

**METHOD AND MARKERS FOR IDENTIFICATION AND RELATIVE
QUANTIFICATION OF NUCLEIC ACID SEQUENCE, MUTATION, COPY NUMBER,
OR METHYLATION CHANGES USING COMBINATIONS OF NUCLEASE,
LIGATION, DEAMINATION, DNA REPAIR, AND POLYMERASE REACTIONS
WITH CARRYOVER PREVENTION**

[0001] This application claims benefit of U.S. Provisional Patent Application Serial No. 63/019,142, filed on May 1, 2020, which is hereby incorporated by reference in its entirety.

FIELD

[0002] The present application relates to methods and markers for identifying and quantifying nucleic acid sequence, mutation, copy number, and/or methylation changes using combinations of nuclease, ligation, deamination, DNA repair and polymerase reactions with carryover prevention.

BACKGROUND

[0003] Cancer is the leading cause of death in developed countries and the second leading cause of death in developing countries. Cancer kills 580,000 patients annually in the US, 1.3 million in Europe, and 2.8 million in China (Siegel et al., "Cancer Statistics, 2016," *CA Cancer J. Clin.* 66(1):7-30 (2016)). Cancer is now the biggest cause of mortality worldwide, with an estimated 8.2 million deaths from cancer in 2012 (Torre et al., "Global Cancer Statistics, 2012," *CA Cancer J. Clin.* 65(2):87-108 (2015)). Cancer cases worldwide are forecast to rise by 75% and reach close to 25 million over the next two decades. The lifetime risk of a woman dying from an invasive cancer is 19%, for a man it is 23%. With total annual costs of cancer care in the U.S. exceeding \$400 billion, there is no other medical issue that so urgently needs intelligent solutions.

[0004] In the U.S., new cancer cases among men are dominated by prostate (21%), lung (14%), colorectal (8%), urinary bladder (7%), melanoma (6%), non-Hodgkin lymphoma (5%), renal (5%), head and neck (4%), leukemia (4%), and liver and bile cancer (3%). Among women, most of the newly diagnosed cancers are breast (29%), lung (13%), colorectal (8%), uterine corpus (7%), thyroid (6%), non-Hodgkin lymphoma (4%), melanoma (3%), leukemia (3%), pancreatic (3%), and renal cancer (3%). The leading causes of cancer deaths are lung cancer (27%), prostate cancer (8%), colorectal cancer (8%), and lung cancer (26%), breast cancer (14%), colorectal cancer (8%), for men and women, respectively. These cancers are driven by different biological processes, and while there have been exciting advancements in the treatment

of some cancers, such as the emergence of targeted therapeutics and immunotherapy, most cancers are found at later stage, where survival is poor. Due to lack of reliable and inexpensive early detection tests, many cancer types are diagnosed at later stages, where survival rates for some cancers drop to below 10%. The current screening technologies are failing due to low patient compliance, high expense, and low sensitivity and specificity rates (Das et al., "Predictive and Prognostic Biomarkers in Colorectal Cancer: A Systematic Review of Recent Advances and Challenges," *Biomedicine & Pharmacotherapy* 87:8-19 (2016)). For example, the high cost, discomfort, and invasiveness of colonoscopy are significant impediments to patient compliance for CRC screening (Beydoun et al., "Predictors of Colorectal Cancer Screening Behaviors Among Average-risk Older Adults in the United States," *Cancer Causes & Control: CCC* 19(4):339-359 (2008)). Likewise, patient distaste for handling feces has limited the success of FOBT/FIT, and eliminated stool-based tests as a remedy for low compliance. In contrast, the current proposal addresses these problems by developing a blood test with the potential to become widely adopted. Increasing patient compliance for CRC testing will lead to earlier detection and, ultimately, increased patient survival.

[0005] Ultimately, there is an urgent need to develop non-invasive, highly sensitive, highly specific, and cost-effective tests which will detect early-stage cancers. Two relatively recent developments in cancer research serve as the guiding principles for these tasks. First, is the use of modern genomic tools (such as genome-wide sequencing, transcriptional, and methylation profiling). Public accessibility to vast databases generated from these studies has accelerated the discovery of a wider list of molecular markers (such as promoter methylation, mutation, copy number, or expression levels of mRNA, microRNA, non-coding RNA (ncRNA), and long non-coding RNA (lncRNA) associated with cancer progression. Second is the discovery that nucleic acids can be released by the cancer cells into the patient's bloodstream. Cancer cells may undergo apoptosis (triggered cell death), which releases cell free DNA (cfDNA) into the patients' blood (Salvi et al., "Cell-free DNA as a Diagnostic Marker for Cancer: Current Insights," *OncoTargets and Therapy* 9:6549-6559 (2016)). The levels of cfDNA in serum from patients with cancer vary from vanishingly small to high, but do not correlate with cancer stage (Perlin et al., "Serum DNA Levels in Patients With Malignant Disease," *American Journal of Clinical Pathology* 58(5):601-602 (1972); Leon et al., "Free DNA in the Serum of Cancer Patients and the Effect of Therapy," *Cancer Res.* 37(3):646-650 (1977)). Moreover, exosomes (lipid vesicles ranging from 30 to 100 nm), which are released into the blood by cancer cells, can contain the same RNA molecules which serve as transcriptional signatures of the tumors. Exosomes, or tumor associated vesicles, shield mRNA,

lncRNA, ncRNA, and even mutant tumor DNA from exogenous nucleases, and, as such, the markers are in a protected state. Other protected states include, but are not limited to, DNA, RNA, and proteins within circulating tumor cells (CTCs), within other non-cellular membrane containing vesicles or particles, within nucleosomes, or within Argonaute or other protein complexes. cfDNA in particular, contains the same molecular aberrations as the solid tumors, such as mutations hyper/hypo methylation, copy number changes, or chromosomal rearrangements (Ignatiadis et al., "Circulating Tumor Cells and Circulating Tumor DNA for Precision Medicine: Dream or Reality?" *Ann. Oncol.* 25(12):2304-2313 (2014)).

[0006] Tumor-specific CpG methylations have been detected in the plasma from patients with a variety of solid tumors (Pratt VM, "Are We Ready for a Blood-Based Test to Detect Colon Cancer?" *Clinical Chemistry* 60(9):1141-1142 (2014); Warton et al., "Methylation of Cell-free Circulating DNA in the Diagnosis of Cancer," *Frontiers in Molecular Biosciences* 2:13 (2015)), through various techniques involving bisulfite conversion of unmethylated cytosines, methylation-sensitive enzymes, or immunoprecipitation of 5-methylcytosines (Jorda et al., "Methods for DNA methylation analysis and applications in colon cancer," *Mutat. Res.* 693(1-2):84-93 (2010)). Methylation signatures have better specificity towards a particular cancer type likely because methylation patterns are highly tissue specific (Issa JP, "DNA Methylation as a Therapeutic Target in Cancer," *Clin. Cancer Res.* 13(6):1634-1637 (2007)). The best studied blood-based methylation markers for CRC detection are located in the promoter region of the SEPT9 gene (Church et al., "Prospective Evaluation of Methylated SEPT9 in Plasma for Detection of Asymptomatic Colorectal Cancer," *Gut* 63(2):317-325 (2014); Lofton-Day et al., "DNA Methylation Biomarkers for Blood-Based Colorectal Cancer Screening," *Clinical Chemistry* 54(2):414-423 (2008); Potter et al., "Validation of a Real-time PCR-based Qualitative Assay for the Detection of Methylated SEPT9 DNA in Human Plasma," *Clinical Chemistry* 60(9):1183-1191 (2014); Ravegnini et al., "Simultaneous Analysis of SEPT9 Promoter Methylation Status, Micronuclei Frequency, and Folate-Related Gene Polymorphisms: The Potential for a Novel Blood-Based Colorectal Cancer Biomarker," *International Journal of Molecular Sciences* 16(12):28486-28497 (2015); Toth et al., "Detection of Methylated SEPT9 in Plasma is a Reliable Screening Method for Both Left- and Right-sided Colon Cancers," *PloS One* 7(9):e46000 (2002); Toth et al., "Detection of Methylated Septin 9 in Tissue and Plasma of Colorectal Patients with Neoplasia and the Relationship to the Amount of Circulating Cell-Free DNA," *PloS One* 9(12):e115415 (2014); Warren et al., "Septin 9 Methylated DNA is a Sensitive and Specific Blood Test for Colorectal Cancer," *BMC Medicine* 9:133 (2011)), and other potential markers for CRC diagnostics include CpG sites on promoter regions of THBD

(Lange et al., “Genome-scale Discovery of DNA-methylation Biomarkers for Blood-Based Detection of Colorectal Cancer,” *PloS One* 7(11):e50266 (2012)), C9orf50 (Lange et al., “Genome-scale Discovery of DNA-methylation Biomarkers for Blood-Based Detection of Colorectal Cancer,” *PloS One* 7(11):e50266 (2012)), ZNF154 (Margolin et al., “Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development,” *The Journal of Molecular Diagnostics* 18(2):283-298 (2016)), and AGBL4, FLI1 and TWIST1 (Lin et al., “Clinical Relevance of Plasma DNA Methylation in Colorectal Cancer Patients Identified by Using a Genome-Wide High-Resolution Array,” *Ann. Surg. Oncol.* 22 Suppl 3:S1419-1427 (2015)). In breast cancer, methylation at promoter regions of tumor suppressor genes (including ATM, BRCA1, RASSF1, APC, and RAR β) has been detected in patients’ cfDNAs (Tang et al., “Blood-based DNA Methylation as Biomarker for Breast Cancer: a Systematic Review,” *Clinical Epigenetics* 8:115 (2016)). A caveat for using methylation markers is that bisulfite conversion tends to destroy DNA, and thus decreases the overall signal that can be detected. Methylation detection techniques may also lead to false-positive signals due to incomplete conversion of unmethylated cytosines. As described herein, an extensive bioinformatics analysis of public databases has been performed to identify CRC-specific, and tissues-specific methylation markers suitable for detection of cancer in the plasma. The methylation marker detection assays enable a higher level of multiplexing with single-molecule detection capabilities, which are predicted to allow for higher sensitivity and specificity across a broad spectrum of cancers.

[0007] The challenge to develop reliable diagnostic and screening tests is to distinguish those markers emanating from the tumor that are indicative of disease (*e.g.*, early cancer) vs. presence of the same markers emanating from normal tissue (which would lead to a false-positive signal). There is also a need to balance the number of markers examined and the cost of the test, with the specificity and sensitivity of the assay. Comprehensive molecular profiling (mRNA, methylation, copy number, miRNA, mutations) of thousands of tumors by The Cancer Genome Atlas Consortium (TCGA), has revealed that colorectal tumors are as different from each other as they are from breast, prostate, or other epithelial cancers (TCGA “Comprehensive Molecular Characterization of Human Colon and Rectal Cancer *Nature* 487:330-337 (2014)). Further, those few markers they share in common are also present in multiple cancer types, hindering the ability to pinpoint the tissue of origin. BRAF mutations frequently occur in melanoma (42%) and thyroid cancer (41%), while KRAS is also highly mutated in pancreatic (55%) and lung (16%) cancers (Forbes et al., “COSMIC: Exploring the World's Knowledge of Somatic Mutations in Human Cancer,” *Nucleic Acids Res.* 43(Database issue):D805-811 (2015)).

In general, CRC mutation markers such as those of *KRAS* and *BRAF* are found in late-stage primary cancers and metastases (Spindler et al., “Circulating free DNA as Biomarker and Source for Mutation Detection in Metastatic Colorectal Cancer,” *PloS One* 10(4):e0108247 (2015); Gonzalez-Cao et al., “BRAF Mutation Analysis in Circulating Free Tumor DNA of Melanoma Patients Treated with BRAF Inhibitors,” *Melanoma Res.* 25(6):486-495 (2015); Sakai et al., “Extended RAS and BRAF Mutation Analysis Using Next-Generation Sequencing,” *PloS One* 10(5):e0121891 (2015)). For early cancer detection, the nucleic acid assay should serve primarily as a screening tool, requiring the availability of secondary diagnostic follow-up (e.g., colonoscopy for colorectal cancer).

