in lung cancer mutation hotspots

This example describes application of quantitative NaMe-PrO for mutation enrichment in lung cancer mutation hotspots. Following fragmentation of genomic DNA, DNA molecules are ligated to oligonucleotides containing unique identifiers (molecular barcodes), followed by pre-amplification and target selection for DNA targets containing potential mutation hotspots. 292 hotspots that are known to be recurrently mutated in lung cancer[1, 2] will be focused upon. qNAME-PRO is then applied to enrich mutations by >100-fold over all targets tested, followed by NGS. Serial dilutions of mutant-to-wild-type DNA, 5% - 0.01% abundance are tested. Computational techniques are employed to provide mutation quantification based on allelic barcode-count diversity (abed) and filter-out sequencing and PCR errors. Accuracy and selectivity of qNaME-PrO-NGS is evaluated and compared to existing NGS practices.

NaME-PrO (FIG. 1) employs a double-strand-DNA-specific nuclease (DSN) and overlapping oligonucleotide-probes interrogating WT-DNA targets and guiding nuclease digestion to these sites. Mutation-containing DNA creates probe-DNA mismatches that inhibit digestion, thus subsequent DNA-amplification magnifies DNA-alterations at all selected targets. NaME-PrO is a single step approach that operates at the level of genomic DNA prior to performing DNA amplification hence it is not impacted by polymerase errors. NaMe-PrO can achieve ~50-fold to several-hundred-fold mutation-enrichment, depending on conditions

applied[8], in diverse human samples on multiple clinically-relevant targets including tumor samples and circulating-DNA in 10-plex or 50-plex reactions[8]. Enrichment enables routine mutation detection at 0.01% abundance while by adjusting conditions it is possible to sequence mutations down to 0.00003% abundance, or to scan tumor-suppressor genes for rare-mutations[8]. NaME-PrO introduces a simple and highly parallel process to remove un-informative DNA sequences and unmask clinically and biologically useful alterations. The method creates the potential for massively parallel mutation enrichment prior to sequencing and engenders a new paradigm whereby rare mutations do not require deep sequencing for their detection. Representative examples of the application of NaME-PrO in clinical specimens and cfDNA are shown in FIG. 8, while additional data and details are in [8].

Recovery of tumor fingerprint in plasma-circulating DNA following NaME-PrO.

In another demonstration of the power of the technique, NaME-PrO was applied followed by MiSeq targeted re-sequencing to cfDNA from a breast CA patient whose tumor and cfDNA were both sequenced over the whole exome (150x read-depth). There were 50 DNA targets found mutated in the tumor but non-mutated in plasma. It was hypothesized that at least some of these mutations are present in cfDNA but are at too low level to be identified by standard exome sequencing. To examine the hypothesis, linkers were ligated to cfDNA followed by 10 cycles ligation-mediated PCR using Q5 high fidelity polymerase (as in step 3, FIG. 2), followed by application of 50-plex NaME-PrO. Next, a 50-plex PCR was applied to provide target enrichment, followed by Illumina adaptor attachment and MiSeq sequencing. The data (FIG. 9) revealed 31 somatic mutations in cfDNA matching the 'missing' tumor mutations. Most of these were at 0.5-2% levels, however a few of these cfDNA mutations were at original mutation levels of 0.03-0.1% and were deemed un-reliable due to the 'noise'. Following NaME-PrO these mutations were clearly scored. These data explain why the mutations were not found via exome sequencing of cfDNA and demonstrate the ability of NaME-PrO-MiSeq to recover the mutation fingerprint in cfDNA.

Study design

Sources of DNA, clinical samples.

(a) DNA from cell lines: DNA with defined mutations are obtained from Horizon diagnostics, UK, and also from human cancer cell lines with fully sequenced genome (COSMIC mutation database, Cell Line Project, Sanger Institute, UK, http://cancer.sanger.ac.uk/cancergenome/projects/cell_lines/). This is serially diluted (5-0.01%) into wild-type human male genomic DNA. The resulting mutation abundance is verified via digital droplet PCR, ddPCR using the known methods[57]. To validate mutation enrichment in regions for which no mutation-containing cell line is available, error-prone PCR will be employed for the first 10 cycles of LMPCR in FIG. 2.

(b) DNA from clinical cancer samples and plasma-circulating-DNA: Resected lung tumor specimens and serial plasma samples previously collected from 100 cancer patients undergoing radio-chemo- therapy treatment are used. Blood obtained 1 week before therapy, at initiation of therapy, weekly throughout the course of therapy, and every 2-4 months after the end of therapy was processed for plasma and DNA extraction within 1h of collection. Lymphocyte DNA and Qiagen-extracted circulating-DNA was stored at -80°C.

Selection of lung cancer hotspot mutation and full length tumor suppressor gene databases.

Hotspot mutations have been previously identified from large-scale sequencing of patients with lung cancer; in fact, screening for these mutations in cell-free DNA of patients with lung cancer using hybrid capture has demonstrated broad patient coverage [1, 2]. To enable benchmarking with recent, sophisticated approaches for hybrid capture and sequencing, focus will be put on panel development on the same hotspots and genes covered by the CAPP-Seq / iDES approach[1, 2]. The CAPP-Seq/iDES database includes 292 hotspots recurrently mutated in lung cancer as well as 521 full exons of 139 genes including well-known tumor suppressor genes, TP53, ALK, etc. The lung cancer patient coverage of the CAPP-Seq panel includes more than 96% of patients with lung adenocarcinoma or squamus cell carcinoma and identifies a median of 4 mutations per patient[1, 2].

Synthesis of molecular barcodes containing common 'tags' and ligation to randomly-fragmented genomic DNA

To synthesize random molecular barcodes that ensure that each molecule of a given DNA allele is ligated to a different barcode, an excess of random barcodes over anticipated allelic copies must be used. For example, if the starting genomic DNA is lOOng (~3xl0 ^{4} allelic copies), 10^{8} different barcodes during ligation with the randomly fragmented DNA so that the chances of the same barcode ligating to two alleles of the same sequence is exceedingly small. (If the same barcode ligates to two completely different sequences there is no concern). For a sequence diversity of -10 ^{8}, oligonucleotide synthesis with a 14-mer region containing at each position a randomly chosen nucleotide plus a common ~18-mer oligonucleotide 'tag' on the 5-end, comprising part of the Illumina sequencing adaptor is used. FIG. 3A demonstrates the steps involved. Randomers are synthesized by Trilink Technologies, http://www.trilinkbiotech.com/products/oligo/randomers.asp. An extension reaction is then used to synthesize the opposite strand, resulting to double-stranded adaptors that will be prepared for end repair and ligation to fragments obtained by genomic DNA fragmentation.

For fragmentation of genomic DNA an enzymatic approach using Shearase™ is applied, as previously reported [8], followed by enzymatic end-repair to obtain blunt-ended 5'P-containing dsDNA. The fragmentation is adjusted to create random genomic DNA fragments in the range 100-200 bp which is appropriate for mutation enrichment via qNaME-PrO. Ligation to double-stranded barcoded adaptors is done also via the Illumina protocol. Following ligation, a 10-cycle PCR pre-amplification using the common 'tag' is applied to yield multiple copies of allelic-families, each family containing its own barcode. Following target enrichment using customized target enrichment on beads (EZ-kit, Roche-Nimblegen), NaME-PrO followed by 10 cycles PCR using the common tags is applied to enrich mutation-containing fragments, along with their barcodes. The NGS-ready library is sequenced. Following qNAME-PRO, DNA is processed for paired-end sequencing using the Illumina MiSeq platform. Although the method is applicable to all platforms, the modest cost of MiSeq instrument makes it accessible to many groups that do not have access to a high-end facility, hence a combination of NaME-Pro with MiSeq was chosen for highest impact. Mutations within WT DNA comprise 5% down to 0.01% abundance, using serial dilutions of genomic DNA with defined mutations. Sensitivity and selectivity of NAME-PRO-NGS is evaluated in triplicate and compared to (a) using conventional PCR-NGS,

without molecular barcode corrections; and (b) using NGS corrected via molecular barcode technology. The molecular barcodes are employed as described below.

Strategy for deriving the original mutation abundance via allelic barcode-count diversity, 'abed'

The abed approach, described in FIG. 2, is a novel use of molecular barcodes to derive the true mutation abundance in the original DNA without being affected by potentially variable mutation enrichment during NaME-PrO. abed is a general strategy to retain strict quantification following any technology for mutation enrichment. Briefly, in the simplified example of FIG. 3A with 1 mutated allele in a total of 6, barcodes are attached via ligation before any other manipulation of the sample. Following 10 cycles amplification of barcoded alleles there are multiple copies of allelic families ('polonies'), each polony containing a distinct barcode (FIG. 3A). The mutation abundance at this stage is equal to the fraction of distinct barcodes with mutations over all distinct barcodes. When qNaME-PrO is applied, the majority (but not all[8]), WT DNA from each allelic family are rendered un-amplifiable, hence they are reduced following a subsequent 10-cycle PCR step (FIG. 3A). In contrast, allelic families with mutations are amplified and their frequency is increased by >100fold over allelic families containing WT. Despite enrichment of mutations over WT, the quantification of the original mutation abundance remains unaffected, as this is tied to the fraction of distinct barcodes with mutations over all distinct barcodes. Accordingly, during NGS the allelic barcode-count diversity, abed, is an absolute measure of the original mutation abundance.

FIG. 3B demonstrates the effectiveness of the abed approach. The workflow of Fig 2 was applied to a set of 33 DNA targets with mutations in circulating DNA from a breast cancer patient. Following sequencing and analysis, allelic frequencies (MAF, y axis) for each of the 33 targets (X-axis) were measured with (light grey bars) and without NaME-PrO enrichment (dark grey bars). The left panel of FIG. 3B shows MAF when the mutation quantification does not take into account the unique molecular barcodes on each molecule, i.e. without using the abed approach. The MAF quantification is distorted due to the NaME-PrO mutation enrichment, i.e. the MAF recorded after NaME-PrO application is very different from the true quantification of each mutated target as shown by the dark grey bars, which indicate the unbiased MAF. In contrast and as shown in the right panel of FIG. 3B, when the abed approach is implemented by counting unique molecular barcodes, the mutant target allele abundance is less distorted by

NaME-PrO enrichment. That is, in the right panel of FIG. 3B, the light grey and dark grey bars are closer together for most of the 33 targets (see comparison within boxes where the light grey and corresponding dark grey bars are almost equal). By counting unique molecular barcodes using the abed approach, the distortion of data caused by mutation enrichment is corrected or overcome.

Elimination of PCR/sequencing errors using molecular barcodes

In addition to using the molecular barcodes for deriving the original mutation abundance, the same barcodes are applied to eliminate PCR or sequencing errors, as reported[4-7]. In this approach[6], if a mutation is present in a given colony at least 90% of mutated fragments will be required to contain the same barcode for each polony. If this is not the case, the mutation is considered to be the result of a sequencing error or upstream PCR-error. Using this principle it is possible to reduce drastically the apparent sequencing errors during NGS4-7. Software to implement the abed approach to derive mutation abundance, as well as the reduction of errors via molecular barcodes is developed.

Example 3: qNaME-PrO for mutation enrichment in targeted genomic fractions corresponding to full-length exons of recurrently mutated genes in lung cancer

qNaME-PrO is adapted for 'mutation scanning' to enrich all possible mutations along 521 full exons of 139 genes that have been identified to be recurrently mutated in lung cancer[1, 2]. The methodologies of Example 1 are employed to validate qNaME-PrO application in full length genes.

Enrichment of multiple mutations over extended sequence regions by 'tiling' NaME-PrO oligonucleotide probes

It was demonstrated that in order to enrich multiple contiguous mutations along tumor suppressor genes like TP53 it is possible to 'tile' the overlapping 15bp region of NaME-PrO probes (FIG. 10A and [8]) and perform several NaME-PrO reactions in parallel, followed by mixing the products and performing a single sequencing run. It was demonstrated that a contiguous probe tiling arrangement (FIG. 10B) can enrich multiple mutations along a (longer) 50bp DNA section. This contiguous probe arrangement results to enrichments of 30-40-fold and

provides a balance between mutation enrichment and number of probes needed to cover a given DNA fragment. Of note, lung CA gene fusions (e.g., EML4-ALK fusion [1]) that result to breakpoints within the mapped exons are also enriched by NaME-PrO, since probes corresponding to the WT cfDNA fragment do not bind adequately the fusion breakpoint and DSN will not digest those fragments, resulting to enrichment.

qNaME-PrO application to full exons of genes significantly mutated in lung cancer

Following barcode ligation, 10-cycle pre-amplification and bead-based target enrichment as per FIG. 2 is performed, and the sample is split in 4 reactions. Each reaction separately employs qNaME-PrO using distinct groups of probes enriching mutations along sequential 50-bp-long sequence regions, as per FIG. 10B, thereby 'scanning' for mutation enrichment over the entire length of all targets. For example, any given 50bp regions within a ~150bp cfDNA fragment is enriched for mutations in one of the 4 distinct groups of probes. This is then tagged with a separate Illumina barcode-adaptor. All 4 NaME-PrO reactions are mixed and sequenced in a single MiSeq sequencing run.

The general methodologies employed in Example 2 is also applied in this Example to compare qNaME-PrO-NGS with current NGS approaches. Thus, sensitivity and selectivity of NAME-PRO-NGS sequencing is evaluated in triplicate and compared to (a) using conventional PCR-NGS, without molecular barcode corrections; and (b) using NGS corrected via molecular barcode technology.

Other considerations

(a) Amount of starting cfDNA material needed and anticipated lowest mutation abundance detectable via qNaME-PrO-NGS. Three main barriers limit NGS-based rare variant detection: **(i)** the intrinsic error frequency of high-throughput sequencing. As reported [36] using molecular barcode corrections allows mutation abundances as low as 0.01% to be discriminated from noise. **(ii)** The number of reads a sequencing platform can produce. As per FIG. 11 by using mutation enrichment via NaME-PrO the efficiency of sequencing increases by >100fold with equivalent gains in throughput and cost. Therefore items (i) and (ii) are uniquely addressed by the novel qNaME-PrO technology that is applied in this work. **(iii)** The amount of starting material needed for carrying out the protocol in FIG. 2 is ~30ng of cfDNA, corresponding to about 10,000

genomic copies. Considering that part of this cfDNA is highly fragmented and not readily amplifiable, it is estimated conservatively that about 3,000 genomic copies will make it to the mutation enrichment stage followed by sequencing. This enables mutation detection down to 0.03%, in good agreement with breast CA cfDNA exome-sequenced samples described above, FIG. 9. We anticipate using 30-60ng cfDNA to enable mutation detection to 0.01-0.03%.