[0008] Compounding the biological problem is the need to reliably quantify mutation, CpG methylation, or DNA or RNA copy number from either a very small number of initial cells (i.e. from CTCs), or when the cancer signal is from cell-free DNA (cfDNA) in the blood and diluted by an excess of nucleic acid arising from normal cells, or inadvertently released from normal blood cells during sample processing (Mateo et al., “The Promise of Circulating Tumor Cell Analysis in Cancer Management,” *Genome Biol.* 15:448 (2014); Haque et al., “Challenges in Using ctDNA to Achieve Early Detection of Cancer,” *BioRxiv.* 237578 (2017)).

[0009] Some cancer IVD companies have developed commercially available methylation detection tests. The aforementioned SEPT9 methylation is the basis for Epi proColon test, a CRC-detection assay by Epigenomics (Lofton-Day et al., “DNA Methylation Biomarkers for Blood-based Colorectal Cancer Screening,” *Clinical Chemistry* 54(2):414-423 (2008)). While initial results on smaller sample sets showed promise, large-scale studies with 1,544 plasma samples showed a sensitivity of 64% for stage I-III CRC, and a specificity of 78%-82%, effectively sending 180 to 220 out of 1,000 individuals to unnecessary colonoscopies (Potter et al., “Validation of a Real-time PCR-based Qualitative Assay for the Detection of Methylated SEPT9 DNA in Human Plasma,” *Clinical Chemistry* 60(9):1183-1191 (2014)). Clinical Genomics is currently developing blood based CRC detection test based on the methylation of the BCAT1 and IKZF1 genes (Pedersen et al., “Evaluation of an Assay for Methylated BCAT1 and IKZF1 in Plasma for Detection of Colorectal Neoplasia,” *BMC Cancer* 15:654 (2015)). Large-scale studies using 2,105 plasma samples of this two-marker test showed an overall sensitivity of 66%, with 38% for stage I CRC, and an impressive specificity of 94% (Young et al., “A Cross-sectional Study Comparing a Blood Test for Methylated BCAT1 and IKZF1 Tumor-derived DNA with CEA for Detection of Recurrent Colorectal Cancer,” *Cancer Medicine* 5(10): 2763-2772 (2016)). Exact Sciences and collaborators have slightly improved the sensitivity of CRC fecal tests (Bosch et al., “Analytical Sensitivity and Stability of DNA

Methylation Testing in Stool Samples for Colorectal Cancer Detection,” *Cell Oncol. (Dordr)* 35(4):309-315 (2012); Hong et al., “DNA Methylation Biomarkers of Stool and Blood for Early Detection of Colon Cancer,” *Genet. Test. Mol. Biomarkers* 17(5):401-406 (2013); Imperiale et al., “Multitarget Stool DNA Testing for Colorectal-Cancer Screening,” *N. Engl. J. Med.* 370(14):1287-1297 (2014); Xiao et al., “Validation of Methylation-Sensitive High-Resolution Melting (MS-HRM) for the Detection of Stool DNA Methylation in Colorectal Neoplasms,” *Clin. Chim. Acta* 431:154-163 (2014); Yang et al., “Diagnostic Value of Stool DNA Testing for Multiple Markers of Colorectal Cancer and Advanced Adenoma: a Meta-Analysis,” *Can. J. Gastroenterol.* 27(8):467-475 (2013)), by adding K-ras mutation as well as BMP3 and NDRG4 methylation markers (Lidgard et al., “Clinical Performance of an Automated Stool DNA Assay for Detection of Colorectal Neoplasia,” *Clin. Gastroenterol. Hepatol.* 11(10):1313-1318 (2013)). Large-scale studies on 12,500 stool samples claims 93% sensitivity, yet specificity is still only 85%, essentially sending 150 out of 1,000 individuals to unnecessary colonoscopies. Despite logistical issues in handling feces, Exact Sciences recently sold their millionth test. The Cologuard website states the test result has both false-positives and false-negatives, and the test should not be used if the patient has hemorrhoids, menstrual period, or blood in the stool. The Cologuard website also warns that the test is not for use by patients with Ulcerative Colitis (UC), Crohn’s disease (CD), Inflammatory Bowel Disease (IBD), or with a family history of cancer. In other words, Exact Sciences excludes the very patients who would most benefit from an accurate CRC detection test. More recently, Laboratory for Advanced Medicine (based in Irvine, CA with ties to various Chinese academic institutions) demonstrated the potential of interrogating the methylation status of a single CpG site (cg10673833) for blood-based detection of colorectal cancer (Luo et al., “Circulating Tumor DNA Methylation Profiles Enable Early Diagnosis, Prognosis Prediction, and Screening for Colorectal Cancer,” *Science Translational Medicine* 12:(524) (2020)).

A continuum of diagnostic needs will require a continuum of diagnostic tests.

[0010] The majority of current molecular diagnostics efforts in cancer have centered on: (i) prognostic and predictive genomics, *e.g.*, identifying inherited mutations in cancer predisposition genes, such as BrCA1, BrCA2, (Ford et al., *Am. J. Hum. Genet.* 62:676-689 (1998)) (ii) individualized treatment, *e.g.*, mutations in the EGFR gene guiding personalized medicine (Sequist and Lynch, *Ann. Rev. Med.* 59:429-442 (2008)), and (iii) recurrence monitoring, *e.g.*, detecting emerging KRAS mutations in patients developing resistance to drug treatments (Hiley et al., *Genome Biol.* 15: 453 (2014); Amado et al., *J. Clin. Oncol.* 26:1626-

1634 (2008)). Yet, this misses major opportunities in the cancer molecular diagnostics continuum: (i) more frequent screening of those with a family history, (ii) screening for detection of early disease, and (iii) monitoring treatment efficacy. To address these three unmet needs, a new metric for blood-based detection termed “cancer marker load”, analogous to viral load is herein proposed.

[0011] DNA sequencing provides the ultimate ability to distinguish all nucleic acid changes associated with disease. However, the process still requires multiple up-front sample and template preparation, and consequently, DNA sequencing is not always cost-effective. DNA microarrays can provide substantial information about multiple sequence variants, such as SNPs or different RNA expression levels, and are less costly than sequencing; however, they are less suited for obtaining highly quantitative results, nor for detecting low abundance mutations. On the other end of the spectrum is the TaqMan™ reaction, which provides real-time quantification of a known gene, but is less suitable for distinguishing multiple sequence variants or low abundance mutations.

[0012] NGS requires substantial up-front sample preparation to polish ends and append linkers, and the current error rates of 0.7% are too high to identify 2-3 molecules of mutant sequence in a 10,000-fold excess of wild-type molecules. “Deep sequencing” protocols have been developed to overcome this deficiency by appending unique molecular identifiers to both strands of an individual fragment. These approaches are known as: Tam-Seq & CAPP-Seq (Roche), Circle-Seq (Guardant Health), Safe-SeqS (Personal Genome Diagnostics), ThruPlex (Rubicon Genomics), NEBNext (New England Biolabs), QIAseq (Qiagen), Oncomine (ThermoFisher), Duplex Barcoding (Schmitt), SMRT (Pacific Biosciences), SiMSen-Seq (Stahlberg), and smMIP (Shendure). However, these methods require a 30 to 100-fold depth per mutant strand to verify each mutation and distinguish from other types of sequencing errors. Recent work from MSKCC demonstrates that 60,000-fold coverage is required to accurately identify mutations in plasma from metastatic cancer patients (91% sensitivity, 508-gene panel, 60,000x). Compounding the challenge, a recent paper from NEB has called into question the quality of the most widely used databases for rare variant and somatic mutations (Chen et al., “DNA Damage is a Pervasive Cause of Sequencing Errors, Directly Confounding Variant Identification,” *Science* 355(6326):752-756 (2017)).

[0013] It is critical to match each unmet diagnostic need with the appropriate diagnostic test – one that combines the divergent goals of achieving both high sensitivity (*i.e.*, low false-negatives) and high specificity (*i.e.*, low false-positives) at a low cost. For example, direct sequencing of EGFR exons from a tumor biopsy to determine treatment for non-small cell lung

cancer (NSCLC) is significantly more accurate and cost effective than designing TaqMan™ probes for the over 180 known mutations whose drug response is already catalogued (Jia et al., *Genome Res.* 23:1434-1445 (2013)). The most sensitive technique for detecting point mutations, such as “BEAMing” (Dressman et al., *Proc. Natl. Acad. Sci. USA* 100: 8817-8822 (2003)), rely on prior knowledge of which mutations to look for, and thus are best suited for monitoring for disease recurrence, rather than for early detection. Likewise, to monitor blood levels of Bcr-Abl translocations when treating CML patients with Gleevec (Jabbour et al., *Cancer* 112: 2112-2118 (2008)), a simple quantitative reverse-transcription PCR assay is far preferable to sequencing the entire genomic DNA in 1 ml of blood (9 million cells x 3 GB = 27 million Gb of raw data).

[0014] Sequencing 2.1 Gb each of cell-free DNA (cfDNA) isolated from NSCLC patients was used to provide 10,000-fold coverage on 125 kb of targeted DNA (Kandoth et al., *Nature* 502: 333-339 (2013)). This approach correctly identified mutations present in matched tumors, although only 50% of stage 1 tumors were covered. The approach has promise for NSCLC, where samples average 5 to 20 mutations/Mb, however targeted NGS would not be cost effective for other cancers such as breast and ovarian, that average less than 1 to 2 mutations per Mb. Current up-front ligation, amplification, and/or capture steps required for highly accurate targeted deep sequencing are still more complex than multiplexed PCR-TaqMan™ or PCR-LDR assays.

[0015] Deep sequencing of cfDNAs for 58 cancer-related genes at 30,000-fold coverage is capable of detecting Stage 1 or 2 cancer at moderately high sensitivity but missed 29% of CRC, 41% of breast, 41% of lung, and 32% of ovarian cancer, respectively (Phallen et al., “Direct Detection of Early-stage Cancers Using Circulating Tumor DNA,” *Science Translational Medicine* 9(403) (2017)). An alternative strategy relied on targeted sequencing of an average of 30 bases in 61 segments to interrogate “hot-spot” mutations in 16 genes including TP53, KRAS, APC, PIK3CA, PTEN, missed more early cancers (Cohen et al., “Detection and Localization of Surgically Resectable Cancers with a Multi-analyte Blood Test,” *Science* (2018)). To extend the sensitivity of mutation sequencing, the Hopkins team very recently combined NGS with quantitation of serum protein markers (such as CA-125, CA19-9, CEA, HGF, Myeloperoxidase, OPN, Prolactin, TIMP-1) and improved detection of five cancer types (ovary, liver, stomach, pancreas, and esophagus) at sensitivities ranging from 69% to 98% (Cohen et al. “Detection and Localization of Surgically Resectable Cancers with a Multi-analyte Blood Test,” *Science* (2018)). One caveat of using these protein markers is that prior large-scale studies with age-matched controls (n= 22,000) have not shown clinical utility (Jacobs et al., “Prevalence Screening for Ovarian Cancer in Postmenopausal Women by CA 125 Measurement and Ultrasonography,”

BMJ 306(6884):1030-1034 (1993)). Thus, in a 2018 JAMA report, “The USPSTF recommends against [CA-125] screening for ovarian cancer in asymptomatic women. This recommendation applies to asymptomatic women who are not known to have a high-risk hereditary cancer syndrome” (USPSTF et al., “Screening for Ovarian Cancer: US Preventive Services Task Force Recommendation Statement” *JAMA* 319(6):588-594 (2018)). Another caveat of using these protein markers is that they reflect tissue damage and are likely to also appear in patients with inflammatory diseases such as arthritis (Kaiser, “Liquid Biopsy for Cancer Promises Early Detection,” *Science* 359(6373):259 (2018)). With the growing obesity epidemic and an aging population in the U.S., the risk of false-positives from protein markers increases with obesity and age-driven inflammation.

[0016] More recently, the NGS sequencing companies (Grail, Guardant Health, Natera, Freenome) have moved aggressively to expand their targeted sequencing panels to also now include whole genome sequencing (WGS) and whole genome bisulfite sequencing (Bis-WGS). The recent results from Grail, abstract published at ASCO 2018 (Klein et al., “Development of a Comprehensive Cell-free DNA (cfDNA) Assay for Early Detection of Multiple Tumor Types: The Circulating Cell-free Genome Atlas (CCGA) Study,” ASCO Annual Meeting 2018, Chicago, IL; Abstract 12021#134)) reveal that while sensitivity claims of detecting “early” CRC are at 63%, that is based on only 27 samples, most of which are Stage III. Even mutation rich lung cancer gives sensitivity at 50%, again with most samples at Stage III. When most of the samples are Stage I & II, such as prostate cancer, the sensitivity for “early cancer” detection drops to < 5%. When attempting to detect the most common form of breast cancer (HR+/HER2), the sensitivity drops to < 13%. Worse, those breast cancers diagnosed by screening gave sensitivities of < 11%. In short, the NGS approach fails by consistently missing 30% to 80% of early-stage cancers (i.e. stage I & II). In a research initially reported in 2019 ASCO meeting (Liu et al., “Simultaneous Multi-cancer Detection and Tissue of Origin (TOO) Localization Using Targeted Bisulfite Sequencing Plasma Cell-free DNA (cfDNA),” ASCO Breakthrough Presentation 2019)), and subsequently published in 2020 (Liu et al., “Sensitive and Specific Multi-cancer Detection and Localization Using Methylation Signatures in Cell-free DNA,” *Annals of Oncology*; In Press (2020)), GRAIL indicated that their Multi-Cancer Early Detection Test exhibited an Overall Detection Rate (12 deadly cancer types) of 76 % (99.3 % specificity). A combined analysis of this group of cancers showed robust detection across all stages with detection rates of 39 percent (27-52%), 69 percent (56-80%), 83 percent (75-90%), and 92 percent (86-96%) at stages I (n=62), II (n=62), III (n=102), and IV (n=130), respectively. In another conference, GRAIL and collaborators (Oxnard et al., “Simultaneous Multi-cancer

Detection and Tissue of Origin (TOO) Localization Using Targeted Bisulfite Sequencing of Plasma Cell-free DNA (cfDNA)," ESMO Congress (2019)), reported the results from their analysis of cell-free DNA (DNA that had once been confined to cells but had entered the bloodstream upon the cells' death) in 3,583 blood samples, including 1,530 from patients diagnosed with cancer and 2,053 from people without cancer. The patient samples comprised more than 20 types of cancer, including hormone receptor-negative breast, colorectal, esophageal, gallbladder, gastric, head and neck, lung, lymphoid leukemia, multiple myeloma, ovarian, and pancreatic cancer. The overall specificity was 99.4%, meaning only 0.6% of the results incorrectly indicated that cancer was present. The sensitivity of the assay for detecting a pre-specified high mortality cancer (the percent of blood samples from these patients that tested positive for cancer) was 76%. Within this group, the sensitivity was 32% for patients with stage I cancer; 76% for those with stage II; 85% for stage III; and 93% for stage IV. Sensitivity across all cancer types was 55%, with similar increases in detection by stage. For the 97% of samples that returned a tissue of origin result, the test correctly identified the organ or tissue of origin in 89% of cases. However, another 2019 study (reported by GRAIL and collaborators) questioned the validity of the aforementioned reports (Razavi et al., "High-intensity Sequencing Reveals the Sources of Plasma Circulating Cell-free DNA Variants," *Nat Med* 25(12):1928-1937 (2019)). Through a 2 Mb, 508-gene panel sequencing (60,000x depth), the authors demonstrated the vast majority of cell-free DNA mutations in both non-cancer controls and cancer patients had features consistent with clonal hematopoiesis, a process whereby white blood cells progressively accumulate somatic alterations without necessarily producing a hematological condition or malignancy. Indeed, mutations appeared in 93.6 percent of the white blood cells from individuals without cancer and 99.1 percent of those with cancer. In a recently held conference, GRAIL and their collaborators reported that their blood-based test can detect multiple GI cancers at sensitivity of under 50% for Stage I, and 73% for Stage I-III (Wolpin et al., "Performance of a Blood-based Test for the Detection of Multiple Cancer Types," In: *Gastrointestinal Cancers Symposium 2020* (2020)). As for Freenome, a recent ASCO presentation indicated that their platform (plasma analysis by whole-genome sequencing, bisulfite sequencing, and protein quantification methods) was able to achieve a mean sensitivity of 92% in early-stage (n = 17) and 84% in late-stage (n = 11) at a specificity of 90% for colorectal adenocarcinoma detection. Across all CRC pathological subtypes, the Freenome test achieved a specificity of 90%, and sensitivities of 80% and 83% for early-stage (n = 19) and late-stage (n = 12), respectively. Private discussion with Imran Haque, who just resigned as CSO of Freenome – where he had a \$70 million budget and 30 scientists to sequence the plasma of 817 CRC and matched control

patients – confirmed that Freenome (as well as GRAIL) were overcalling the data, and that none of them had a cogent approach to achieve cost-effective true early cancer detection (Wan et al., “Machine Learning Enables Detection of Early-stage Colorectal Cancer by Whole-genome Sequencing of Plasma Cell-free DNA,” *BioRxiv* 478065 (2018)).

[0017] A comprehensive data analysis of over 600 colorectal cancer samples that takes into account tumor heterogeneity, tumor clusters, and biological/technical false-positives ranging from 3% to 10% per individual marker showed that the optimal early detection screen for colorectal cancer would require at least 5 to 6 positive markers out of 24 markers tested (Bacolod et al., *Cancer Res.* 69:723-727 (2009); Tsafrir et al., *Cancer Res.* 66: 2129-2137 (2006); Weinstein et al., *Nat. Genet.* 45: 1113-1120 (2013); Navin N.E. *Genome Biol.* 15: 452 (2014); Hiley et al., *Genome Biol* 15:453 (2014); Esserman et al. *Lancet Oncol* 15:e234-242 (2014)). Further, marker distribution is biased into different tumor clades, *e.g.*, some tumors are heavily methylated, while others are barely methylated, and indistinguishable from age-related methylation of adjacent tissue. Consequently, a multidimensional approach using combinations of 3-5 sets of mutation, methylation, miRNA, ncRNA, lncRNA, mRNA, copy-variation, alternative splicing, or translocation markers is needed to obtain sufficient coverage of all different tumor clades. Analogous to non-invasive prenatal screening for trisomy, based on sequencing or performing ligation detection on random fragments of cfDNA (Benn et al., *Ultrasound Obstet. Gynecol.* 42(1):15-33 (2013); Chiu et al., *Proc. Natl. Acad. Sci. U S A* 105: 20458 – 20463 (2008); Juneau et al., *Fetal Diagn. Ther.* 36(4) (2014)), the actual markers scored in a cancer screen are secondary to accurate quantification of those positive markers in the plasma.

[0018] As pointed out above, cancer-specific RNA markers (including microRNAs, lncRNAs, and mRNAs) may also be present in blood, either free of any compartment (Souza et al., “Circulating mRNAs and miRNAs as Candidate Markers for the Diagnosis and Prognosis of Prostate Cancer,” *PloS One* 12(9):e0184094 (2017)), or contained in exosomes (Nedaeinia et al., “Circulating Exosomes and Exosomal microRNAs as Biomarkers in Gastrointestinal Cancer,” *Cancer Gene Ther* 24(2):48-56 (2017); Lai et al., “A microRNA Signature in Circulating Exosomes is Superior to Exosomal Glypican-1 Levels for Diagnosing Pancreatic Cancer,” *Cancer Lett* 39:86-93 (2017)) or circulating tumor cells (“CTCs”), and have been tagged as potential indicators of early- stage cancers. Challenges abound regarding the use of plasma-derived nucleic markers in early cancer detection, including the minuscule amount of these markers in blood relative to those derived from surrounding cells. Indeed, these limitations make it appear that these “early” detection assays are more likely to detect late-stage primary and

metastatic cancers (Pantel “Blood-Based Analysis of Circulating Cell-Free DNA and Tumor Cells for Early Cancer Detection,” *PLoS Med* 13(12):e1002205 (2016)).

Technical Challenges of Cancer Diagnostic Test Development.

[0019] Diagnostic tests that aim to find very rare or low-abundance mutant sequences face potential false-positive signal arising from: (i) polymerase error in replicating wild-type target, (ii) DNA sequencing error, (iii) mis-ligation on wild-type target, (iii) target independent PCR product, and (iv) carryover contamination of PCR products arising from a previous positive sample. The profound clinical implications of a positive test result when screening for cancer demand that such a test use all means possible to virtually eliminate false-positives.

[0020] Central to the concept of nucleic acid detection is the selective amplification or purification of the desired cancer-specific markers away from the same or closely similar markers from normal cells. These approaches include: (i) multiple primer binding regions for orthogonal amplification and detection, (ii) affinity selection of CTC's or exosomes, and (iii) spatial dilution of the sample.

[0021] The success of PCR-LDR, which uses 4 primer-binding regions to assure sensitivity and specificity, has previously been demonstrated. Desired regions are amplified using pairs or even tandem pairs of PCR primers, followed by orthogonal nested LDR primer pairs for detection. One advantage of using PCR-LDR is the ability to perform proportional PCR amplification of multiple fragments to enrich for low copy targets, and then use quantitative LDR to directly identify cancer-specific mutations. Biofire/bioMerieux has developed a similar technology termed “film array”; wherein initial multiplexed PCR reaction products are redistributed into individual wells, and then nested real-time PCR performed with SYBR Green Dye detection.

[0022] Affinity purification of CTC's using antibody or aptamer capture has been demonstrated (Adams et al., *J. Am. Chem. Soc.* 130: 8633-8641 (2008); Dharmasiri et al., *Electrophoresis* 30:3289-3300 (2009); Soper et al. *Biosens. Bioelectron.* 21:1932-1942 (2006)). Peptide affinity capture of exosomes has been reported in the literature. Enrichment of these tumor-specific fractions from the blood enables copy number quantification, as well as simplifying screening and verification assays.

[0023] The last approach, spatial dilution of the sample, is employed in digital PCR as well as its close cousin known as BEAMing (Vogelstein and Kinzler, *Proc. Natl. Acad. Sci. US A.* 96(16):9236-41 (1999); Dressman et al., *Proc. Natl. Acad. Sci. USA* 100:8817-8822 (2003)). The rationale for digital PCR is to overcome the limit of enzymatic discrimination when the

sample comprises very few target molecules containing a known mutation in a 1,000 to 10,000-fold excess of wild-type DNA. By diluting input DNA into 20,000 or more droplets or beads to distribute less than one molecule of target per droplet, the DNA may be amplified via PCR, and then detected via probe hybridization or TaqMan™ reaction, giving in essence a 0/1 digital score. The approach is currently the most sensitive for finding point mutations in plasma, but it does require prior knowledge of the mutations being scored, as well as a separate digital dilution for each mutation, which would deplete the entire sample to score just a few mutations (Alcaide et al., “A Novel Multiplex Droplet Digital PCR Assay to Identify and Quantify KRAS Mutations in Clinical Specimens,” *J. Mol. Diagn.* 21:28-33 (2019); Guibert et al., “Liquid Biopsy of Fine-Needle Aspiration Supernatant for Lung Cancer Genotyping,” *Lung Cancer* 1768:193-207 (2018); Yoshida et al., “Highly Sensitive Detection of ALK Resistance Mutations in Plasma Using Droplet Digital PCR,” *BMC Cancer* 18:1136 (2018)).

[0024] When developing multiplexed assays, there is a tricky balance between performing enough preliminary cycles of PCR or other amplification techniques to generate sufficient copies of each mutant or methylated region such that when diluting into uniplex qPCR, multiplex qPCR, uniplex droplet PCR or multiplexed droplet PCR there are sufficient copies to get a signal if true positive; and performing too many PCR cycles such that some markers over-amplify while others are suppressed, or relative quantification is lost.

[0025] The present application is directed at overcoming these and other deficiencies in the art.