(b) Presence of SNPs in the same target region with a mutation. If there are SNPs within a few bp from targeted somatic mutations and within the overlap region of either the upper or lower strand probes, during NaME-PrO, such SNPs may prevent DSN from digesting WT samples effectively, thereby leading to reduced mutation enrichment. To circumvent this hurdle, provided these SNPs are included in public databases, additional probes that match each of SNP versions are included in the assay[8], such that irrespective of what SNP is present the WT sample will always have a fully-matched probe[8]. Furthermore, since normal lymphocyte DNA is available in the samples, presence of 'private' germline mutations will be accounted for and additional NaME-PrO probes can be included on a sample-to-sample basis, as needed.

Example 4: tumor fingerprints in plasma and application to plasma circulating DNA

Tumor and circulating-DNA collected during the course of radio-chemo-therapy from a cohort of lung cancer patients are screened via qNaME-PrO-NGS, targeting both hotspot mutations as well as full length tumor suppressor gene and oncogene lung CA mutations. By tracing mutations over numerous DNA targets, the advantage of following multiple mutations rather than a single mutation in circulating-DNA is evaluated, thus circumventing biomarker sensitivity and specificity problems associated with tumor heterogeneity, uneven cfDNA release, or occult metastatic sites.

Application of NAME-PRO-NGS uniquely enables maximum efficiency sequencing for tracing of the mutated portion of the circulating-DNA in sequential samples obtained before, during and after completion of radio-chemo-therapy of lung cancer patients with few sequencing reads. Plasma-circulating DNA at regular time points obtained during radio-chemo-therapy from 100 lung cancer patients is analyzed via qNaME-PrO-NGS and the results assessed for concordance with the mutations identified in tumor tissue from each patient. Quantitative tracing

of tumor DNA in the circulation during therapy can help assess therapeutic response[19, 20, 58]. Furthermore, periodic quantitative tracing of tumor DNA after completion of therapy can be used for surveillance, monitoring tumor remission or early detection of relapse[34].

The approach described in FIG. 1 is followed by omitting the fragmentation step since cfDNA is fragmented. The anticipated advantage of tracing the entire tumor mutational fingerprint at mutation abundances down to 0.01%-0.1%, as opposed to tracing a single mutation in plasma at the same lowest abundance, will be examined systematically in the sequential plasma samples. Reproducibility will be assessed via triplicate independent experiments from the same plasma samples. Mutations in key genes is independently verified by genotyping techniques (HRM, ddPCR). Concordance of plasma-identified mutations with the primary tumor will help assess the hypothesis that DNA from different tumor clones enters circulation at different stages[46, 48] during therapy.

Statistical analysis

The ability of qNAME-PRO to identify a tumor signature in serial plasma samples from 100 lung cancer patients with paired tumor tissue is examined. The advantage of tracing a tumor fingerprint over tracing a single mutation (current practice) is assessed. This is interrogated in several ways. The qNaME-PrO-NGS approach as well as standard NGS to identify a tumor signature in plasma is used. It is assumed that the performance in tumor tissue represents the best outcome for plasma, and that either method may lose sensitivity. For this initial assessment, it is determined whether mutations were identified. The McNemar test is used to assess the ability of qNAME-PRO-NGS to identify mutations that are not identified by standard NGS. The mutations identified in tumor tissue are compared with those identified in plasma. It is anticipated that in doing this with NGS, it may be a binomial outcome—mutations detected in tumor may not be found in plasma due to sensitivity issues. The McNemar test is again used to assess the sensitivity of NGS in the plasma. With qNAME-PRO-NGS, the findings in plasma are compared with those in tumor, to identify when the full mutational fingerprint can be identified in plasma and when only a partial fingerprint can be found. Whether the number of mutations identified by qNAME-PRO changes over time in the serial samples is also examined. Each assessment will also be examined as part of a longitudinal model in which it is possible to

detect tumor mutations in plasma by both, one, or neither method over time. This is modeled by using generalized estimating equations, allowing inclusion of patient characteristics in the model.

Power calculation

It is desired to compare the ability to identify a tumor fingerprint in plasma samples based on a single mutation compared with multiple mutations. This data is analyzed using the McNemar test for paired assessments on the sample. For the purposes of calculation, it is assumed that the tumor fingerprint based on a single mutation will be detected with probability 0.50, and that the fingerprint based on multiple mutations will be detected 80% of the time. It is further assumed that half the patients who are detectable based on multiple mutations are also detectable based on single mutations, and therefore 40 samples will be detectable by multiple mutations only. It is assumed as well that 10 samples will be detectable only by the single mutation, and will be missed by the multiple mutations. In this setting, for the 100 paired samples that will be used the McNemar test would have 99% power testing at the 0.05 significance level to find that the multiple mutation technique is more sensitive. If it is assumed that either better performance for the multiple mutation approach or poorer performance for the single mutation approach will occur, it is found that the constraints of the 2 x 2 table force even greater divergence between the rate of discordant pairs, thereby increasing the power of the design to more than 99% for a range of assumptions. For example even if neither approach identifies tumor in plasma samples from 45% of the patients (which corresponds to an identification rate of 50% for the full fingerprint and 20% for the single marker), there is still 99% power to identify superiority of the full fingerprint with 100 samples. Even the scenario in which the single marker identifies disease in plasma 70% of the time, compared to 90% for the full fingerprint, has better than 99% power with 100 samples.

References for Examples 1, 2 and 3

[1] Newman AM, Bratman SV, To J, Wynne JF, Eclov NC, Modlin LA, Liu CL, Neal JW, Wakelee HA, Merritt RE, Shrager JB, Loo BW, Jr., Alizadeh AA, Diehn M: An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. Nat Med 2014, 20:548-54.

[2] Newman AM, Lovejoy AF, Klass DM, Kurtz DM, Chabon JJ, Scherer F, Stehr H, Liu CL, Bratman SV, Say C, Zhou L, Carter JN, West RB, Sledge GW, Jr., Shrager JB, Loo BW, Jr., Neal JW, Wakelee HA, Diehn M, Alizadeh AA: Integrated digital error suppression for improved detection of circulating tumor DNA. Nat Biotechnol 2016, 34:547-55.

[3] Adalsteinsson: Recent literature: listed in bty://personal.broadiostirui:e.org/viktor/publications.htrnl. 2016.

[4] Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B: Detection and quantification of rare mutations with massively parallel sequencing. Proc Natl Acad Sci U S A 201 1, 108:9530-5.
[5] Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA: Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 2012.
[6] Gregory MT, Bertout JA, Ericson NG, Taylor SD, Mukherjee R, Robins HS, Drescher CW, Bielas JH: Targeted single molecule mutation detection with massively parallel sequencing. Nucleic Acids Res 2016, 44:e22.
[7] Jee J, Rasouly A, Shamovsky I, Akivis Y, S RS, Mishra B, Nudler E: Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. Nature 2016.
[8] Song C, Liu Y, Fontana R, Makrigiorgos A, Mamon H, Kulke MH, Makrigiorgos GM: Elimination of unaltered DNA in mixed clinical samples via nuclease-assisted minor-allele enrichment. Nucleic Acids Res 2016.
[9] Song C, Liu Y, Fontana R, Makrigiorgos A, Mamon H, Kulke MH, Makrigiorgos GM: Elimination of un-altered DNA in clinical samples via nuclease-assisted minor allele enrichment. Nucleic Acids Res, In Press 2016.
[10] Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, Thornton K, Agrawal N, Sokoll L, Szabo SA, Kinzler KW, Vogelstein B, Diaz LA, Jr.: Circulating mutant DNA to assess tumor dynamics. Nat Med 2008, 14:985-90.
[11] Thomas RK, Nickerson E, Simons JF, Janne PA, Tengs T, Yuza Y, Garraway LA, Laframboise T, Lee JC, Shah K, O'Neill K, Sasaki H, Lindeman N, Wong KK, Borras AM, Gutmann EJ, Dragnev KH, Debiasi R, Chen TH, Glatt KA, Greulich H, Desany B, Lubeski CK, Brockman W, Alvarez P, Hutchison SK, Leamon JH, Ronan MT, Turenchalk GS, Egholm M, Sellers WR, Rothberg JM, Meyerson M: Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. Nat Med 2006, 12:852-5.
[12] Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, Severson T, Taylor GA, Teschendorff AE, Tse K, Turashvili G, Varhol R, Warren RL, Watson P, Zhao Y, Caldas C, Huntsman D, Hirst M, Marra MA, Aparicio S: Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature 2009, 461:809-13.
[13] Reis-Filho J: Next-generation sequencing. Breast Cancer Research 2009, 11:S12.
[14] Milbury CA, Correll M, Quackenbush J, Rubio R, Makrigiorgos GM: COLD-PCR enrichment of rare cancer mutations prior to targeted amplicon resequencing. Clin Chem 2012, 58:580-9.
[15] Gray J: Cancer: Genomics of metastasis. Nature 2010, 464:989-90.
[16] Tang S, Huang T: Characterization of mitochondrial DNA heteroplasmy using a parallel sequencing system. Biotechniques, 48:287-96.
[17] Qin W, Kozlowski P, Taillon BE, Bouffard P, Holmes AJ, Janne P, Camposano S, Thiele E, Franz D, Kwiatkowski DJ: Ultra deep sequencing detects a low rate of mosaic mutations in tuberous sclerosis complex. Hum Genet, 127:573-82.
[18] Schmitt MW, Kennedy SR, Salk JJ, Fox EJ, Hiatt JB, Loeb LA: Detection of ultra-rare mutations by next-generation sequencing. Proc Natl Acad Sci U S A 2012, 109: 14508-13.
[19] Misale S, Yaeger R, Hobor S, Scala E, Janakiraman M, Liska D, Valtorta E, Schiavo R, Buscarino M, Siravegna G, Bencardino K, Cercek A, Chen CT, Veronese S, Zanon C, Sartore-Bianchi A, Gambacorta M, Gallicchio M, Vakiani E, Boscaro V, Medico E, Weiser M, Siena S, Di Nicolantonio F, Solit D, Bardelli A: Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. Nature 2012, 486:532-6.
[20] Diaz LA, Jr., Williams RT, Wu J, Kinde I, Hecht JR, Berlin J, Allen B, Bozic I, Reiter JG, Nowak MA, Kinzler KW, Oliner KS, Vogelstein B: The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. Nature 2012, 486:537-40.
[21] Kuang Y, Rogers A, Yeap BY, Wang L, Makrigiorgos M, Vetrand K, Thiede S, Distel RJ, Janne PA: Noninvasive detection of EGFR T790M in gefitinib or erlotinib resistant non-small cell lung cancer. Clin Cancer Res 2009, 15:2630-6.
[22] Taniguchi K, Uchida J, Nishino K, Kumagai T, Okuyama T, Okami J, Higashiyama M, Kodama K, Imamura F, Kato K: Quantitative detection of EGFR mutations in circulating tumor DNA derived from lung adenocarcinomas. Clin Cancer Res 201 1, 17:7808-15.
[23] Girotti MR, Gremel G, Lee R, Galvani E, Rothwell D, Viros A, Mandal AK, Lim KH, Saturno G, Furney SJ, Baenke F, Pedersen M, Rogan J, Swan J, Smith M, Fusi A, Oudit D, Dhomen N, Brady G, Lorigan P, Dive C, Marais R: Application of Sequencing, Liquid Biopsies, and Patient-Derived Xenografts for Personalized Medicine in Melanoma. Cancer Discov 2016, 6:286-99.
[24] Thress KS, Paweletz CP, Felip E, Cho BC, Stetson D, Dougherty B, Lai Z, Markovets A, Vivancos A, Kuang Y, Ercan D, Matthews SE, Cantarini M, Barrett JC, Janne PA, Oxnard GR: Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. Nat Med 2015, 21:560-2.

[25] Oxnard GR, Paweletz CP, Kuang Y, Mach SL, O'Connell A, Messineo MM, Luke JJ, Butaney M, Kirschmeier P, Jackman DM, Janne PA: Noninvasive Detection of Response and Resistance in EGFR-Mutant Lung Cancer Using Quantitative Next-Generation Genotyping of Cell-Free Plasma DNA. Clin Cancer Res 2014.
[26] Schwaederle M, Husain H, Fanta PT, Piccioni DE, Kesari S, Schwab RB, Patel SP, Harismendy O, Ikeda M, Parker BA, Kurzrock R: Use of Liquid Biopsies in Clinical Oncology: Pilot Experience in 168 Patients. Clin Cancer Res 2016.
[27] Thierry AR, Mouliere F, El Messaoudi S, Mollevi C, Lopez-Crapez E, Rolet F, Gillet B, Gongora C, Dechelotte P, Robert B, Del Rio M, Lamy PJ, Bibeau F, Nouaille M, Loriot V, Jarrousse AS, Molina F, Mathonnet M, Pezet D, Ychou M: Clinical validation of the detection of KRAS and BRAF mutations from circulating tumor DNA. Nat Med 2014, 20:430-5.
[28] Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, Dunning MJ, Gale D, Forshew T, Mahler-Araujo B, Rajan S, Humphray S, Becq J, Halsall D, Wallis M, Bentley D, Caldas C, Rosenfeld N: Analysis of circulating tumor DNA to monitor metastatic breast cancer. N Engl J Med 2013, 368: 1199-209.
[29] Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, Shovlin M, Jaffe ES, Staudt LM, Lai C, Steinberg SM, Chen CC, Zheng J, Willis TD, Faham M, Wilson WH: Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. The lancet oncology 2015, 16:541-9.
[30] Higgins MJ, Jelovac D, Barnathan E, Blair B, Slater S, Powers P, Zorzi J, Jeter SC, Oliver GR, Fetting J, Emens L, Riley C, Stearns V, Diehl F, Angenendt P, Huang P, Cope L, Argani P, Murphy KM, Bachman KE, Greshock J, Wolff AC, Park BH: Detection of tumor PIK3CA status in metastatic breast cancer using peripheral blood. Clin Cancer Res 2012, 18:3462-9.
[3 1] Murphy DM, Bejar R, Stevenson K, Neuberg D, Shi Y, Cubrich C, Richardson K, Eastlake P, Garcia-Manero G, Kantarjian H, Ebert BL, Mike Makrigiorgos G: NRAS mutations with low allele burden have independent prognostic significance for patients with lower risk myelodysplastic syndromes. Leukemia 2013, 27:2077-8 1.
[32] Reckamp KL, Melnikova VO, Karlovich C, Sequist LV, Camidge DR, Wakelee H, Perol M, Oxnard GR, Kosco K, Croucher P, Samuelsz E, Vibat CR, Guerrero S, Geis J, Berz D, Mann E, Matheny S, Rolfe L, Raponi M, Erlander MG, Gadgeel S: A Highly Sensitive and Quantitative Test Platform for Detection of NSCLC EGFR Mutations in Urine and Plasma. J Thorac Oncol 2016.
[33] Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW: High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Anal Chem 201 1, 83:8604-10.
[34] Diaz LA, Jr., Bardelli A: Liquid biopsies: genotyping circulating tumor DNA. J Clin Oncol 2014, 32:579-86.
[35] Li J, Wang L, Mamon H, Kulke MH, Berbeco R, Makrigiorgos GM: Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. Nat Med 2008, 14:579-84.
[36] Milbury CA, Li J, Makrigiorgos GM: Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. Nucleic Acids Res 201 1, 39:e2.
[37] How-Kit A, Lebbe C, Bousard A, Daunay A, Mazaleyrat N, Daviaud C, Mourah S, Tost J: Ultrasensitive detection and identification of BRAF V600 mutations in fresh frozen, FFPE, and plasma samples of melanoma patients by E-ice-COLD-PCR. Anal Bioanal Chem 2014, 406:55 13-20.
[38] Narayan A, Carriero NJ, Gettinger SN, Kluytenaar J, Kozak KR, Yock TI, Muscato NE, Ugarelli P, Decker RH, Patel AA: Ultrasensitive measurement of hotspot mutations in tumor DNA in blood using error-suppressed multiplexed deep sequencing. Cancer Res 2012, 72:3492-8.
[39] Forshew T, Murtaza M, Parkinson C, Gale D, Tsui DW, Kaper F, Dawson SJ, Piskorz AM, Jimenez-Linan M, Bentley D, Hadfield J, May AP, Caldas C, Brenton JD, Rosenfeld N: Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. Sci Transl Med 2012, 4:136ra68.
[40] Lou DI, Hussmann JA, McBee RM, Acevedo A, Andino R, Press WH, Sawyer SL: High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing. Proc Natl Acad Sci U S A 2013, 110: 19872-7.
[41] Shagin DA, Rebrikov DV, Kozhemyako VB, Altshuler IM, Shcheglov AS, Zhulidov PA, Bogdanova EA, Staroverov DB, Rasskazov VA, Lukyanov S: A novel method for SNP detection using a new duplex-specific nuclease from crab hepatopancreas. Genome Res 2002, 12: 1935-42.