SUMMARY

[0026] A first aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil residues to produce a treated sample. One or more enzymes capable of digesting deoxyuracil

(dU)-containing nucleic acid molecules are provided, and one or more primary oligonucleotide primer sets are provided. Each primary oligonucleotide primer set comprises (a) a first primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer, wherein the first or second primary oligonucleotide primer further comprises a 5' primer-specific portion. The treated sample, the one or more first primary oligonucleotide primers of the primer sets, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof. One or more oligonucleotide probe sets are then provided. Each probe set comprises (a) a first oligonucleotide probe having a 5' primer-specific portion and a 3' DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and (b) a second oligonucleotide probe having a 5' DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion and a 3' primer-specific portion, and wherein the first and second oligonucleotide probes of a probe set are configured to hybridize, in a base specific manner, on a complementary nucleotide sequence of a first polymerase chain reaction product. The first

polymerase chain reaction products are blended with a ligase and the one or more oligonucleotide probe sets to form one or more ligation reaction mixtures. The one or more ligation reaction mixtures are subjected to one or more ligation reaction cycles whereby the first and second oligonucleotide probes of the one or more oligonucleotide probe sets are ligated together, when hybridized to their complementary sequences, to form ligated product sequences in the ligation reaction mixture wherein each ligated product sequence comprises the 5' primer-specific portion, the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and the 3' primer-specific portion. The method further involves providing one or more secondary oligonucleotide primer sets. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the ligated product sequence and (b) a second secondary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the ligated product sequence. The ligated product sequences, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further comprises detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0027] Another aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to

convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample. The method further involves providing one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, and providing one or more first primary oligonucleotide primer(s) that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more methylated or hydroxymethylated residue. The treated sample, the one or more first primary oligonucleotide primers, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. One or more secondary oligonucleotide primer sets are provided. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that is complementary to a portion of the polymerase extension product formed from the first primary oligonucleotide primer and (b) a second secondary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first secondary oligonucleotide primer. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures, and conditions suitable for carrying out two or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising a 5' primer-specific portion of the first secondary oligonucleotide primer, a DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and a complement of the 5' primer-specific portion of the second secondary oligonucleotide primer. The method

further comprises providing one or more tertiary oligonucleotide primer sets. Each tertiary oligonucleotide primer set comprises (a) a first tertiary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first polymerase chain reaction products and (b) a second tertiary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the first polymerase chain reactions product sequence. The first polymerase chain reaction products, the one or more tertiary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU) containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further involves detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0028] Another aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample. One or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules present in the sample, and one or more primary oligonucleotide primer sets are provided. Each primary oligonucleotide primer set comprises (a) a first primary oligonucleotide primer that comprises a nucleotide sequence that is

complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer, wherein the first or second primary oligonucleotide primer further comprises a 5' primer-specific portion. The treated sample, the one or more first primary oligonucleotide primers of the primer sets, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules in the reaction mixture, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof. One or more secondary oligonucleotide primer sets are then provided. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer having a 3' portion that is complementary to a portion of a first polymerase chain reaction product formed from the first primary oligonucleotide primer and (b) a second secondary oligonucleotide primer having a 3' portion that comprises a nucleotide sequence that is complementary to a portion of a first polymerase chain reaction product formed from the first secondary oligonucleotide primer. The first polymerase chain reaction products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase

chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out two or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further comprises detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0029] Another aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample. One or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules present in the sample are provided, and one or more primary oligonucleotide primer sets are provided. Each primary oligonucleotide primer set comprises (a) a first primary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer. The treated sample, the one or more first primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are

subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules in the reaction mixture, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reactions products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof. One or more secondary oligonucleotide primer sets are then provided. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first polymerase chain reaction products or their complements and (b) a second secondary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the first polymerase chain reaction products or their complements. The first polymerase chain reaction products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further involves detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0030] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 6-12 markers, 12-24 markers, 24-36 markers, 36-48 markers, 48-72 markers, 72-96 markers, or > 96 markers. Each marker in a given set is selected by having any one or more of the following criteria: present, or above a cutoff level, in > 50% of biological samples of the disease cells or tissue from individuals diagnosed with the disease state; absent, or below a cutoff level, in > 95% of biological samples of the normal cells or tissue from individuals without the disease state; present, or above a cutoff level, in > 50% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with the disease state; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without the disease state; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with the disease state; and, wherein at least 50% of the markers in a set each comprise one or more methylated or hydroxymethylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated or hydroxymethylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with the disease state. The method involves obtaining the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. Nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are

carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with the disease state if a minimum of 2 or 3 markers are present or above a cutoff level in a marker set comprising from 6-12 markers; or a minimum of 3, 4, or 5 markers are present or above a cutoff level in a marker set comprising from 12-24 markers; or a minimum of 3, 4, 5, or 6 markers are present or above a cutoff level in a marker set comprising from 24-36 markers; or a minimum of 4, 5, 6, 7, or 8 markers are present or above a cutoff level in a marker set comprising from 36-48 markers; or a minimum of 6, 7, 8, 9, 10, 11, or 12 markers are present or above a cutoff level in a marker set comprising from 48-72 markers, or a minimum of 7, 8, 9, 10, 11, 12 or 13 markers are present or above a cutoff level in a marker set comprising from 72-96 markers, or a minimum of 8, 9, 10, 11, 12, 13 or "n"/12 markers are present or above a cutoff level in a marker set comprising 96 to "n" markers, when "n" > 168 markers.

[0031] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 48-72 total cancer markers, 72-96 total cancer markers or ≥ 96 total cancer markers, wherein on average greater than one quarter such markers in a given set cover each of the aforementioned major cancers being tested. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in > 50% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer present, or above a cutoff level, in > 50% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily

fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining a biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 48-72 total cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 72-96 total cancer markers; or a minimum of 6 or "n"/18 markers are present or are above a cutoff level in a marker set comprising 96 to "n" total cancer markers, when "n" > 96 total cancer markers.

[0032] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of and identifying the most likely specific tissue(s) of origin of a solid tissue cancer in the following groups: Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 36-48 group-specific cancer markers, 48-64 group-specific cancer markers or ≥ 64 group-specific cancer markers, wherein on average greater than one third such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in $> 50\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in $> 50\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the

cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dioxuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 36-48 group-specific cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 48-64 group-specific cancer markers; or a minimum of 6 or "n"/12 markers are present or are above a cutoff level in a marker set comprising 64 to "n" group-specific cancer markers, when "n" > 64 group-specific cancer markers.

[0033] Another aspect of the present application relates to a method of diagnosing or prognosing a disease state of a gastrointestinal cancer including colorectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 6-12 markers, 12-18 markers, 18-24 markers, 24-36 markers, 36-48 markers or ≥ 48 markers. Each marker is selected by having any one or more of the following criteria for gastrointestinal cancer: present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with gastrointestinal cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without gastrointestinal cancer; present, or above a cutoff level, in > 75% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with gastrointestinal cancer; absent, or below a

cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without gastrointestinal cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with gastrointestinal cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with gastrointestinal cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with gastrointestinal cancer if a minimum of 2, 3 or 4 markers are present or are above a cutoff level in a marker set comprising from 6-12 markers; or a minimum of 2, 3, 4, or 5 markers are present or are above a cutoff level in a marker set comprising from 12-18 markers; or a minimum of 3, 4, 5, or 6 markers are present or are above a cutoff level in a marker set comprising from 18-24 markers; or a minimum of 3, 4, 5, 6, 7, or 8 markers are present or are above a cutoff level in a marker set comprising from 24-36 markers; or a minimum of 4, 5, 6, 7, 8, 9, or 10 markers are present or are above a cutoff level in a marker

set comprising from 36-48 markers; or a minimum of 5, 6, 7, 8, 9, 10, 11, 12, or “n”/12 markers are present or are above a cutoff level in a marker set comprising 48 to “n” markers, when “n” > 48 markers.

[0034] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 36-48 total cancer markers, 48-64 total cancer markers, or ≥ 64 total cancer markers. On average greater than half of such markers in a given set cover each of the aforementioned major cancers being tested. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample

including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dioxouracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 36-48 total cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 48-64 total cancer markers; or a minimum of 6 or "n"/12 markers are present or are above a cutoff level in a marker set comprising 64 to "n" total cancer markers, when "n" > 96 total cancer markers.

[0035] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of and identifying the most likely specific tissue(s) of origin of a solid tissue cancer in the following groups: Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half

of such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and

distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 24-36 group-specific cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 36-48 group-specific cancer markers; or a minimum of 6 or " n "/8 markers are present or are above a cutoff level in a marker set comprising 48 to " n " group-specific cancer markers, when " n " > 48 group-specific cancer markers.

[0036] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state to guide and monitor treatment of a solid tissue cancer in one or more of the following groups; Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in > 75% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions

thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise of one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals with a given tissue-specific cancer will on average have from approximately one-quarter to about one-half or more of the markers scored as present, or are above a cutoff level in the tested marker set, wherein to guide and monitor subsequent treatment, a portion or all of the identified markers scored as present or the identified markers as above a cutoff level in the tested marker set are deemed the “patient-specific marker set”, and retested on a subsequent biological sample from the individual during the treatment protocol, to monitor for loss of marker signal, wherein if a minimum of 3 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 12-24 markers; or if a minimum of 4 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 24-36 markers; or a minimum of 5 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 36-48 markers; or a minimum of 6 or “n”/8 markers remain present or remain above a cutoff level in a patient-specific marker set

comprising 48 to “n” markers, when “n” > 48 markers after the treatment protocol has been administered, then the continuing presence of said markers may guide a decision to change the cancer treatment therapy.

[0037] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state for recurrence of a solid tissue cancer in one or more of the following groups; Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in > 75% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise of one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions

thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals with a given tissue-specific cancer will on average have from approximately one-quarter to about one-half or more of the markers scored as present, or are above a cutoff level in the tested marker set, wherein to monitor for recurrence, a portion or all of the markers scored as being present, or the markers scored as above a cutoff level in the tested marker set are deemed the "patient-specific marker set", and retested on subsequent biological samples from the individual after a successful treatment, to monitor for gain of marker signal, wherein if a minimum of 3 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 12-24 markers; or if a minimum of 4 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 24-36 markers; or a minimum of 5 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 36-48 markers; or a minimum of 6 or "n"/8 markers reappear or rise above a cutoff level in a patient-specific marker set comprising 48 to "n" markers, when "n" > 48 markers after the treatment protocol has been administered, then the reappearance or rise or rise above a cutoff level in a patient-specific marker set may guide a decision to resume the cancer treatment therapy or change to a new cancer treatment therapy.

[0038] Another aspect of the present application relates to a two-step method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or

level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The method involves obtaining a biological sample that includes exosomes, tumor-associated vesicles, markers within other protected states, cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. A first step is applied to the biological samples with an overall sensitivity of $> 80\%$ and an overall specificity of $> 90\%$ or an overall Z-score of > 1.28 to identify individuals more likely to be diagnosed or prognosed with the disease state. A second step is then applied to biological samples from those individuals identified in the first step with an overall specificity of $> 95\%$ or an overall Z-score of > 1.65 to diagnose or prognose individuals with the disease state. The first step and/or the second step are carried out using a method of the present application.

[0039] The present application describes a number of approaches for detecting mutations, expression, splice variant, translocation, copy number, and/or methylation changes in target nucleic acid molecules using nuclease, ligase, and polymerase reactions. The present application solves the problems of carry over prevention, as well as allowing for spatial multiplexing to provide relative quantification, similar to digital PCR. Such technology may be utilized for non-invasive early detection of cancer, non-invasive prognosis of cancer, and monitoring for cancer recurrence from plasma or serum samples.

[0040] The present application provides a comprehensive roadmap of nucleic acid methylation, miRNA, lncRNA, ncRNA, mRNA Exons, as well as cancer-associated protein markers that are specific for solid-tissue cancers and matched normal tissues. The present application teaches the art of selecting the desired number of markers and types of markers for both pan-oncology and specific cancers (*i.e.* colorectal cancer) to guide the physician to improve the treatment of the patient. Details on primer design and optimized primer sequences are provided to enable rapid validation of these tests for both pan-oncology and specific cancers. The two-step procedure is designed to cast a wide net to initially identify most of the individuals harboring an early cancer, followed by a more stringent second step to improve specificity and narrow the patients to those most likely to harbor a hidden cancer, who are then sent for imaging and followup. The advantage of this 2-step approach is that it not only identifies the potential tissue of origin, but it is designed to provide the highest positive predictive value (PPV). Thus, when a result for a rare cancer comes back as presumptive positive (*i.e.* early ovarian cancer) the physician can focus her attention on providing imaging and followup to those patients who need

it the most, while the test minimizes the false-positives that create unnecessary anxiety and unwanted invasive procedures.