[42] Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C: Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 2009, 27: 182-9.
[43] Mertes F, Elsharawy A, Sauer S, van Helvoort JM, van der Zaag PJ, Franke A, Nilsson M, Lehrach H, Brookes AJ: Targeted enrichment of genomic DNA regions for next-generation sequencing. Briefings in functional genomics 201 1, 10:374-86.
[44] Wagner PD, Verma M, Srivastava S: Challenges for Biomarkers in Cancer Detection. Annals of the New York Academy of Sciences 2004, 1022:9-16.
[45] Kopreski MS, Benko FA, Borys DJ, Khan A, McGarrity TJ, Gocke CD: Somatic mutation screening: identification of individuals harboring K-ras mutations with the use of plasma DNA. Journal of the National Cancer Institute 2000, 92:918-23.
[46] Alix-Panabieres C, Pantel K: Clinical Applications of Circulating Tumor Cells and Circulating Tumor DNA as Liquid Biopsy. Cancer Discov 2016, 6:479-91.
[47] Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, Bartlett BR, Wang H, Luber B, Alani RM, Antonarakis ES, Azad NS, Bardelli A, Brem H, Cameron JL, Lee CC, Fecher LA, Gallia GL, Gibbs P, Le D, Giuntoli RL, Goggins M, Hogarty MD, Holdhoff M, Hong SM, Jiao Y, Juhl HH, Kim JJ, Siravegna G, Laheru DA, Lauricella C, Lim M, Lipson EJ, Marie SK, Netto GJ, Oliner KS, Olivi A, Olsson L, Riggins GJ, Sartore-Bianchi A, Schmidt K, Shih 1M, Oba-Shinjo SM, Siena S, Theodorescu D, Tie J, Harkins TT, Veronese S, Wang TL, Weingart JD, Wolfgang CL, Wood LD, Xing D, Hruban RH, Wu J, Allen PJ, Schmidt CM, Choti MA, Velculescu VE, Kinzler KW, Vogelstein B, Papadopoulos N, Diaz LA, Jr.: Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci Transl Med 2014, 6:224ra24.
[48] Hibi K, Robinson CR, Booker S, Wu L, Hamilton SR, Sidransky D, Jen J: Molecular detection of genetic alterations in the serum of colorectal cancer patients. Cancer Res 1998, 58: 1405-7.
[49] Richardson AL, Iglehart JD: BEAMing up personalized medicine: mutation detection in blood. Clin Cancer Res 2012, 18:3209-1 1.
[50] Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, Kiezun A, Hammerman PS, McKenna A, Drier Y, Zou L, Ramos AH, Pugh TJ, Stransky N, Helman E, Kim J, Sougnez C, Ambrogio L, Nickerson E, Shefler E, Cortes ML, Auclair D, Saksena G, Voet D, Noble M, DiCara D, Lin P, Lichtenstein L, Heiman DI, Fennell T, Imielinski M, Hernandez B, Hodis E, Baca S, Dulak AM, Lohr J, Landau DA, Wu CJ, Melendez-Zajgla J, Hidalgo-Miranda A, Koren A, McCarroll SA, Mora J, Lee RS, Crompton B, Onofrio R, Parkin M, Winckler W, Ardlie K, Gabriel SB, Roberts CW, Biegel JA, Stegmaier K, Bass AJ, Garraway LA, Meyerson M, Golub TR, Gordenin DA, Sunyaev S, Lander ES, Getz G: Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013, 499:214-8.
[5 1] Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J: Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res 2013, 23:843-54.
[52] Flaherty P, Natsoulis G, Muralidharan O, Winters M, Buenrostro J, Bell J, Brown S, Holodniy M, Zhang N, Ji HP: Ultrasensitive detection of rare mutations using next-generation targeted resequencing. Nucleic Acids Res 2012, 40:e2.
[53] Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L: VarScan: variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 2009, 25:2283-5.
[54] Bansal V, Libiger O, Torkamani A, Schork NJ: Statistical analysis strategies for association studies involving rare variants. Nat Rev Genet 2010, 11:773-85.
[55] Bidard FC, Weigelt B, Reis-Filho JS: Going with the flow: from circulating tumor cells to DNA. Sci Transl Med 2013, 5:207psl4.
[56] Ozsolak F: Third-generation sequencing techniques and applications to drug discovery. Expert Opin Drug Discov 2012, 7:23 1-43.
[57] Song C, Castellanos-Rizaldos E, Bejar R, Ebert BL, Makrigiorgos GM: DMSO Increases Mutation Scanning Detection Sensitivity of High-Resolution Melting in Clinical Samples. Clin Chem 2015, 6 1:1354-62.
[58] Siravegna G, Mussolin B, Buscarino M, Corti G, Cassingena A, Crisafulli G, Ponzetti A, Cremolini C, Amatu A, Lauricella C, Lamba S, Hobor S, Avallone A, Valtorta E, Rospo G, Medico E, Motta V, Antoniotti C, Tatangelo F, Bellosillo B, Veronese S, Budillon A, Montagut C, Racca P, Marsoni S, Falcone A, Corcoran RB, Di Nicolantonio F, Loupakis F, Siena S, Sartore-Bianchi A, Bardelli A: Clonal evolution and resistance to EGFR blockade in the blood of colorectal cancer patients. Nat Med 2015.

OTHER EMBODIMENTS

All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

From the above description, one skilled in the art can easily ascertain the essential characteristics of the present disclosure, and without departing from the spirit and scope thereof, can make various changes and modifications of the disclosure to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

EQUIVALENTS

While several inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary meanings of the defined terms.

All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles "a" and "an," as used herein in the specification and in the claims, unless clearly indicated to the contrary, should be understood to mean "at least one."

The phrase "and/or," as used herein in the specification and in the claims, should be understood to mean "either or both" of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with "and/or" should be construed in the same fashion, i.e., "one or more" of the elements so conjoined. Other elements may optionally be present other than the elements specifically identified by the "and/or" clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to "A and/or B", when used in conjunction with open-ended language such as "comprising" can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, "or" should be understood to have the same meaning as "and/or" as defined above. For example, when separating items in a list, "or" or "and/or" shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as "only one of or "exactly one of," or, when used in the claims, "consisting of," will refer to the inclusion of exactly one element of a number or list of elements. In general, the term "or" as used herein shall only be interpreted as indicating exclusive alternatives (i.e. "one or the other but not both") when preceded by terms of exclusivity, such as "either," "one of," "only one of," or "exactly one of." "Consisting essentially of," when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase "at least one," in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase "at least one" refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, "at least one of A and B" (or, equivalently, "at least one of A or B," or, equivalently "at least one of A and/or B") can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as "comprising," "including," "carrying," "having," "containing," "involving," "holding," "composed of," and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases "consisting of and "consisting essentially of shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03. It should be appreciated that embodiments described in this document using an open-ended transitional phrase (e.g., "comprising") are also contemplated, in alternative embodiments, as "consisting of and "consisting essentially of the feature described by the open-ended transitional phrase. For example, if the disclosure describes "a composition comprising A and B", the disclosure also contemplates the alternative embodiments "a composition consisting of A and B" and "a composition consisting essentially of A and B".

What is claimed is:

CLAIMS

1. A method of determining original abundance of mutant alleles of target nucleic acid in a sample of genomic DNA, the method comprising

providing a sample of double-stranded genomic DNA fragments anticipated to comprise mutant alleles of target nucleic acid and wild-type alleles of the target nucleic acid, wherein each terminus of the genomic DNA fragments is attached to a unique double stranded barcode and a double stranded common sequence tag, wherein the common sequence tag is located upstream of the unique barcode;

amplifying a portion of the double-stranded genomic DNA fragments using primers that are complementary to the common sequence tag;

enriching in the amplified portion the mutant alleles of the target nucleic acids relative to the wild-type alleles of the target nucleic acids;

obtaining a measure of total number of unique barcodes associated with mutant target nucleic acid sequence in the enriched portion;

obtaining a measure of total number of unique barcodes associated with the sample,

wherein the original abundance of the mutant target nucleic acid in the sample of genomic DNA is represented by the ratio of the total number of unique barcodes associated with mutant target nucleic acid sequence to the total number of unique barcodes associated with the sample.

2. The method of claim 1, wherein the total number of unique barcodes associated with the sample is the total number of unique barcodes associated with mutant and wild-type target nucleic acid sequence in the enriched portion.

3. The method of claim 1, further comprising:

obtaining a measure of total number of unique barcodes associated with wild-type target nucleic acid sequence in an un-enriched control sample,

wherein the total number of unique barcodes associated with the sample is the sum of the total number of unique barcodes associated with mutant target nucleic acid sequence in the enriched portion and the total number of unique barcodes associated with wild-type target nucleic acid sequence in the un-enriched control sample.

4. The method of claim 3, further comprising amplifying the double-stranded genomic DNA fragments in the un-enriched control sample using primers that are complementary to the common sequence tag prior to obtaining a measure of total number of unique barcodes associated with wild-type target nucleic acid sequence in the un-enriched control sample.

5. The method of claim 3 or 4, wherein the enriched portion and the un-enriched control sample are sequenced in the same sequencing reaction.

6. The method of any one of claims 1-5, further comprising enriching both the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample.

7. The method of claim 6, wherein the enriching the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample comprises contacting the sample with one or more probes that specifically bind target nucleic acids.

8. The method of claim 7, wherein the one or more probes are bound to a solid substrate.

9. The method of claim 8, wherein the solid substrate comprises beads.

10. The method of any of the preceding claims, further comprising fragmenting genomic DNA to form the double-stranded genomic DNA fragments.

11. The method of claim 10, wherein the genomic DNA is fragmented using Shearase or any other DNA-ase type of enzyme that digests DNA randomly.

12. The method of claim 10 or 11, further comprising enzymatic end-repairing of the DNA fragments to obtain blunt ended 5' containing double-stranded DNA.

13. The method of any one of claims 1-9, further comprising forming the double-stranded genomic DNA fragments by multiplex PCR.

14. The method of any one of the preceding claims, wherein the method does not include obtaining a measure of total number of barcodes that are associated with non-target mutant nucleic acids or non-target wild-type nucleic acids in the enriched portion.

15. The method of any one of the preceding claims, further comprising obtaining sequences of the mutant target nucleic acids and wild-type target nucleic acids in the enriched portion.

16. The method of any one of the preceding claims, wherein the double-stranded genomic DNA fragments are 40-200bp in length.

17. The method of any one of the preceding claims, wherein the unique barcode is 8-14bp in length.

18. The method of any one of the preceding claims, wherein the common tag is 16-40 bp in length.

19. The method of any of the preceding claims, further comprising attaching the double-stranded genomic DNA fragments to the unique double-stranded barcodes.

20. The method of claim 19, wherein the attaching of double-stranded fragments of genomic DNA to the unique barcodes is performed at a ratio of 10^7-10^9 unique barcodes to lOOng double-stranded genomic DNA fragments present in the sample.

21. The method of claim 19 or 20, wherein the attaching comprises starting from a single-stranded barcode, synthesizing the opposite strand of the single-stranded barcode using an extension reaction to form a double stranded barcode, and then ligating to an end of the double-stranded genomic DNA fragment

22. The method of claim 19 or 20, wherein the attaching of barcodes comprises PCR using oligonucleotide primers, wherein each oligonucleotide primer comprises a common tag portion, a unique barcode portion, and a target-specific portion.

23. The method of claim 22, wherein the PCR is performed for 1-6 cycles.

24. The method of any one of the preceding claims, wherein the barcoded DNA fragments are amplified using PCR or COLD-PCR.

25. The method of claim 24, wherein the PCR is performed for 3-30 cycles.

26. The method of any one of the preceding claims wherein the common tag is biotinylated.

27. The method of claim 25, wherein the double-stranded genomic DNA fragments attached to barcodes and common tags are bound to streptavidin-beads.

28 The method of any one of the preceding claims, wherein the genomic DNA is obtained from a biological sample.

29. The method of claim 28, wherein the biological sample is selected from tissue, blood, plasma, serum, urine, saliva, and cerebrospinal fluid.