[0041] The present application provides robust approaches for detecting markers of cancer (e.g., mutations, expression, splice variants, translocations, copy number, and/or methylation changes) using either qPCR or dPCR readout using protocols that are amenable to automation and work on readily available commercial instruments. The approach provides advantages in being integrated and convenient for laboratory setup, allowing for cost reduction, scalability, and fit with medical and laboratory flow in a CLIA-compatible automated setting. The benefit in lives saved world-wide would be of incalculable value.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] Figures 1A-D illustrate a conditional logic tree for an early detection colorectal cancer test based on analysis of a patient's blood sample. Figure 1A illustrates a one-step colorectal cancer assay using 12 markers at average sensitivity of 75%. Figure 1B illustrates a two-step colorectal cancer assay using 12 markers at average sensitivity of 75% in the first step, and 24 markers at average sensitivity of 75% in the second step. Figure 1C illustrates a one-step colorectal cancer assay using 18 markers at average sensitivity of 75%. Figure 1D illustrates a two-step colorectal cancer assay using 18 markers at average sensitivity of 75% in the first step, and 36 markers at average sensitivity of 75% in the second step. Figures 1E-L illustrate a conditional logic tree for a two-step assay for an early detection pan-oncology cancer test based on analysis of a patient's blood sample. Figure 1E illustrates a two-step pan-oncology assay using 96 group-specific markers at average sensitivity of 50% in the first step, followed by 1 or 2 groups of 64 type-specific markers each at average sensitivity of 50% in the second step. Figure 1F illustrates a two-step pan-oncology assay using 96 group-specific markers at average sensitivity of 50% in the first step, followed by 1 or 2 groups of 48 group-specific markers each at average sensitivity of 75% in the second step. Figure 1G illustrates a two-step pan-oncology assay using 48 cancer-specific markers at average sensitivity of 75% in the first step, followed by 96 type-specific markers each at average sensitivity of 50% in the second step. Figure 1H illustrates a two-step pan-oncology assay using 64 cancer-specific markers at average sensitivity of 75% in the first step, followed by 96 type-specific markers each at average sensitivity of 50% in the second step. Figure 1I illustrates a two-step pan-oncology assay using 96 group-specific markers at average sensitivity of 66% in the first step, followed by 1 or 2 groups of 64 type-specific markers each at average sensitivity of 66% in the second step. Figure 1J illustrates a two-step pan-oncology assay using 96 group-specific markers at average sensitivity of 66% in

the first step, followed by 1 or 2 groups of 48 group-specific markers each at average sensitivity of 75% in the second step. Figure 1K illustrates a two-step pan-oncology assay using 48 cancer-specific markers at average sensitivity of 75% in the first step, followed by 96 type-specific markers each at average sensitivity of 66% in the second step. Figure 1L illustrates a two-step pan-oncology assay using 64 cancer-specific markers at average sensitivity of 75% in the first step, followed by 96 type-specific markers each at average sensitivity of 66% in the second step. Figure 1M illustrates a conditional logic tree for a two-step assay to guide and monitor cancer treatment based on analysis of a patient's blood sample. The sample is analyzed with a targeted cancer-specific gene panel to identify mutations to guide therapy. Meanwhile the tumor or plasma is tested with 48 group-specific markers at average sensitivity of 75% to identify 12-24 markers specific to that patient. These markers are subsequently used to monitor treatment efficacy. Figure 1N illustrates a conditional logic tree for a two-step assay to monitor for cancer recurrence based on analysis of a patient's blood sample. The tumor or plasma is tested with 48 group-specific markers at average sensitivity of 75% to identify 12-24 markers specific to that patient. After the patient is deemed cancer-free, these markers are used to identify early recurrence. Samples from patients that cross a threshold are then subjected to a targeted cancer-specific gene panel to verify presence of original mutations, and identify mutations to guide therapy and treat the recurrence.

[0043] Figure 2 illustrates exPCR-LDR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0044] Figure 3 illustrates exPCR-LDR-qPCR carryover prevention reaction with UniTaq detection to identify or relatively quantify low-level methylation.

[0045] Figure 4 illustrates a variation of exPCR-LDR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0046] Figure 5 illustrates exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0047] Figure 6 illustrates exPCR-qPCR carryover prevention reaction with UniTaq detection to identify or relatively quantify low-level methylation.

[0048] Figure 7 illustrates a variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0049] Figure 8 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0050] Figure 9 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0051] Figure 10 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0052] Figure 11 illustrates another variation of exPCR-LDR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0053] Figure 12 illustrates another variation of exPCR-LDR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0054] Figure 13 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0055] Figure 14 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0056] Figure 15 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0057] Figure 16 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0058] Figure 17 illustrates another variation of exPCR-qPCR carryover prevention reaction with Taqman™ detection to identify or relatively quantify low-level methylation.

[0059] Figures 18A-B illustrate results for calculated overall Sensitivity and Specificity for a 24-marker assay, where the average individual marker sensitivity is 50% (Figure 18A), and the average individual marker false-positive rate is from 2% to 5% (Figure 18B).

[0060] Figures 19A-B illustrate results for calculated overall Sensitivity and Specificity for a 36-marker assay, where the average individual marker sensitivity is 50% (Figure 19A), and the average individual marker false-positive rate is from 2% to 5% (Figure 19B).

[0061] Figures 20A-B illustrate results for calculated overall Sensitivity and Specificity for a 48-marker assay, where the average individual marker sensitivity is 50% (Figure 20A), and the average individual marker false-positive rate is from 2% to 5% (Figure 20B).

[0062] Figures 21A-B illustrate the ROC curve for a 48-marker assay, where the average individual marker sensitivity is 50%, as well as the calculated AUC, when the average number of molecules per marker in the blood ranges from 150 to 600 molecules. For Figures 21A and 21B, the calculations are based on an average individual marker false-positive rate of 2% and 3%, respectively.

[0063] Figures 22A-B illustrate the ROC curve for a 48-marker assay, where the average individual marker sensitivity is 50%, as well as the calculated AUC, when the average number of molecules per marker in the blood ranges from 150 to 600 molecules. For Figures 22A and 22B,

the calculations are based on an average individual marker false-positive rate of 4% and 5%, respectively.

[0064] Figures 23A-B provide a list of blood-based, colon cancer-specific microRNA markers derived through analysis of TCGA microRNA datasets, which may be present in exosomes or other protected state in the blood.

[0065] Figures 24A-X provide a list of blood-based, colon cancer-specific ncRNA and lncRNA markers, which may be present in exosomes or other protected state in the blood.

[0066] Figures 25A-C provide a list of blood-based colon cancer-specific exon transcripts that may be enriched in exosomes or other protected states in the blood.

[0067] Figures 26A-J provide a list of cancer proteins markers, identified through, mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from Colorectal tumors, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0068] Figure 27 provides a list of protein markers that can be secreted by Colorectal tumors into the blood.

[0069] Figures 28A-Y provide a list of primary CpG sites that are Colorectal cancer and colon-tissue specific markers, that may be used to identify the presence of colorectal cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0070] Figures 29A-P provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are Colorectal cancer and colon-tissue specific markers, that may be used to identify the presence of Colorectal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood.

[0071] Figures 30A-B illustrate results for calculated overall Sensitivity and Specificity for a 24-marker assay, where the average individual marker sensitivity is 66% (Figure 30A), and the average individual marker false-positive rate is from 2% to 5% (Figure 30B).

[0072] Figures 31A-B illustrate results for calculated overall Sensitivity and Specificity for a 36-marker assay, where the average individual marker sensitivity is 66% (Figure 31A), and the average individual marker false-positive rate is from 2% to 5% (Figure 31B).

[0073] Figures 32A-B illustrate results for calculated overall Sensitivity and Specificity for a 48-marker assay, where the average individual marker sensitivity is 66% (Figure 32A), and the average individual marker false-positive rate is from 2% to 5% (Figure 32B).

[0074] Figures 33A-B illustrate results for calculated overall Sensitivity and Specificity for a 12-marker assay, where the average individual marker sensitivity is 75% (Figure 33A), and the average individual marker false-positive rate is from 2% to 5% (Figure 33B).

[0075] Figures 34A-B illustrate results for calculated overall Sensitivity and Specificity for a 18-marker assay, where the average individual marker sensitivity is 75% (Figure 34A), and the average individual marker false-positive rate is from 2% to 5% (Figure 34B).

[0076] Figures 35A-B illustrate results for calculated overall Sensitivity and Specificity for a 24-marker assay, where the average individual marker sensitivity is 75% (Figure 35A), and the average individual marker false-positive rate is from 2% to 5% (Figure 35B).

[0077] Figures 36A-B illustrate results for calculated overall Sensitivity and Specificity for a 32-marker assay, where the average individual marker sensitivity is 75% (Figure 36A), and the average individual marker false-positive rate is from 2% to 5% (Figure 36B).

[0078] Figures 37A-B illustrate results for calculated overall Sensitivity and Specificity for a 36-marker assay, where the average individual marker sensitivity is 75% (Figure 37A), and the average individual marker false-positive rate is from 2% to 5% (Figure 37B).

[0079] Figures 38A-B illustrate results for calculated overall Sensitivity and Specificity for a 48-marker assay, where the average individual marker sensitivity is 75% (Figure 38A), and the average individual marker false-positive rate is from 2% to 5% (Figure 38B).

[0080] Figure 39 provides a list of blood-based, solid tumor-specific ncRNA and lncRNA markers, which may be present in exosomes or other protected state in the blood.

[0081] Figures 40A-F provide a list of candidate blood-based solid tumor-specific exon transcripts that may be enriched in in exosomes or other protected state in the blood.

[0082] Figures 41A-H provide a list of cancer proteins markers, identified through, mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from solid tumors, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0083] Figures 42A-S provide a list of primary CpG sites that are Solid-tumor and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0084] Figures 43A-J provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are Solid-tumor and tissue-specific markers, that may be used to

identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0085] Figure 44 provides a list of cancer proteins markers, identified through, mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0086] Figures 45A-S provide a list of primary CpG sites that are colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0087] Figures 46A-J provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0088] Figures 47A-C provide a list of primary CpG sites that are breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0089] Figures 48A-B provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0090] Figure 49 provides a list of primary CpG sites that are lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0091] Figure 50 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are lung adenocarcinoma, lung squamous cell carcinoma, or head & neck

squamous cell carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0092] Figure 51 provides a list of primary CpG sites that are prostate adenocarcinoma or invasive urothelial bladder cancer and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0093] Figure 52 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are prostate adenocarcinoma or invasive urothelial bladder cancer and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0094] Figure 53 provides a list of blood-based, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma-specific ncRNA and lncRNA markers, which may be present in exosomes or other protected state in the blood.

[0095] Figures 54A-E provide a list of candidate blood-based liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma-specific exon transcripts that may be enriched in exosomes or other protected state in the blood.

[0096] Figures 55A-B provide a list of cancer proteins markers, identified through, mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0097] Figures 56A-E provide a list of primary CpG sites that are liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0098] Figures 57A-C provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0099] Figures 58A-D provide a list of primary CpG sites that are Solid-tumor and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0100] Figures 59A-C provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are Solid-tumor and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0101] Figures 60A-D provide a list of primary CpG sites that are colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0102] Figures 61A-D provide a list of chromosomal regions or sub-regions within which are primary CpG sites that are colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0103] Figure 62 provides a list of primary CpG sites that are breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0104] Figure 63 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0105] Figure 64 provides a list of primary CpG sites that are lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0106] Figure 65 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are lung adenocarcinoma, lung squamous cell carcinoma, or head & neck

squamous cell carcinoma and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0107] Figure 66 provides a list of primary CpG sites that are prostate adenocarcinoma or invasive urothelial bladder cancer and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0108] Figure 67 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are prostate adenocarcinoma or invasive urothelial bladder cancer and tissue-specific markers, that may be used to identify the presence of solid-tumor cancer from cfDNA, or DNA within exosomes, or DNA in another protected state (such as within CTCs) within the blood.

[0109] Figure 68 provides a list of blood-based, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma-specific ncRNA and lncRNA markers, which may be present in exosomes or other protected state in the blood.

[0110] Figure 69 provides a list of candidate blood-based liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma-specific exon transcripts that may be enriched in in exosomes or other protected state in the blood.

[0111] Figures 70A-B illustrate the real-time PCR amplification plots obtained in a multiplexed detection of 20 CRC methylation markers by TET-APOBEC-exPCR-LDR-qPCR, using reverse primers with long tails, using 1 µg (starting) of sonicated HT29 cell line DNA, without methyl capture (Figure 70A), and without methyl capture (Figure 70B).