30. The method of claim 29, wherein the tissue sample is a tumor biopsy.

31. The method of claim 28, wherein the biological sample comprises circulating cell-free DNA.

32. The method of any one of the preceding claims, wherein the mutant target nucleic acid in the amplified portion is enriched by 2-300 fold relative to the wild-type target nucleic acid in the amplified portion.

33. The method of any of the preceding claims, wherein the enriching the mutant alleles of the target nucleic acids in the amplified portion relative to the wild-type alleles of the target nucleic acids in the amplified portioncomprises subjecting the amplified portion to one or more of the following: Nuclease-assisted Minor-allele Enrichment using Probe Overlap (NaME-PrO), Coamplification at Lower Denaturation temperature-PCR (COLD-PCR), Improved and Complete Enrichment COLD-PCR (ice-COLD-PCR), Temperature-Tolerant-ice-COLD-PCR (TT-ice-COLD-PCR), toehold PCR blocker-PCR using locked nucleic acids (LNA), peptide nucleic acids (PNA or XNA), or blockers with other modified nucleotides, CRISP-R-mediated mutation enrichment by selecting mutated alleles or removing wild-type alleles and Differential Strand Separation at Critical Temperature (DiSSECT).

34. The method of claim 33, wherein the NaMe-PrO comprises

(a) preparing an amplification reaction mixture comprising the double-stranded mutant and wild-type target nucleic acids, a thermostable double strand-specific nuclease (DSN), PCR amplification components, and a pair of oligonucleotide probes, one of which is complementary to the wild-type nucleic acid top strand and the other is complementary to the wild-type nucleic acid bottom strand, wherein the probes overlap each other by 10-15 probes such that the overlap coincides with the target region;

(b) subjecting the reaction mixture to a denaturing temperature to permit denaturation of the wild-type nucleic acid and the mutant target nucleic acid;

(c) reducing the temperature to permit hybridization of the probes to their corresponding sequences on the wild-type and mutant target nucleic acids thereby forming complementary wild-type-probe duplexes, wherein the DSN cleaves the complementary wild-type-probe duplexes but not the partially complementary target mutant -probe duplexes; and

(d) subjecting the reaction mixture to an amplification condition thereby enriching the uncleaved mutant target nucleic acid relative to the cleaved wild-type nucleic acid.

35. The method of claim 34, wherein the target region that coincides with the overlap of the probes comprises more than one mutation.

36. The method of claims 34 or 35, wherein the probes have a 3'-terminal polymerase block.

37. The method of any one of the claims 34-36, wherein the probes are complementary to SNPs near target mutations.

38. The method of claim 24 or 33, wherein the COLD-PCR comprises

(a) denaturing the double-stranded mutant and wild-type target nucleic acids by subjecting the double-stranded target mutant and wild-type nucleic acids to a first denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(b) forming a target mutant/wild-type strand duplex;

(c) denaturing said mutant/wild-type strand duplex by subjecting the nucleic acid amplified portion to a critical temperature (Tc) that is below the Tm of the wild-type nucleic acids;

(d) annealing a primer pair to the mutant and wild-type target nucleic acid strands; and

(e) extending said primer pair so as to enrich said mutant target sequence relative to said wild-type strand.

39. The method of clam 38, wherein (b) and (c) are repeated 1-19 times before performing (d).

40. The method of claims 38 or 39, wherein (a) to (e) are repeated 1-29 times.

41. The method of claim 33, wherein ice-COLD-PCR comprises

(a) exposing the mutant and wild-type target nucleic acids to a reference sequence that is complementary the target sequence;

(b) denaturing the double-stranded target mutant and wild-type nucleic acids by subjecting the double-stranded mutant and wild-type target nucleic acids to a first denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(c) forming a target mutant/reference strand and target wild-type/reference strand duplexes;

(d) denaturing said mutant/reference strand duplex by subjecting the nucleic acid amplified portion to a critical temperature (Tc) that is below the Tm of the wild-type/reference duplex;

(d) annealing a primer pair to the mutant and wild-type target nucleic acid strands; and

(e) extending said primer pair so as to enrich said mutant target sequence relative to said wild-type target nucleic acid.

42. The method of claim 33, wherein TT-ice-COLD-PCR comprises

(a) exposing the mutant and wild-type target nucleic acids to a reference sequence that is complementary the target sequence;

(b) denaturing the double-stranded target mutant and wild-type target nucleic acids by subjecting the double-stranded target mutant and wild-type nucleic acids to a denaturing temperature that is above the melting temperature of the wild-type nucleic acid;

(c) forming a target mutant/reference strand and target wild-type/reference strand duplexes;

(d) denaturing said mutant/reference strand duplex by subjecting the nucleic acid amplified portion to a first critical temperature (Tc) that is below the Tm of the wild-type/reference duplex;

(e) annealing a primer pair to the mutant and wild-type target nucleic acid strands;

(f) extending said primer pair so as to enrich said mutant target nucleic acid relative to said wild-type target nucleic acid; and

(f) repeating steps (d) to (f) at least once at a second Tc which is above the first Tc.

43. The method of claim 33, wherein the DiSSECT comprises:

(a) allowing mutant and wildtype target nucleic acids to bind to complementary probes which are immobilized to beads, wherein the probes resemble the wild-type nucleic acids;

(b) denaturing the target mutant/probe duplex by subjecting the nucleic acid amplified portion to a critical temperature such that the wild-type/probe duplex does not denature;

(c) collecting the eluate from the beads; and

(d) repeating at least once (a)-(c) using beads on which the probes are unbound to any nucleic acid.

44. The method of any one of the preceding claims, further comprising determining the original abundance of one or more additional mutant alleles of target nucleic acids at different loci on the genomic DNA.

45. The method of claim 44, wherein the mutant alleles of the target nucleic acids are enriched relative to wild-type alleles of the target nucleic acids by subjecting the amplified portion to multiplexed NaME-PrO.

46. The method of claim 45, wherein the multiplexed NaME-PrO uses overlapping probe pairs in parallel reactions, wherein each parallel reaction uses a distinct pair of overlapping probes, such that the totality of overlapping probe pairs span a contiguous region of genomic DNA of interest.

47. The method of claim 45, wherein the multiplexed NaME-PrO uses continuous and non-overlapping probe pairs in parallel reactions, wherein each parallel reaction uses a distinct pair of non-overlapping probes, such that the totality of non-overlapping probe pairs span a contiguous region of genomic DNA of interest.

48. A method of determining disease diagnosis or prognosis in a subject, the method comprising determining original abundance of a mutant target nucleic acid in a sample of genomic DNA according to the method of any one of claims 1-47, wherein the determining is repeated in two or more genomic DNA samples obtained from a subject over a period of time, and wherein an increase in original abundance of the target mutant nucleic acids in the sample over time is indicative of a worsening diagnosis or prognosis in the subject.

49. A method of determining efficacy of a therapeutic treatment in a subject, the method comprising determining original abundance of a mutant target nucleic acid in a sample of genomic DNA according to the method of any one of claims 1-47, wherein the determining is repeated in genomic DNA samples obtained from a subject before and after treatment, and wherein a decrease in original abundance of the target mutant nucleic acids in the sample after treatment is indicative of therapeutic efficacy in the subject.

50. The method of claim 48 or 49, wherein the determining is repeated in 2-10 genomic DNA samples obtained over a period of time.

51. The method of claim 48 or 49, wherein the determining is repeated over a period of time of 2 months to 2 years.

52. A method of obtaining a fingerprint of mutations associated with a disease, the method comprising determining original abundance of mutant alleles of target nucleic acids of a plurality of mutations according to the method of any one of claims 1-47, wherein the determining is performed on one or more samples of genomic DNA obtained from one or more subjects suffering from the disease.

53. The method of claim 6, wherein the enriching the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids in the sample comprises

amplifying the double-stranded genomic DNA fragments in the sample using a first and second primer, wherein the first primer is complementary to the double stranded common sequence tag, and the second primer is a nested gene-specific primer comprising a sequence tag that is different from the double-stranded common sequence tag.

54. The method of claim 53, wherein the enriching the mutant target nucleic acids and wild-type target nucleic acids relative to total nucleic acids is performed after the amplifying of the double-stranded genomic DNA fragments using primers that are complementary to the double-stranded common sequence tag.

55. The method of claim 53 or 54, wherein the amplifying the double-stranded genomic DNA fragments in the sample using a first and second primer is performed for 5-15 cycles.

56. The method of any one of claims 53-55, wherein the first primer comprises a sequencing adapter.

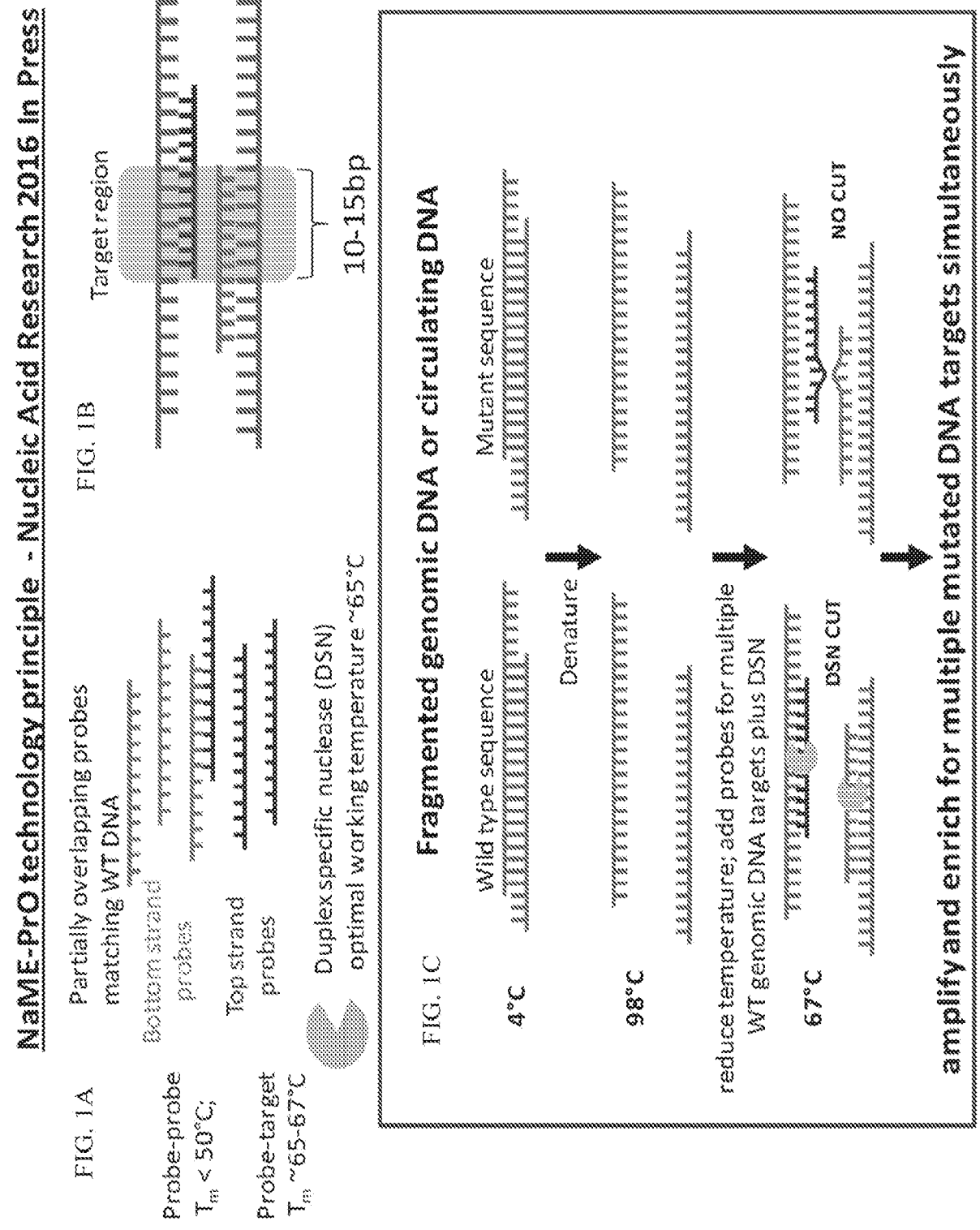

NaME-PrO technology principle - Nucleic Acid Research 2016 in Press
FIG. 1A
Partially overlapping probes matching WT DNA
Bottom strand probes
Top strand probes
Probe-probe T_m < 50°C;
Probe-target T_m ~65-67°C
Duplex specific nuclease (DSN)
optimal working temperature ~65°C
FIG. 1B
Target region
10-15bp
FIG. 1C
Fragmented genomic DNA or circulating DNA
Wild type sequence
Mutant sequence
4°C
Denature
98°C
reduce temperature; add probes for multiple WT genomic DNA targets plus DSN
67°C
DSN CUT
NO CUT
amplify and enrich for multiple mutated DNA targets simultaneously