[0112] Figures 71A-B illustrate the real-time PCR amplification plots obtained in a multiplexed detection of 20 CRC methylation markers by Methyl Captured and TET-APOBEC-exPCR-LDR-qPCR, using reverse primers with long tails, using 1 µg of sonicated HT29 cell line DNA (Figure 71A), and with 1 µg of sonicated normal DNA (Figure 71B).

[0113] Figures 72A-B illustrate the real-time PCR amplification plots obtained in a multiplexed detection of 20 CRC methylation markers by Bisulfite-exPCR-LDR-qPCR, using reverse primers with long tails, using HT29 cell line DNA, with 200 genome equivalents of HT29 cell line DNA in 7,500 genome equivalents of normal, *e.g.* unmethylated DNA (Roche DNA) at 25 nM initial primer concentration for the extension reaction; without (Figure 72A) and with addition of 1,000 nM Universal Primer during the first PCR reaction (Figure 72B).

[0114] Figures 73A-B illustrate the real-time PCR amplification plots obtained in a multiplexed detection of 20 CRC methylation markers by Bisulfite-exPCR-LDR-qPCR, using

reverse primers with long tails, using HT29 cell line DNA, with 200 genome equivalents of HT29 cell line DNA in 7,500 genome equivalents of normal, *e.g.* unmethylated DNA (Roche DNA) at 12 nM initial primer concentration for the extension reaction; without (Figure 73A) and with addition of 1,000 nM Universal Primer during the first PCR reaction (Figure 73B).

DETAILED DESCRIPTION

A Universal Design for Early Detection of Cancer Using “Cancer Marker Load”

[0115] The most cost-effective early cancer detection test may combine an initial multiplexed coupled amplification and ligation assay to determine “cancer load”. For early cancer detection, this would achieve > 95% sensitivity for all cancers (pan-oncology), at > 97% specificity. These design principles may also be extended to include monitoring the efficacy of treatment, as well as detecting early cancer recurrence.

[0116] Several flow charts for cancer tumor load assays are illustrated in Figure 1. In its simplest form, the assay would be a one-step assay to identify individuals with early colorectal cancer (CRC). A blood sample is fractionated into plasma and other components as needed, a set of 12 markers with average sensitivity of 75% are assayed, and the results are recorded (Figure 1A). For example, an initial multiplexed PCR/LDR screening assay scoring for mutation, methylation, miRNA, mRNA, alternative splicing, and/or translocations identifies those samples with positive results. The physician is not concerned with which specific markers are positive but gives a simple directive. Those patients with 0-1 markers positive are told not to worry, go home, you are cancer-free. Those patients with ≥ 3 of 12 markers positive are directed to get a colonoscopy. Those patients with an intermediate number of positive markers (2) are instructed to come back in 3-6 months for retesting. Thus, the test is based on the overall cancer marker load and not dependent on the specific markers that test positive.

[0117] In an advanced version of the test, a two-step assay would be performed to identify if the patient has colorectal cancer. The rationale for a two-step test is initially cast a wide net to maximize sensitivity in identifying the most individuals with potential cancer, followed by a second step only on the positive samples (which contain both true and false-positives) to maximize specificity, eliminate virtually all the false-positives and hone in on those individuals most likely to have cancer. In the first step, a blood sample is fractionated into plasma and other components as needed, followed by an assay to interrogate an initial set of 12 markers with an average sensitivity of 75% (Figure 1B). The first step assay can employ multiplexed PCR/LDR, or digital PCR screening to score for mutation, methylation, miRNA,

mRNA, alternative splicing, and/or translocations events. As in the one-step assay, patients with 0-1 markers positive are presumed to be cancer-free. On the other hand, patients with ≥ 2 markers positive will undergo a second step, wherein 24 (new) markers with 75% sensitivity are assayed and scored as follows: 0-2 positive markers are considered cancer-free; 3 positive markers are advised to come back in 3-6 months for retesting; ≥ 4 positive markers are directed to go get a colonoscopy.

[0118] For higher accuracy of the one step CRC test, after fractionating the blood sample, a set of 18 markers with average sensitivity of 75% are assayed, and the results are recorded (Figure 1C). Those patients with 0-2 markers positive are considered cancer-free; 3 positive markers are advised to come back in 3-6 months for retesting; ≥ 4 positive markers are directed to go get a colonoscopy.

[0119] For higher accuracy of the two step CRC test, after fractionating the blood sample, a set of 18 markers with average sensitivity of 75% are assayed, and the results are recorded (Figure 1D). As in the one-step assay, patients with 0-2 markers positive are presumed to be cancer-free. On the other hand, patients with ≥ 3 markers positive will undergo a second step, wherein 36 (new) markers with 75% sensitivity are assayed and scored as follows: 0-3 positive markers are considered cancer-free; 4 positive markers are advised to come back in 3-6 months for retesting; ≥ 5 positive markers are directed to go get a colonoscopy.

[0120] In a pan-oncology version of the test, in the first step the assay would screen 96 markers, wherein on average ≥ 36 such markers would exhibit an average sensitivity of 50% for most major cancers (see Figure 1E). These cancers would cluster to certain groups, which include: Group 1 (Colorectal, Stomach, Esophagus); Group 2 (Breast, Endometrial, Ovarian, Cervical, Uterine); Group 3 (Lung, Head & Neck); Group 4 (Prostate, Bladder), & Group 5 (Liver, Pancreatic, Gall Bladder). Patients with 0-4 markers positive are presumed to be cancer-free, while patients with ≥ 5 markers positive will undergo a second step. Presumptive positive samples are then assayed in the second step testing 1 or 2 groups, using 64 markers per group, wherein on average ≥ 36 such markers would exhibit an average sensitivity of 50% for each specific types of cancer within that group, including using tissue-specific markers to validate the initial result, and to identify tissue of origin. Results are scored as follows: 0-3 positive markers are considered cancer-free; 4 positive markers are advised to come back in 3-6 months for retesting; ≥ 5 positive markers are directed to go to imaging that matches the type(s) of cancer most likely to be the tissue of origin. For higher sensitivities, both the initial 96 markers in the first step, and the group-specific markers in the second step would have average sensitivity of

66% (Figure 1I). The physician may then order targeted sequencing to further guide treatment decisions for the patient.

[0121] In a variation of the pan-oncology test, in the first step the assay would screen 96 markers, wherein on average ≥ 36 such markers would exhibit an average sensitivity of 50% for most major cancers (see Figure 1F). Patients with 0-4 markers positive are presumed to be cancer-free, while patients with ≥ 5 markers positive will undergo a second step. Presumptive positive samples are then assayed in the second step testing 1 or 2 groups, using 48 markers per group, wherein on average ≥ 36 such markers would exhibit an average sensitivity of 75% for each specific types of cancer within that group. Results are scored as follows: 0-3 positive markers are considered cancer-free; 4 positive markers are advised to come back in 3-6 months for retesting; ≥ 5 positive markers are directed to go to imaging that matches the type(s) of cancer most likely to be the tissue of origin. These sets of group-specific markers may not always identify the exact tissue of origin, but they should narrow it down to a specific group. As an alternative approach to identifying the tissue of origin, methylation markers may be scored using targeted bisulfite sequencing, to access more or additional methylation markers, instead of, or in addition to the step 2 above. For higher sensitivities, the initial 96 markers in the first step would have average sensitivity of 66% (Figure 1J). The physician may then order targeted sequencing to further guide treatment decisions for the patient.

[0122] In a more streamlined version of the pan-oncology test, in the first step the assay would screen 48 markers, wherein on average ≥ 24 such markers would exhibit an average sensitivity of 75% for most major cancers (a pan-oncology test, see Figure 1G). Patients with 0-3 markers positive are presumed to be cancer-free, while patients with ≥ 4 markers positive will undergo a second step. For higher accuracy in the initial screen, the first step the assay would screen 64 markers, wherein on average ≥ 36 such markers would exhibit an average sensitivity of 75% for most major cancers (see Figure 1H). Here, patients with 0-4 markers positive are presumed to be cancer-free, while patients with ≥ 5 markers positive will undergo a second step. Presumptive positive samples are then assayed in the second step using the 96 marker pan-oncology assay, wherein on average ≥ 36 such markers would exhibit an average sensitivity of 50% for each specific types of cancer within that group, including using tissue-specific markers to validate the initial result, and to identify tissue of origin. Results are scored as follows: 0-3 positive markers are considered cancer-free; 4 positive markers are advised to come back in 3-6 months for retesting; ≥ 5 positive markers are directed to go to imaging that matches the type(s) of cancer most likely to be the tissue of origin. For higher sensitivities, the 96 markers in the second step would have average sensitivity of 66% (Figures 1K & 1L). As an alternative

approach to identifying the tissue of origin, methylation markers may be scored using targeted bisulfite sequencing, to access more or additional methylation markers, instead of, or in addition to the step 2 above. The physician may then order targeted sequencing to further guide treatment decisions for the patient.

[0123] The aforementioned 5 groups of 48 markers, with average sensitivity of 75% were designed to also be used to monitor treatment (see Figure 1M). Currently, with a newly diagnosed cancer, cancer tissue (or liquid biopsy) is subjected to targeted sequencing to identify mutations or gene rearrangements that may be used to guide therapy. For a given cancer (*i.e.* stomach cancer in Group 1), the cancer tissue or liquid biopsy may be tested with the 48-marker group (1) panel. If the cancer had been identified in the first place using the 2-step screens identified in Figures 1F or 1J, then they will have already undergone the 48-marker group specific test in step 2 of that assay. Of the 48 markers tested, on average 12-24 would be positive. These may then be bundled together in a patient-specific test to monitor treatment efficacy. The plasma of such a patient would be tested post surgery, and during the treatment regimen. The plasma is monitored for loss of the 12-24 marker signal, but if ≥ 3 positive markers remain positive, then this may guide the physician to change therapy.

[0124] The aforementioned 5 groups of 48 markers were designed to also be used to monitor for recurrence (see Figure 1N). If the cancer had been identified in the first place using the 2-step screens identified in Figures 1F and 1J, and/or was monitored as described in Figure 1M, then they will have already undergone the 48-marker group specific test, for which on average 12-24 would be positive. These may then be bundled together in a patient-specific test to monitor for recurrence. The plasma of such a patient who recovered from the original cancer would be monitored for gain of markers from the 12-24 marker panel. Results are scored as follows: 0-2 positive markers are considered cancer-free; ≥ 3 positive markers are directed to go to the second step. The plasma would be subjected to targeted sequencing to identify mutations or gene rearrangements that may be used to guide therapy of the recurrent tumor.

[0125] The present application is directed to a universal diagnostic approach that seeks to combine the best features of digital polymerase chain reaction (PCR), or quantitative polymerase chain reaction (qPCR), with using TET2 for conversion of 5mC (5-methyl cytosine) and 5hmC (5-hydroxy-methyl cytosine) through a cascade reaction into 5-carboxycytosine [*i.e.* 5-methylcytosine (5mC) $\rightarrow$ 5-hydroxymethylcytosine (5hmC) $\rightarrow$ 5-formylcytosine (5fC) $\rightarrow$ 5-carboxycytosine (5caC)], thus protecting 5mC and 5hmC, but not unmethylated C from deamination by APOBEC, (see Technical Report and Protocol with New England Biolabs product: NEBNext Enzymatic Methyl-seq Kit E7120, which is hereby incorporated by reference

in its entirety), ligation detection reaction (LDR), and quantitative detection of multiple disease markers, *e.g.*, cancer markers.

Multiplexing, Avoiding False-Positives, and Carryover Protection

[0126] There is a technical challenge of distinguishing true signal generated from the desired disease-specific nucleic acid differences vs. false signal generated from normal nucleic acids present in the sample vs. false signal generated in the absence of the disease-specific nucleic acid differences (*i.e.* somatic mutations).

[0127] A number of solutions to these challenges are presented below, but they share some common themes.

[0128] The first theme is multiplexing. PCR works best when primer concentration is relatively high, from 50nM to 500nM, limiting multiplexing. Further, the more PCR primer pairs added, the chances of amplifying incorrect products or creating primer-dimers increase exponentially. In contrast, for LDR probes, low concentrations on the order of 4 nM to 20 nM are used, and probe-dimers are limited by the requirement for adjacent hybridization on the target to allow for a ligation event. Use of low concentrations of gene-specific PCR primers or LDR probes containing universal primer sequence “tails” allows for subsequent addition of higher concentrations of universal primers to achieve proportional amplification of the initial PCR or LDR products. Another way to avoid or minimize false PCR amplicons or primer dimers is to use PCR primers containing a few extra bases and a blocking group, which is liberated to form a free 3'OH by cleavage with a nuclease only when hybridized to the target, *e.g.*, a ribonucleotide base as the blocking group and RNase H2 as the cleaving nuclease.

[0129] The second theme is fluctuations in signal due to low input target nucleic acids. Often, the target nucleic acid originated from a few cells, either captured as CTCs, or from tumor cells that underwent apoptosis and released their DNA as small fragments (140–160 bp) in the serum. Under such conditions, it is preferable to perform some level of proportional amplification to avoid missing the signal altogether or reporting inaccurate copy number due to fluctuations when distributing small numbers of starting molecules into individual wells (for real-time, or droplet PCR quantification). As long as these initial amplifications are kept at a reasonable level (approximately 12 to 20 cycles), the risk of carryover contamination during opening of the tube and distributing amplicons for subsequent detection/quantification (using real-time, or droplet PCR) is minimized. Other schemes use even lower amounts of limited amplifications (approximately 8 to 12 cycles).

[0130] The third theme is target-independent signal, also known as “No Template Control” (NTC). This arises from either polymerase or ligase reactions that occur in the absence

of the correct target. Some of this signal may be minimized by judicious primer design. For ligation reactions, the 5' → 3' nuclease activity of polymerase may be used to liberate the 5' phosphate of the downstream ligation primer (only when hybridized to the target), so it is suitable for ligation. Further specificity for distinguishing presence of a low-level mutation using LDR may be achieved by: (i) using upstream mutation-specific LDR probes containing a mismatch in the 2nd or 3rd position from the 3' OH base, (ii) using LNA or PNA probes to wild-type sequence that would reduce hybridization of mutation-specific LDR probes to wild-type sequences, (iii) using LDR probes to wild-type sequence that (optionally) ligate but do not undergo additional amplification, and (iv) using upstream LDR probes containing a few extra bases and a blocking group, which is liberated to form a free 3' OH by cleavage with a nuclease only when hybridized to the complementary target (*e.g.*, RNase H2 and a ribonucleotide base). Similar approaches for improving the specificity for distinguishing presence of a low-level mutation using PCR may be achieved by: (i) using mutation-specific PCR primers containing a mismatch in the 2nd or 3rd position from the 3' OH base, (ii) using LNA or PNA probes to wild-type sequence that would reduce hybridization of mutation-specific PCR primers to wild-type sequences, (iii) using PCR primers to wild-type sequence that are blocked and do not undergo additional amplification, and (iv) using upstream PCR primers containing a few extra bases and a blocking group, which is liberated to form a free 3' OH by cleavage with a nuclease only when hybridized to the complementary target (*e.g.*, RNase H2 and a ribonucleotide base).

[0131] The fourth theme is either suppressed (reduced) amplification or incorrect (false) amplification due to unused primers in the reaction. One approach to eliminate such unused primers is to capture genomic or target or amplified target DNA on a solid support, allow ligation probes to hybridize and ligate, and then remove probes or products that are not hybridized. Alternative solutions include pre-amplification, followed by subsequent nested LDR and/or PCR steps, such that there is a second level of selection in the process.

[0132] The fifth theme is carryover prevention. Carryover signal may be eliminated by standard uracil incorporation during the universal PCR amplification step, and by using UDG (and optionally AP endonuclease) in the pre-amplification workup procedure. Incorporation of carryover prevention is central to the methods of the present application as described in more detail below. The initial PCR amplification is performed using incorporation of uracil. The LDR reaction is performed with LDR probes lacking uracil. Thus, when the LDR products are subjected to real-time PCR quantification, addition of UDG destroys the initial PCR products, but not the LDR products. Further, since LDR is a linear process and the tag primers use sequences absent from the human genome, accidental carryover of LDR products back to the

original PCR will not cause template-independent amplification. Additional schemes to provide carryover prevention with methylated targets include use of restriction endonucleases to destroy unmethylated DNA prior to PCR amplification, or by capturing and enriching methylated DNA using methyl-specific DNA binding proteins or antibodies.

[0133] The sixth theme is achieving even amplification of many mutation-specific or methylation-specific targets in the multiplexed reaction. One approach, as already described above, is to perform limited initial PCR amplifications (8 to 12, or 12 to 20 cycles). However, sometimes different products amplify at different rates, especially when using mutation- or methylation-specific primers, or when using blocking LNA or PNA probes or other means to suppress amplification of wild-type DNA. This is because a regular PCR reaction has both forward and reverse primers working simultaneously. Although there may be preferential amplification using as an example a reverse methylation-specific primer (*i.e.* after using TET2 and APOBEC treatment), the forward primer will amplify both methylated and un-methylated DNA (again, after using TET2 and APOBEC treatment), and thus will magnify differences in initial rates of forward primer amplification. Further, and this also holds when using mutation-specific forward primers, the use of non-selecting reverse primers means that initial amplification products still contain substantial amounts of wild-type DNA sequence, which may lead to undesired false-positives in subsequent amplification steps. One approach is to perform an initial single-sided linear amplification, using primers that amplify only one strand of target DNA. This is particularly useful when amplifying TET2 and APOBEC-treated DNA, where the two resultant strands are no longer complementary to each other. An important variation of this theme destroys the initial target DNA after the linear amplification step. This may be achieved by incorporating one or more modified nucleotides, such as α -thio-dNTPs, that protect the initial extension products (but not the original cfDNA or genomic DNA) from exonuclease I digestion.

Methods of Identifying Cancer Methylation and Hydroxymethylation Markers

[0134] A first aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to

convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil residues to produce a treated sample. One or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules are provided, and one or more primary oligonucleotide primer sets are provided. Each primary oligonucleotide primer set comprises (a) a first primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer, wherein the first or second primary oligonucleotide primer further comprises a 5' primer-specific portion. The treated sample, the one or more first primary oligonucleotide primers of the primer sets, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof. One or more oligonucleotide probe sets are then provided. Each probe set comprises (a) a first oligonucleotide probe having a 5' primer-specific portion and a 3' DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and (b) a second oligonucleotide probe having a 5' DNA repair

enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion and a 3' primer-specific portion, and wherein the first and second oligonucleotide probes of a probe set are configured to hybridize, in a base specific manner, on a complementary nucleotide sequence of a first polymerase chain reaction product. The first polymerase chain reaction products are blended with a ligase and the one or more oligonucleotide probe sets to form one or more ligation reaction mixtures. The one or more ligation reaction mixtures are subjected to one or more ligation reaction cycles whereby the first and second oligonucleotide probes of the one or more oligonucleotide probe sets are ligated together, when hybridized to their complementary sequences, to form ligated product sequences in the ligation reaction mixture wherein each ligated product sequence comprises the 5' primer-specific portion, the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and the 3' primer-specific portion. The method further involves providing one or more secondary oligonucleotide primer sets. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the ligated product sequence and (b) a second secondary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the ligated product sequence. The ligated product sequences, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further comprises detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0135] Figures 2 and 3 illustrate exPCR-LDR-qPCR carryover prevention reaction to detect low-level methylation in accordance with this aspect of the present application. After isolating the genomic or cfDNA, it is optionally treated with a DNA repair kit (Figures 2 and 3, Step A). Subsequently, the DNA is treated with ten-eleven translocation (TET2) dioxygenase

for conversion of 5mC (5-methyl cytosine) and 5hmC (5-hydroxy-methyl cytosine) through a cascade reaction into 5caC (5-carboxycytosine), thus protecting 5mC and 5hmC, but not unmethylated C from deamination by apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC cytidine deaminase), (see Technical Report and Protocol with New England Biolabs product: NEBNext Enzymatic Methyl-seq Kit E7120, which is hereby incorporated by reference in its entirety). DNA Polymerase inserts an “A” base opposite the deaminated C (in other words, dU) but a “G” opposite the 5caC, which is resistant to deamination by APOBEC. Thus, the combination of TET2 followed by APOBEC effectively converts C, but not 5mC or 5hmC to “T” in the DNA sequence. The regions of interest are selectively extended using locus-specific downstream primers comprising 5’ universal primer sequences and 3’ target-specific sequences, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3’ cleavable blocking group (Blk 3’, *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3’OH group which is suitable for polymerase extension (Figures 2 or 3, step B; *see e.g.*, Dobosy et. al. “RNase H-Dependent PCR (rhPCR): Improved Specificity and Single Nucleotide Polymorphism Detection Using Blocked Cleavable Primers,” *BMC Biotechnology* 11(80):1011 (2011), which is hereby incorporated by reference in its entirety). The sample is treated with UDG or similar enzyme to remove dU containing TET2-APOBEC treated input DNA. Suitable enzymes include, without limitation, *E. coli* uracil DNA glycosylase (UDG), Antarctic Thermolabile UDG, or Human single-strand-selective monofunctional uracil-DNA Glycosylase (hSMUG1). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the locus-specific upstream primers are added, followed by limited (8 to 20 cycles) or full (20-40 cycles) PCR using a deoxynucleotide mix that includes dUTP (Figure 2, step C). Upon target-specific hybridization, RNase H removes the RNA base to liberate a 3’OH group which is a few bases upstream of the TET2 converted methylated (or hydroxymethylated) target base, and suitable for polymerase extension (Figure 2, step C). An optional blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated sequence (or its complement) that partially overlaps with the upstream PCR primer will preferentially compete for binding to the TET2-APOBEC converted unmethylated sequence over the upstream primer, thus suppressing amplification of TET2-APOBEC converted unmethylated sequence DNA during each round of PCR. The downstream primers contain identical universal primer tails to prevent primer dimers. Further, such tails provide the option for including Universal primer during the PCR step. This may

assist in generating more equal amounts of products in a multiplexed PCR reaction. The amplified products contain dU as shown in Figure 2, step D, which allows for subsequent treatment with UDG or a similar enzyme for carryover prevention.

[0136] As shown in Figure 2, step E, target-specific oligonucleotide probes are hybridized to the amplified products and ligase (filled circle) covalently seals the two oligonucleotides together when hybridized to their complementary sequence. In this embodiment, the upstream oligonucleotide probe having a sequence specific for detecting the 5-methyl-C or 5-hydroxymethyl-C region of interest further contains a 5' primer-specific portion (Ai) to facilitate subsequent detection of the ligation product. Once again, the presence of blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated sequence suppresses ligation to TET2-APOBEC converted unmethylated target sequence if present after the enrichment of methylated or hydroxymethylated sequence during the PCR amplification step. The downstream oligonucleotide probe, having a sequence common to both TET2-APOBEC converted methylated and unmethylated sequences contains a 3' primer-specific portion (Ci') that, together with the 5' primer specific portion (Ai) of the upstream probe having a sequence specific for detecting the methylated or hydroxymethylated region, permit subsequent amplification and detection of only the desired ligation products. As illustrated in step E of Figure 2, another layer of specificity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream ligation probe. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to generate a ligation competent 3'OH group (Figure 2, step E).

[0137] As shown in Figure 2, step F, target-specific oligonucleotide probes are hybridized to the amplified products and ligase (filled circle) covalently seals the two oligonucleotides together when hybridized to their complementary sequence. The upstream oligonucleotide probe contains a 5' primer-specific portion (Ai) and the downstream oligonucleotide probe contains a 3' primer-specific portion (Ci') that permits subsequent amplification of the ligation product. Following ligation, the ligation products are aliquoted into separate wells, micro-pores or droplets containing one or more tag-specific primer pairs, each pair comprising matched primers Ai and Ci, treated with UDG or similar enzyme to remove dU containing amplification products or contaminants, PCR amplified, and detected. As shown in Figures 2, steps G & H, detection of the ligation product can be carried out using traditional TaqMan™ detection assay (*see* U.S. Patent No. 6,270,967 to Whitcombe et al., and U.S. Patent No. 7,601,821 to Anderson et al., which are hereby incorporated by reference in their entirety). For detection using TaqMan™ an oligonucleotide probe spanning the ligation junction is used in

conjunction with primers suitable for hybridization on the primer-specific portions of the ligation products for amplification and detection. The TaqMan™ probe contains a fluorescent reporter group on one end (F1) and a quencher molecule (Q) on the other end that are in close enough proximity to each other in the intact probe that the quencher molecule quenches fluorescence of the reporter group. During amplification, the TaqMan™ probe and upstream primer hybridize to their complementary regions of the ligation product. The 5' → 3' nuclease activity of the polymerase extends the hybridized primer and liberates the fluorescent group of the TaqMan™ probe to generate a detectable signal (Figure 2, step H). In a preferred embodiment, the Taqman probe contains a second quencher group (ZEN) about 9 bases in from the fluorescent reporter group, and the probe is designed such that the ZEN group is at or adjacent to the mutant base. Use of dUTP during the amplification reaction generates products containing dU, which can subsequently be destroyed using UDG for carryover prevention.