FIG. 2

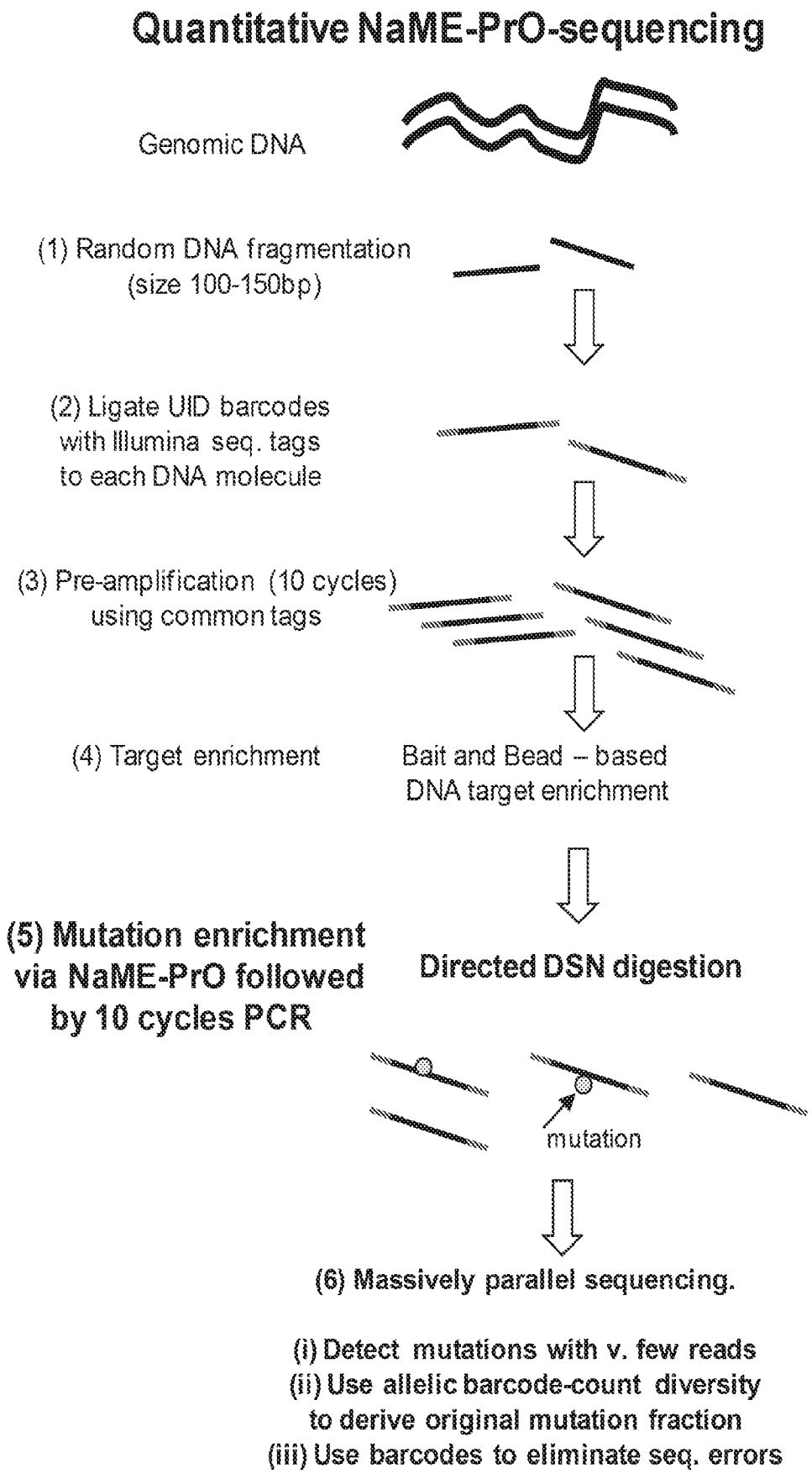

Quantitative NaME-PrO-sequencing
Genomic DNA
(1) Random DNA fragmentation (size 100-150bp)
(2) Ligate UID barcodes with Illumina seq. tags to each DNA molecule
(3) Pre-amplification (10 cycles) using common tags
(4) Target enrichment
Bait and Bead – based DNA target enrichment
(5) Mutation enrichment via NaME-PrO followed by 10 cycles PCR
Directed DSN digestion
mutation
(6) Massively parallel sequencing.
(i) Detect mutations with v. few reads
(ii) Use allelic barcode-count diversity to derive original mutation fraction
(iii) Use barcodes to eliminate seq. errors

FIG. 3A

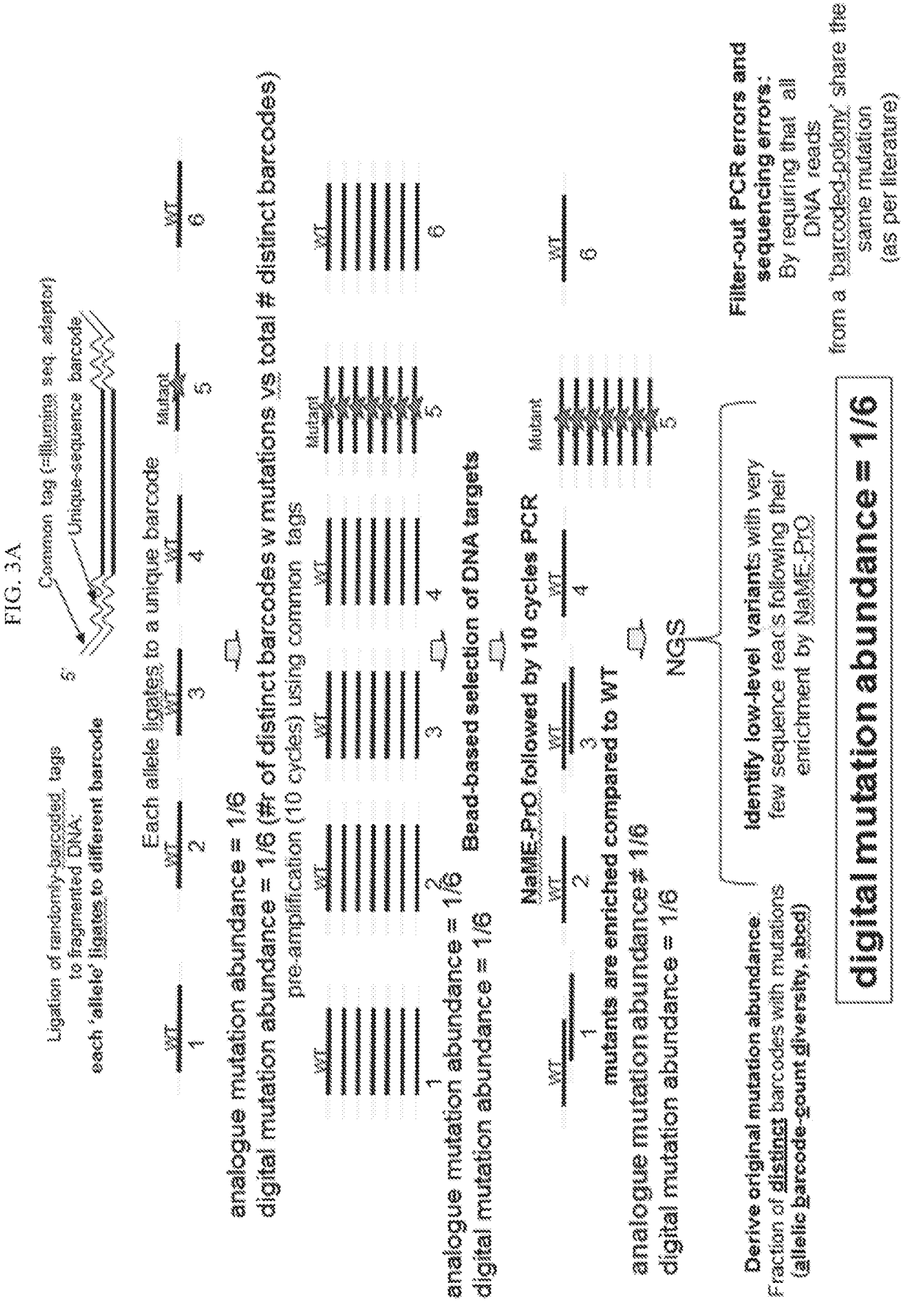

Ligation of randomly-barcoded tags to fragmented DNA: each 'allele' ligates to different barcode
Common tag (=Illumina seq. adaptor)
Unique-sequence barcode
5'
Each allele ligates to a unique barcode
WT
Mutant
analogue mutation abundance = 1/6
digital mutation abundance = 1/6 (#r of distinct barcodes w mutations vs total # distinct barcodes)
pre-amplification (10 cycles) using common tags
analogue mutation abundance = 1/6
digital mutation abundance = 1/6
Bead-based selection of DNA targets
NaME-PrO followed by 10 cycles PCR
mutants are enriched compared to WT
analogue mutation abundance≠ 1/6
digital mutation abundance = 1/6
NGS
Derive original mutation abundance:
Fraction of distinct barcodes with mutations
(allelic barcode-count diversity, abcd)
Identify low-level variants with very few sequence reads following their enrichment by NaME-PrO
Filter-out PCR errors and sequencing errors:
By requiring that all DNA reads
from a 'barcoded-colony' share the same mutation
(as per literature)
digital mutation abundance = 1/6

FIG. 3B

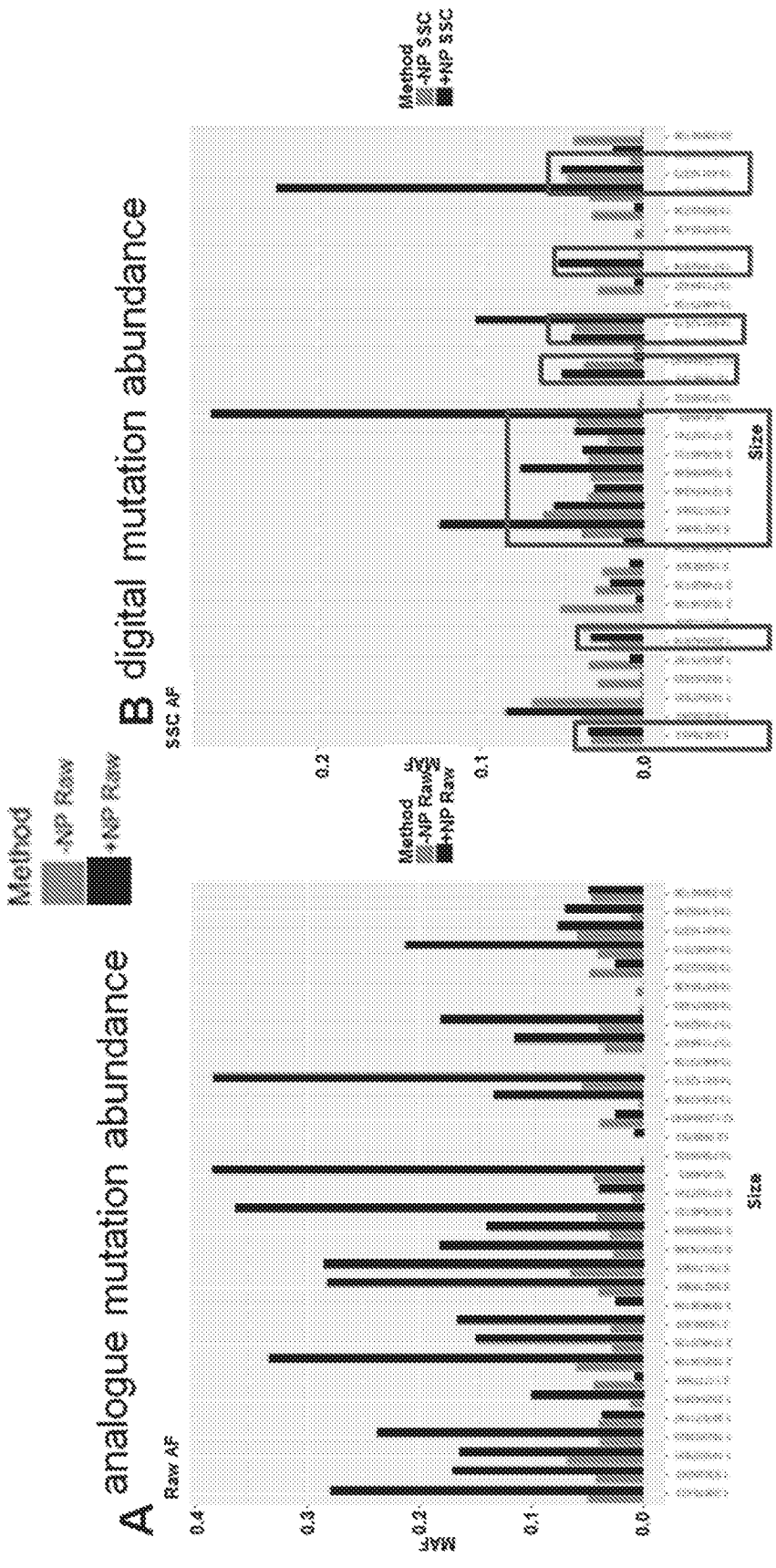

A analogue mutation abundance
Raw AF
0.4
0.3
0.2
0.1
0.0
Size
Method
-NP Raw
+NP Raw
B digital mutation abundance
SSC AF
0.2
0.1
0.0
Size
Method
-NP SSC
+NP SSC

FIG. 4

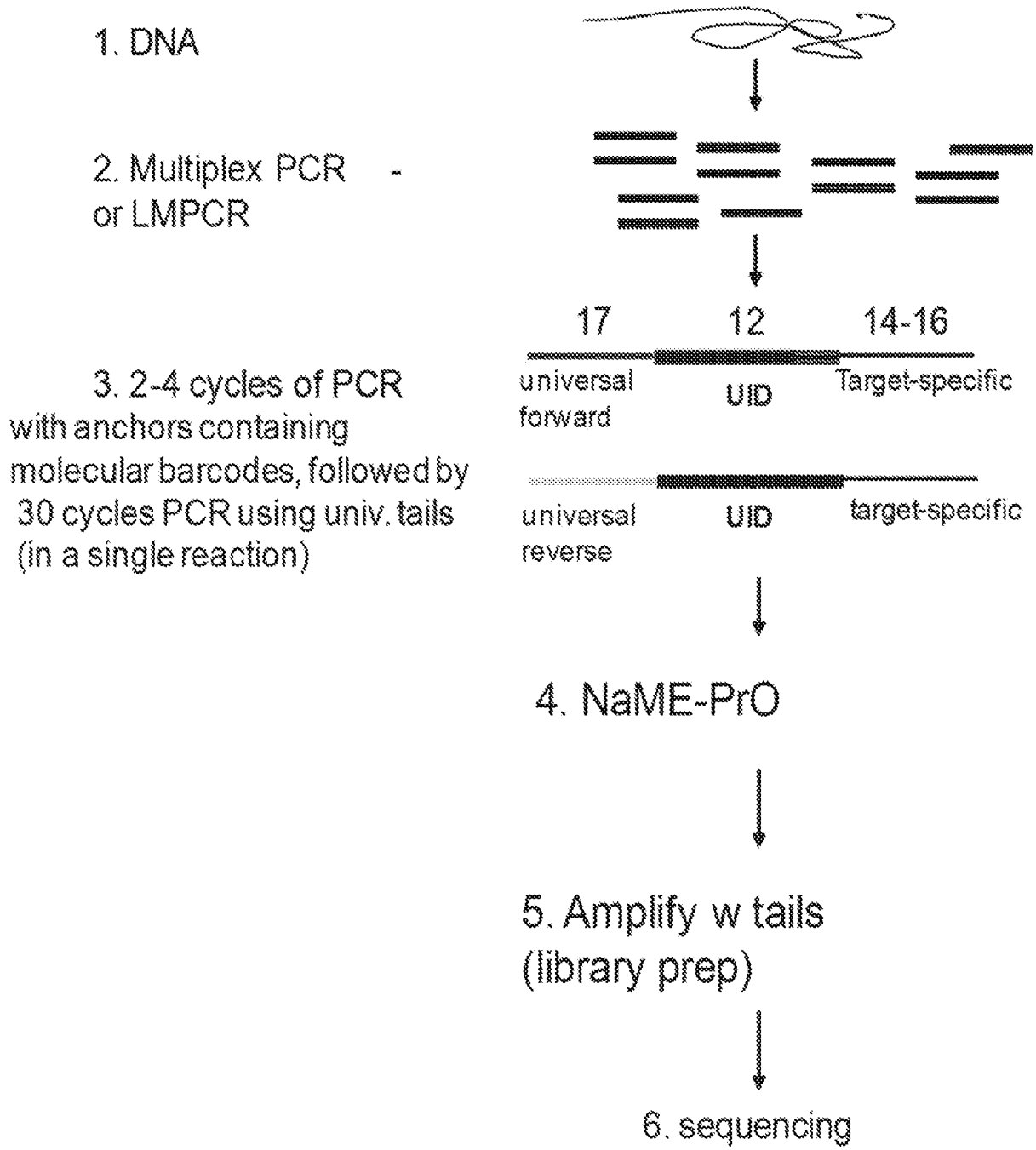

1. DNA
2. Multiplex PCR - or LMPCR
3. 2-4 cycles of PCR with anchors containing molecular barcodes, followed by 30 cycles PCR using univ. tails (in a single reaction)
17
12
14-16
universal forward
UID
Target-specific
universal reverse
UID
target-specific
4. NaME-PrO
5. Amplify w tails (library prep)
6. sequencing

FIG. 5

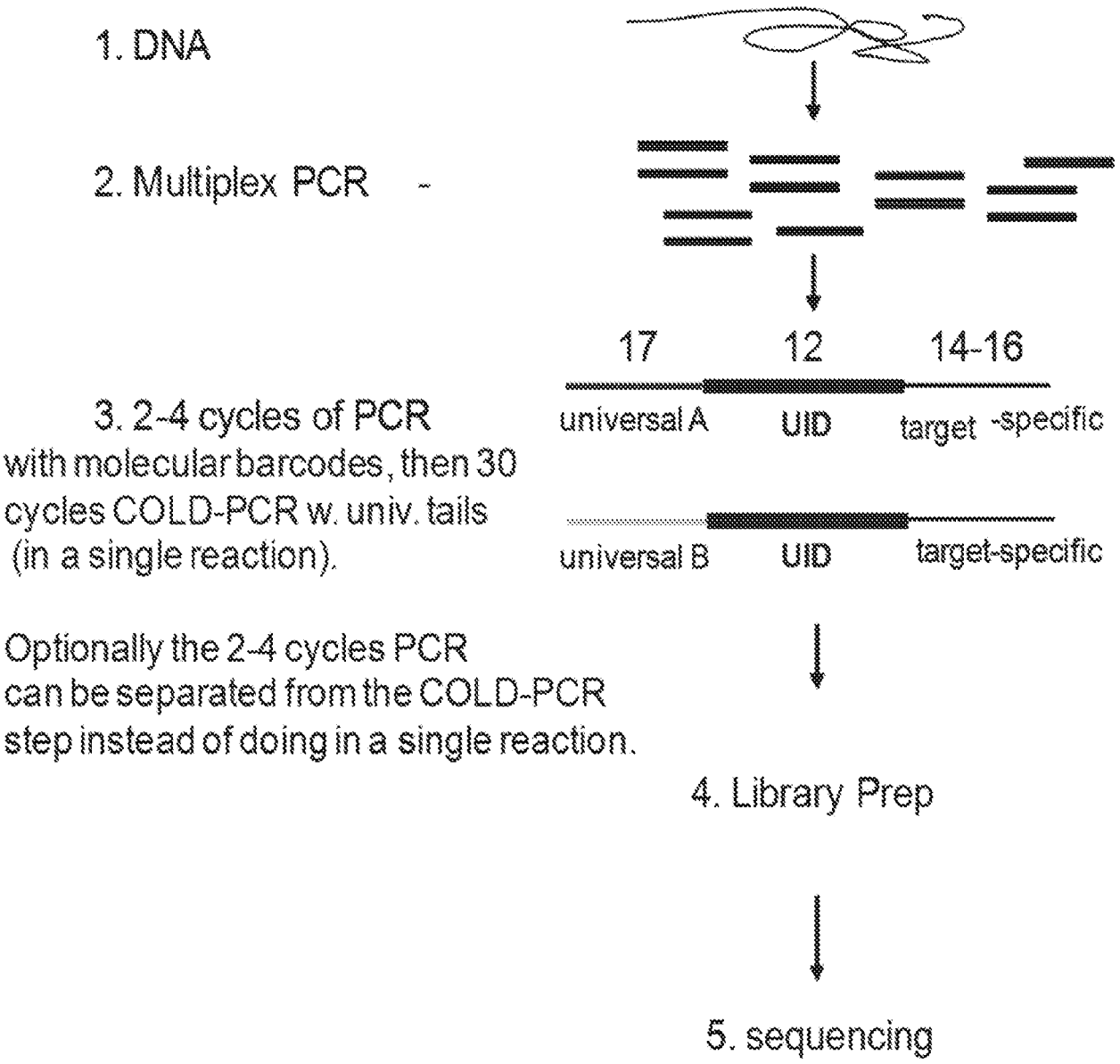

1. DNA
2. Multiplex PCR
17
12
14-16
universal A
UID
target -specific
3. 2-4 cycles of PCR with molecular barcodes, then 30 cycles COLD-PCR w. univ. tails (in a single reaction).
universal B
UID
target-specific
Optionally the 2-4 cycles PCR can be separated from the COLD-PCR step instead of doing in a single reaction.
4. Library Prep
5. sequencing

FIG. 6

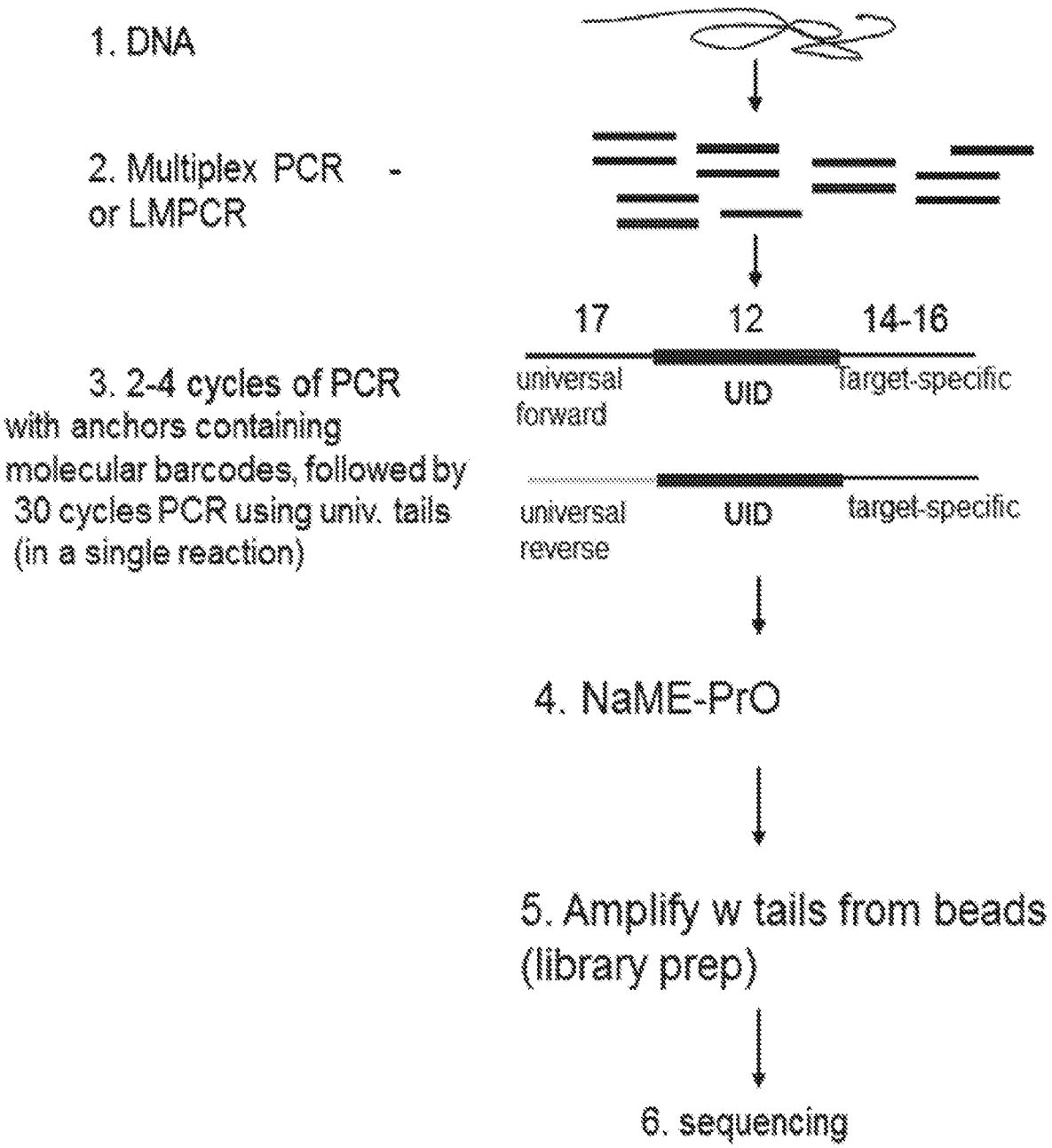

1. DNA
2. Multiplex PCR - or LMPCR
3. 2-4 cycles of PCR with anchors containing molecular barcodes, followed by 30 cycles PCR using univ. tails (in a single reaction)
17
12
14-16
universal forward
UID
target-specific
universal reverse
UID
target-specific
4. NaME-PrO
5. Amplify w tails from beads (library prep)
6. sequencing

FIG. 7A

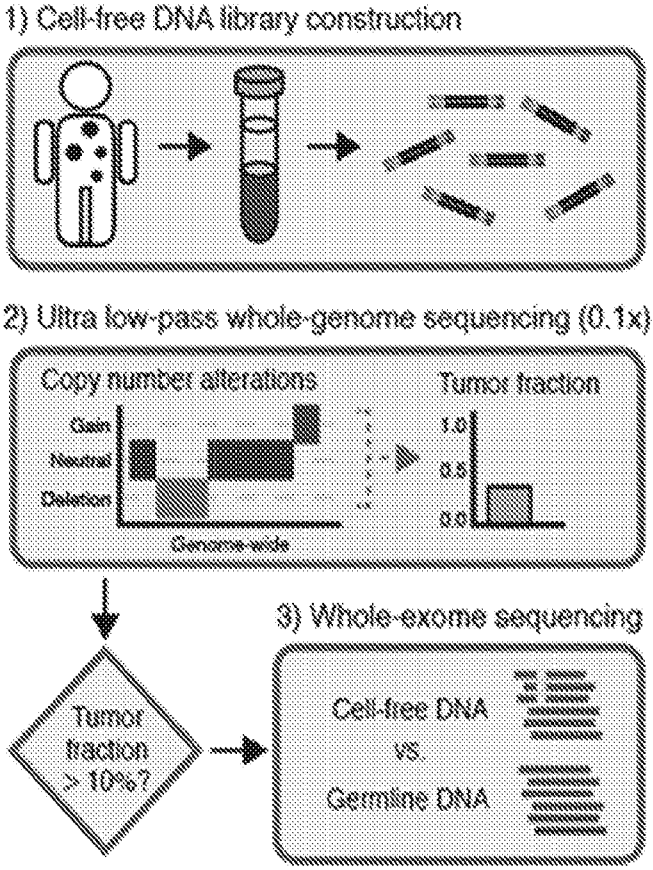

1) Cell-free DNA library construction
2) Ultra low-pass whole-genome sequencing (0.1x)
Copy number alterations
Gain
Neutral
Deletion
Genome-wide
Tumor fraction
1.0
0.5
0.0
3) Whole-exome sequencing
Tumor fraction > 10%?
Cell-free DNA
vs.
Germline DNA

FIG. 7B

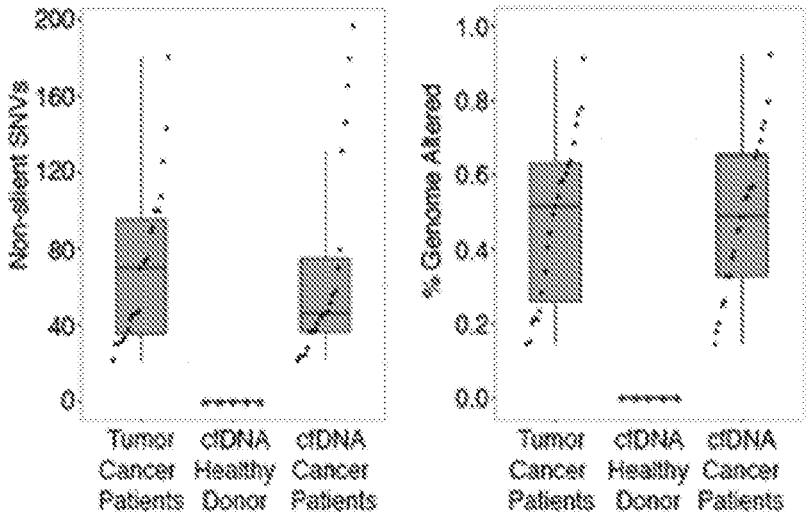

Non-silent SNVs
200
160
120
80
40
0
% Genome Altered
1.0
0.8
0.6
0.4
0.2
0.0
Tumor Cancer Patients
cfDNA Healthy Donor
cfDNA Cancer Patients

FIG. 7C

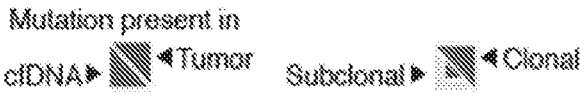
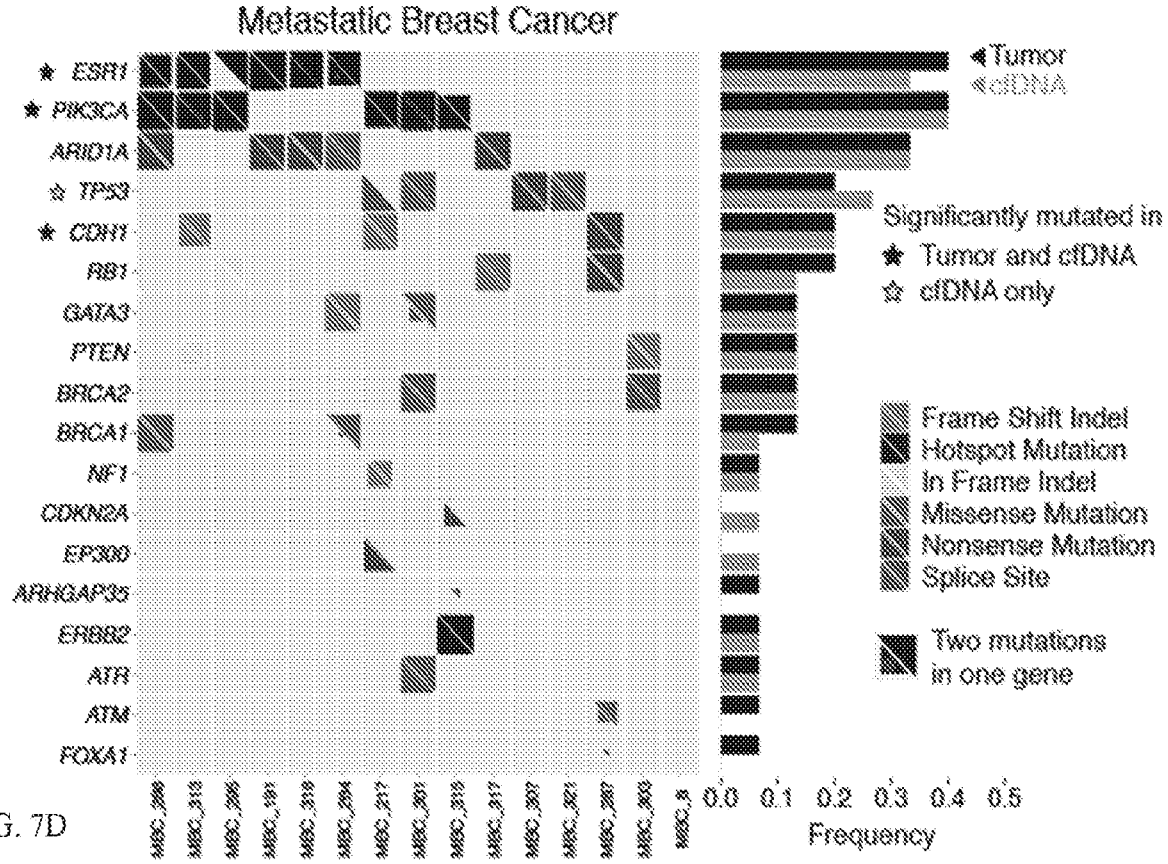
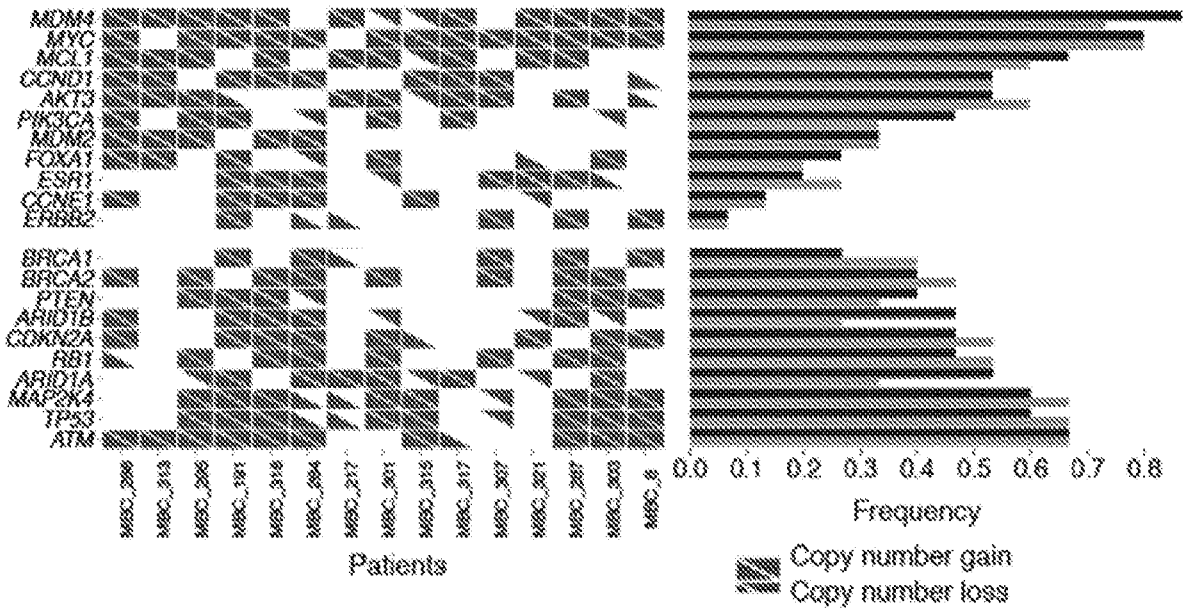

Mutation present in
cfDNA
Tumor
Subclonal
Clonal

FIG. 7D


Metastatic Breast Cancer
ESR1
PIK3CA
ARID1A
TP53
CDH1
RB1
GATA3
PTEN
BRCA2
BRCA1
NF1
CDKN2A
EP300
ARHGAP35
ERBB2
ATR
ATM
FOXA1
Tumor
cfDNA
Significantly mutated in
Tumor and cfDNA
cfDNA only
Frame Shift Indel
Hotspot Mutation
In Frame Indel
Missense Mutation
Nonsense Mutation
Splice Site
Two mutations in one gene
0.0 0.1 0.2 0.3 0.4 0.5
Frequency


MDM4
MYC
MCL1
CCND1
AKT3
PIK3CA
MDM2
FOXA1
ESR1
CCNE1
ERBB2
BRCA1
BRCA2
PTEN
ARID1B
CDKN2A
RB1
ARID1A
MAP2K4
TP53
ATM
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8
Frequency
Patients
Copy number gain
Copy number loss

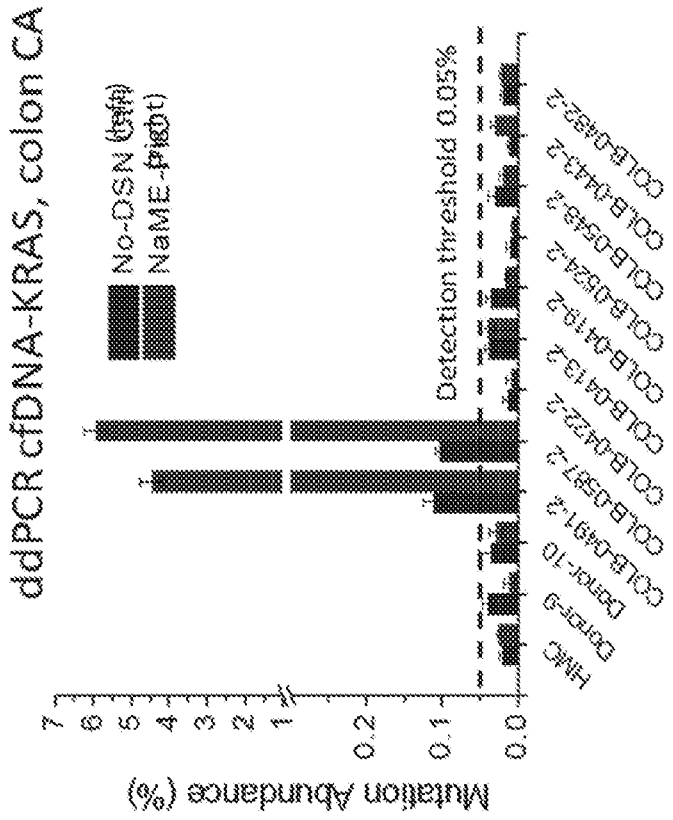

ddPCR cfDNA-KRAS, colon CA
No-DSN (left)
NaME-PrO (right)
Detection threshold 0.05%
Mutation Abundance (%)
7
6
5
4
3
2
1
0.2
0.1
0.0
HMC
Donor-9
Donor-10
COLB-0491-2
COLB-0587-2
COLB-0422-2
COLB-0413-2
COLB-0419-2
COLB-0524-2
COLB-0589-2
COLB-0444-2
COLB-0462-2

FIG. 8B

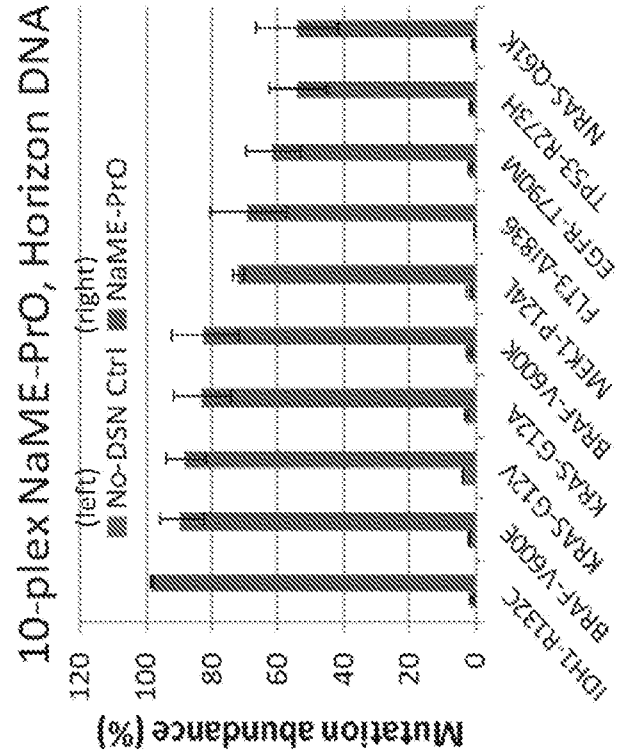

10-plex NaME-PrO, Horizon DNA
(left) (right)
No-DSN Ctrl NaME-PrO
Mutation abundance (%)
120
100
80
60
40
20
0
IDH1-R132C
BRAF-V600E
KRAS-G12V
KRAS-G12A
BRAF-V600K
MEK1-P124L
FLT3-D835
EGFR-T790M
TP53-R273H
NRAS-Q61K

FIG. 8A

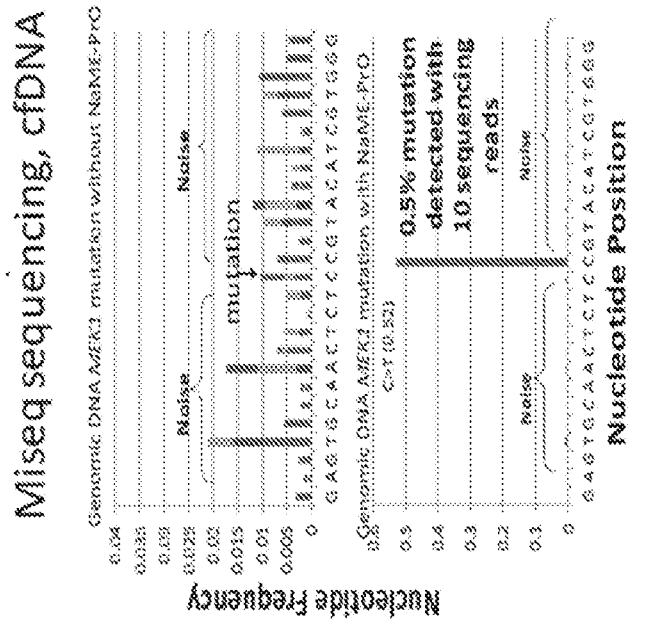

Miseq sequencing, cfDNA
Genomic DNA AZK1 mutation without NaME-PrO
Noise
mutation
Genomic DNA AZK1 mutation with NaME-PrO
0.5% mutation detected with 10 sequencing reads
Nucleotide Position
Nucleotide frequency

FIG. 8D

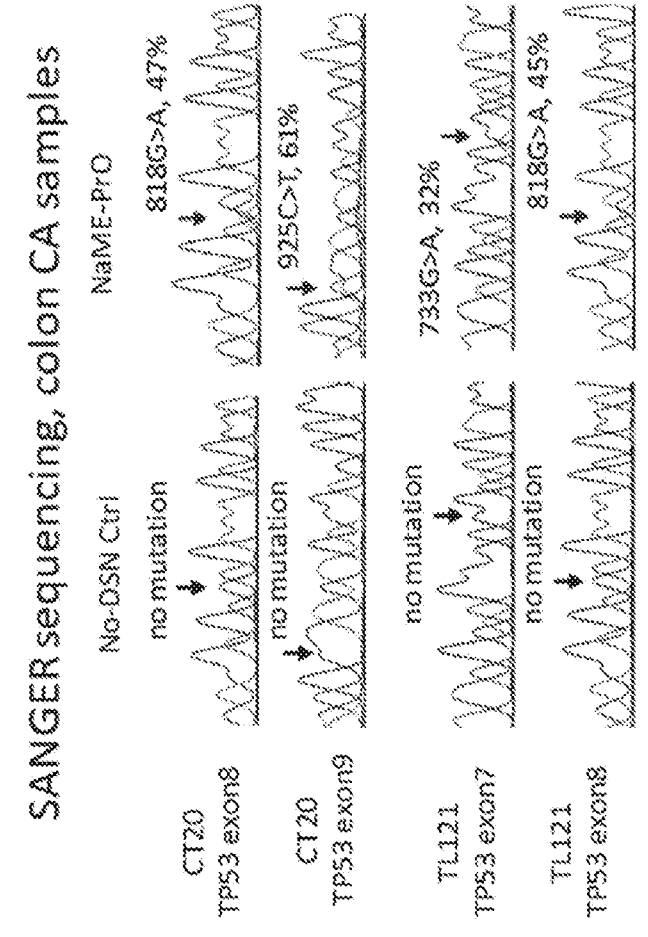

SANGER sequencing, colon CA samples
No-OSN Ctrl
NaME-PrO
CT20
TP53 exon8
no mutation
818G>A, 47%
CT20
TP53 exon9
no mutation
925C>T, 61%
TL121
TP53 exon7
no mutation
733G>A, 32%
TL121
TP53 exon8
no mutation
818G>A, 45%

FIG. 8C

FIG. 9

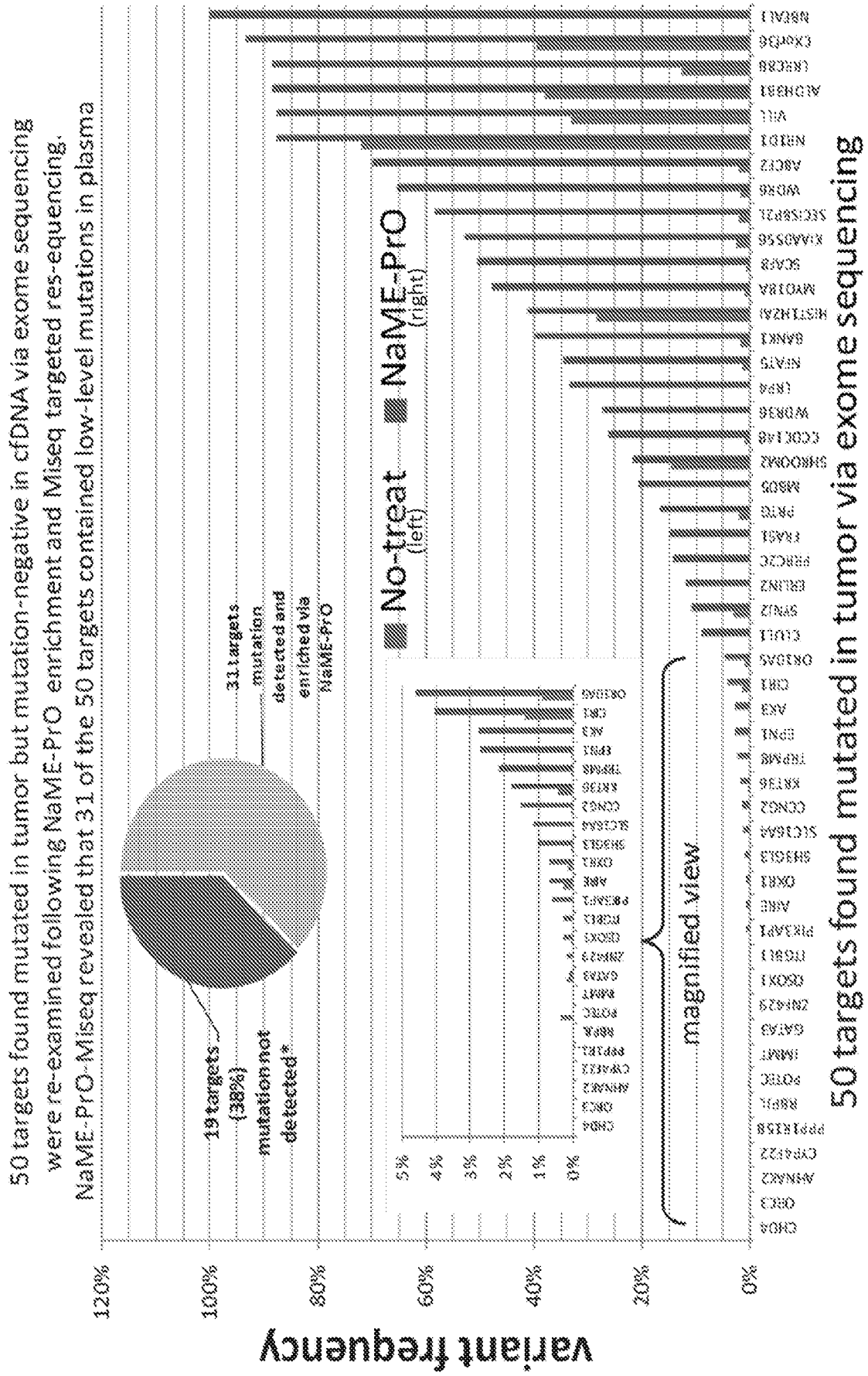

50 targets found mutated in tumor but mutation-negative in cfDNA via exome sequencing were re-examined following NaME-PrO enrichment and Miseq targeted res-equencing. NaME-PrO-Miseq revealed that 31 of the 50 targets contained low-level mutations in plasma
19 targets (38%) mutation not detected*
31 targets mutation detected and enriched via NaME-PrO
No-treat (left)
NaME-PrO (right)
magnified view
50 targets found mutated in tumor via exome sequencing
variant frequency
120%
100%
80%
60%
40%
20%
0%
5%
4%
3%
2%
1%
0%

FIG. 10A

FIG. 10B

To cover an average cfDNA fragment size of ~150bp, 3-4 parallel NaME-PrO reactions with contiguous tiling probes are needed. Therefore 3-4 groups of NaME-PrO probes are needed to enrich multiple mutations on all exons targeted.

FIG. 11

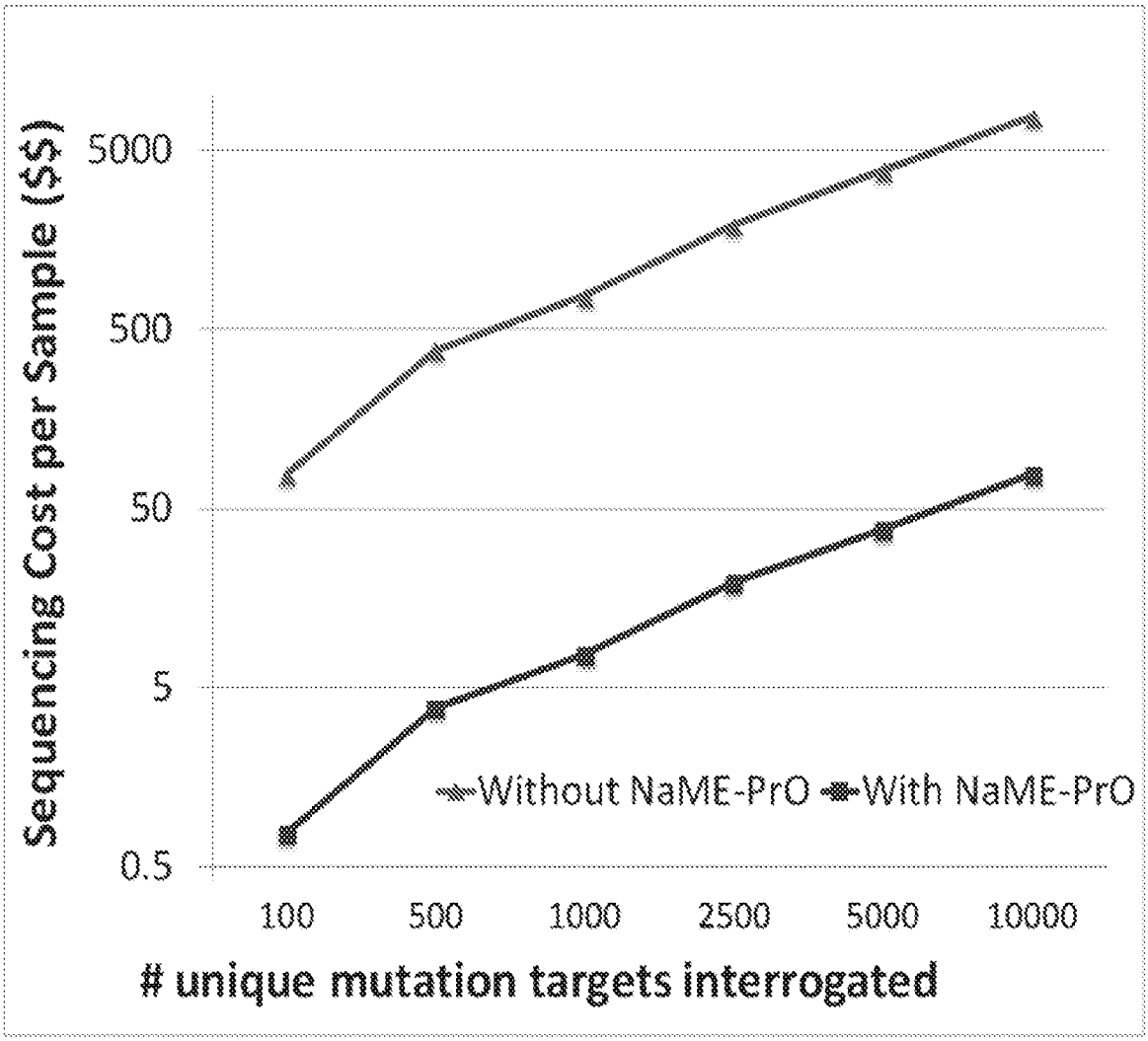
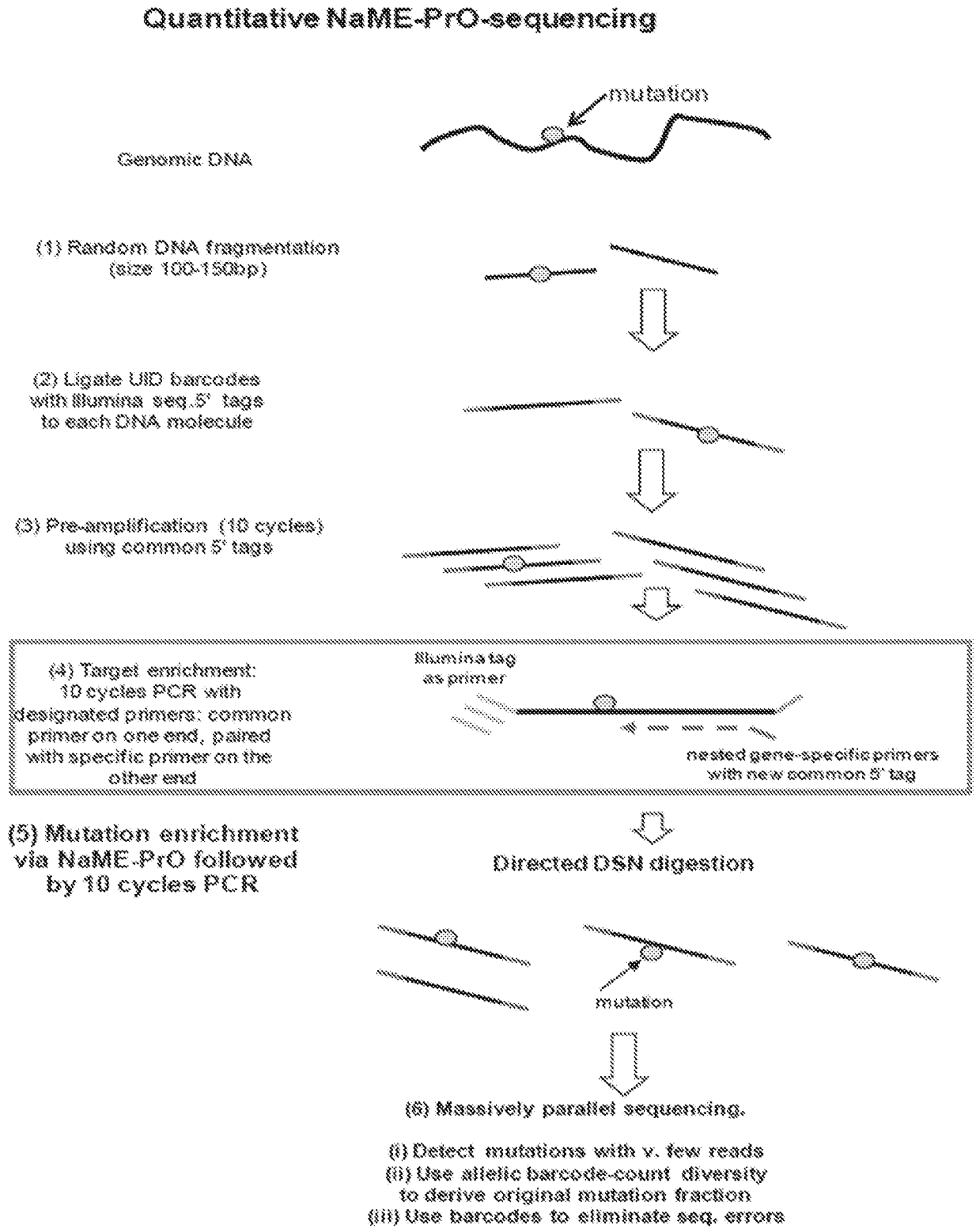

Sequencing Cost per Sample ($$)
5000
500
50
5
0.5
Without NaME-PrO
With NaME-PrO
100
500
1000
2500
5000
10000
unique mutation targets interrogated

FIG. 12


Quantitative NaME-PrO-sequencing
mutation
Genomic DNA
(1) Random DNA fragmentation (size 100-150bp)
(2) Ligate UID barcodes with Illumina seq.5' tags to each DNA molecule
(3) Pre-amplification (10 cycles) using common 5' tags
(4) Target enrichment: 10 cycles PCR with designated primers: common primer on one end, paired with specific primer on the other end
Illumina tag as primer
nested gene-specific primers with new common 5' tag
(5) Mutation enrichment via NaME-PrO followed by 10 cycles PCR
Directed DSN digestion
mutation
(6) Massively parallel sequencing.
(i) Detect mutations with v. few reads
(ii) Use allelic barcode-count diversity to derive original mutation fraction
(iii) Use barcodes to eliminate seq. errors

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US20 17/065747

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/10; C12N 15/11; C12P 19/34; C12Q 1/68 (2018.01)

CPC - C12N 15/10; C12P 19/34; C12Q 1/68; C12Q 1/6827; C12Q 1/6846; C12Q 1/6858; C12Q 1/6874; C12Q 2600/156 (2018.02)

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

USPC - 435/6; 435/91.2; 506/2; 506/26; 536/23.1; 536/24.3; 536/24.33 (keyword delimited)

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2005/0266417 A1 (BARANY et al) 01 December 2005 (01.12.2005) entire document	1-5
Y	US 2016/0046986 A1 (GUARDANT HEALTH, INC.) 18 February 2016 (18.02.2016) entire document	1-5
A	CASTELLANOS-RIZALDOS et al. "Enrichment of mutations in multiple DNA sequences using COLD-PCR in emulsion," PLoS One, 06 December 2012 (06.12.2012), Vol. 7, e51362, Pg. 1-7. entire document	1-5
A	"Paired-End Sequencing Sample Preparation Guide," Illumina Catalog #PE-930-1001 , 01 September 2009 (01.09.2009), Pgs. 1-34. Retrieved from the Internet: <http://mmjggl.caltech.edu/sequencing/Paired-End_SamplePrep_Guide_1005063_B.pdf> on 01 February 2018 (01.02.2018). entire document	1-5
A	US 2010/0203532 A1 (MAKRIGIORGOS) 12 August 2010 (12.08.2010) entire document	1-5
A	US 2016/0289670 A1 (RAINDANCE TECHNOLOGIES, INC) 06 October 2016 (06.10.2016) entire document	1-5

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
01 February 2018	06 MAR 2018
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/065747

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

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

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

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

3. ☒ Claims Nos.: 6-56
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

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

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

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.