[0138] As shown in Figure 3, step D, target-specific oligonucleotide probes are hybridized to the amplified products and ligase (filled circle) covalently seals the two oligonucleotides together when hybridized to their complementary sequence. In this embodiment, the upstream oligonucleotide probe having a sequence specific for detecting the mutation of interest further contains a 5' primer-specific portion (Ai) to facilitate subsequent detection of the ligation product. Once again, the presence of blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated sequence suppresses ligation to TET2-APOBEC converted unmethylated target sequence if present after the enrichment of TET2-APOBEC converted methylated or hydroxymethylated sequence during the PCR amplification step. The downstream oligonucleotide probe, having a sequence common to both TET2-APOBEC converted unmethylated and methylated (or hydroxymethylated) sequences contains a 3' primer-specific portion (Bi-Ci') that, together with the 5' primer specific portion (Ai) of the upstream probe having a sequence specific for detecting the methylated or hydroxymethylated region, permit subsequent amplification and detection of only the desired ligation products. As illustrated in step D of Figure 3, another layer of specificity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream ligation probe. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to generate a ligation competent 3'OH group (Figure 3, step D).

[0139] In this embodiment, the ligation probes are designed to contain UniTaq primer and tag sequences to facilitate detections. The UniTaq system is fully described in U.S. Patent Application Publication No. 2011/0212846 to Spier, which is hereby incorporated by reference in its entirety. The UniTaq system involves the use of three unique "tag" sequences, where at

least one of the unique tag sequences (Ai) is present in the first oligonucleotide probe, and the second and third unique tag portions (Bi' and Ci') are in the second oligonucleotide probe sequence as shown in Figure 3, step D & E. Upon ligation of oligonucleotide probes in a probe set, the resulting ligation product will contain the Ai sequence—target specific sequences—Bi' sequence—Ci' sequence. The essence of the UniTaq approach is that both oligonucleotide probes of a ligation probe set need to be correct in order to get a positive signal, which allows for highly multiplexed nucleic acid detection. For example, and as described herein, this is achieved by requiring hybridization of two parts, *i.e.*, two of the tags, to each other.

[0140] Prior to detecting the ligation product, the sample is treated with UDG to destroy original target amplicons allowing only authentic ligation products to be detected. Following ligation, the ligation products are aliquoted into separate wells, micro-pores or droplets containing one or more tag-specific primer pairs. For the detection step, the ligation product containing Ai (a first primer-specific portion), Bi' (a UniTaq detection portion), and Ci' (a second primer-specific portion) is primed on both strands using a first oligonucleotide primer having the same nucleotide sequence as Ai, and a second oligonucleotide primer that is complementary to Ci' (*i.e.*, Ci). The first oligonucleotide primer also includes a UniTaq detection probe (Bi) that has a detectable label F1 on one end and a quencher molecule (Q) on the other end (F1-Bi-Q-Ai). Optionally positioned proximal to the quencher is a polymerase-blocking unit, *e.g.*, HEG, THF, Sp-18, ZEN, or any other blocker known in the art that is sufficient to stop polymerase extension. In another embodiment, a ZEN quencher group is also positioned about 9 bases from the fluorescent reporter group to assure more complete quenching. PCR amplification results in the formation of double stranded products as shown in Figure 3, step G). In this example, a polymerase-blocking unit prevents a polymerase from copying the 5' portion (Bi) of the first universal primer, such that the bottom strand of product cannot form a hairpin when it becomes single-stranded. Formation of such a hairpin would result in the 3' end of the stem annealing to the amplicon such that polymerase extension of this 3' end would terminate the PCR reaction.

[0141] The double stranded PCR products are denatured, and when the temperature is subsequently decreased, the upper strand of product forms a hairpin having a stem between the 5' portion (Bi) of the first oligonucleotide primer and portion Bi' at the opposite end of the strand (Figure 3, step H). Also, during this step, the second oligonucleotide primer anneals to the 5'-primer specific portion (Ci') of the hairpinned product. Upon extension of the second universal primer in step H, 5' nuclease activity of the polymerase cleaves the detectable label D1 or the

quencher molecule from the 5' end of the amplicon, thereby increasing the distance between the label and the quencher and permitting detection of the label.

[0142] The ligation reaction used in the methods of the present application is well known in the art. Ligases suitable for ligating oligonucleotide probes of a probe set together (optionally following cleavage of a 3' ribose and blocking group on the first oligonucleotide probe, or the 5' flap on the second oligonucleotide probe) include, without limitation *Thermus aquaticus* ligase, *E. coli* ligase, T4 DNA ligase, T4 RNA ligase, *Taq* ligase, 9 N ligase, and *Pyrococcus* ligase, or any other thermostable ligase known in the art. In accordance with the present application, the nuclease-ligation process of the present application can be carried out by employing an oligonucleotide ligation assay (OLA) reaction (*see* Landegren, et al., "A Ligase-Mediated Gene Detection Technique," *Science* 241:1077-80 (1988); Landegren, et al., "DNA Diagnostics -- Molecular Techniques and Automation," *Science* 242:229-37 (1988); and U.S. Patent No. 4,988,617 to Landegren et al., which are hereby incorporated by reference in their entirety), a ligation detection reaction (LDR) that utilizes one set of complementary oligonucleotide probes (*see e.g.*, WO 90/17239 to Barany et al., which is hereby incorporated by reference in its entirety), or a ligation chain reaction (LCR) that utilizes two sets of complementary oligonucleotide probes *see e.g.*, WO 90/17239 to Barany et al., which is hereby incorporated by reference in its entirety).

[0143] The oligonucleotide probes of a probe sets can be in the form of ribonucleotides, deoxynucleotides, modified ribonucleotides, modified deoxyribonucleotides, peptide nucleotide analogues, modified peptide nucleotide analogues, modified phosphate-sugar-backbone oligonucleotides, nucleotide analogs, and mixtures thereof.

[0144] The hybridization step in the ligase detection reaction, which is preferably a thermal hybridization treatment, discriminates between nucleotide sequences based on a distinguishing nucleotide at the ligation junctions. The difference between the target nucleotide sequences can be, for example, a single nucleic acid base difference, a nucleic acid deletion, a nucleic acid insertion, or rearrangement. Such sequence differences involving more than one base can also be detected. Preferably, the oligonucleotide probe sets have substantially the same length so that they hybridize to target nucleotide sequences at substantially similar hybridization conditions.

[0145] Ligase discrimination can be further enhanced by employing various probe design features. For example, an intentional mismatch or nucleotide analogue (*e.g.*, Inosine, Nitroindole, or Nitropyrrole) can be incorporated into the first oligonucleotide probe at the 2nd or 3rd base from the 3' junction end to slightly destabilize hybridization of the 3' end if it is

perfectly matched at the 3' end, but significantly destabilize hybridization of the 3' end if it is mis-matched at the 3' end. This design reduces inappropriate misligations when mutant probes hybridize to wild-type target. Alternatively, RNA bases that are cleaved by RNases can be incorporated into the oligonucleotide probes to ensure template-dependent product formation. For example, Dobosy et al., "RNase H-Dependent PCR (rhPCR): Improved Specificity and Single Nucleotide Polymorphism Detection Using Blocked Cleavable Primers," *BMC Biotechnology* 11(80):1011 (2011), which is hereby incorporated by reference in its entirety, describes using an RNA-base close to the 3' end of an oligonucleotide probe with 3'-blocked end, and cutting it with RNase H2 generating a PCR-extendable and ligatable 3'-OH. This approach can be used to generate either ligation-competent 3'OH (for standard DNA ligases), or 5'-P, or both, in the latter case, provided a ligase that can ligate 5'-RNA base is utilized.

[0146] Other possible modifications included abasic sites, e.g., internal abasic furan or oxo-G. These abnormal "bases" are removed by specific enzymes to generate ligation-competent 3'-OH or 5'P sites. Endonuclease IV, Tth EndoIV (NEB) will remove abasic residues after the ligation oligonucleotides anneal to the target nucleic acid, but not from a single-stranded DNA. Similarly, one can use oxo-G with Fpg or inosine/uracil with EndoV or Thymine glycol with EndoVIII.

[0147] Ligation discrimination can also be enhanced by using the coupled nuclease-ligase reaction described in WO2013/123220 to Barany et al. or U.S. Patent Application Publication No. 2006/0234252 to Anderson et al., which are hereby incorporated by reference in their entirety. In this embodiment, the first oligonucleotide probe bears a ligation competent 3' OH group while the second oligonucleotide probe bears a ligation incompetent 5' end (*i.e.*, an oligonucleotide probe without a 5' phosphate). The oligonucleotide probes of a probe set are designed such that the 3'-most base of the first oligonucleotide probe is overlapped by the immediately flanking 5'-most base of the second oligonucleotide probe that is complementary to the target nucleic acid molecule. The overlapping nucleotide is referred to as a "flap". When the overlapping flap nucleotide of the second oligonucleotide probe is complementary to the target nucleic acid molecule sequence and the same sequence as the terminating 3' nucleotide of the first oligonucleotide probe, the phosphodiester bond immediately upstream of the flap nucleotide of the second oligonucleotide probe is discriminatingly cleaved by an enzyme having flap endonuclease (FEN) or 5' nuclease activity. That specific FEN activity produces a novel ligation competent 5' phosphate end on the second oligonucleotide probe that is precisely positioned alongside the adjacent 3' OH of the first oligonucleotide probe to allow ligation of the two probes to occur. In accordance with this embodiment, flap endonucleases or 5' nucleases that

are suitable for cleaving the 5' flap of the second oligonucleotide probe prior to ligation include, without limitation, polymerases with 5' nuclease activity such as *E.coli* DNA polymerase and polymerases from *Taq* and *T. thermophilus*, as well as T4 RNase H. In another embodiment, the second probe of the probe set has a 3' primer-specific portion, a target specific portion, and a 5' nucleotide sequence, where the 5' nucleotide sequence is complementary to at least a portion of the 3' primer-specific portion, and where the 5' nucleotide sequence hybridizes to its complementary portion of the 3' primer-specific portion to form a hair-pinned second oligonucleotide probe when the second probe is not hybridized to a target nucleotide sequence.

[0148] Alternatively, as shown in Figure 4, the regions of interest are selectively extended using locus-specific upstream primers, an optional blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated (or its complement), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', e.g. C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated target base, and suitable for polymerase extension (Figure 4, step B). An optional blocking LNA or PNA probe comprising the TET2-APOBEC converted unmethylated sequence (or its complement) that partially overlaps with the upstream PCR primer will preferentially compete for binding to the TET2-APOBEC converted unmethylated sequence over the upstream primer, thus suppressing amplification of TET2-APOBEC converted unmethylated sequence DNA during each round of extension. Add UDG, which destroys the TET2-APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the locus-specific downstream primers are added, followed by limited (8 to 20 cycles) or full (20-40 cycles) PCR using a deoxynucleotide mix that includes dUTP (Figure 4, step C). The downstream primers contain identical universal primer tails to prevent primer dimers. Further, such tails provide the option for including Universal primer during the PCR step. This may assist in generating more equal amounts of products in a multiplexed PCR reaction.

[0149] For Figure 4, methylation-specific upstream and locus-specific downstream probes containing tails (Ai, Ci') enable formation of a ligation product in the presence of TET2-APOBEC converted methylated (or hydroxymethylated) base-containing PCR products. Following ligation, the ligation products can be detected using pairs of matched primers Ai and

Ci, and TaqMan™ probes that span the ligation junction as described *supra* for Figure 2 (see Figures 2, steps E-H), or using other suitable means known in the art.

[0150] Alternatively, methylation-specific upstream and locus-specific downstream probes containing tails (Ai, Bi'-Ci') enable formation of a ligation product in the presence of TET2 and APOBEC converted methylated (or hydroxymethylated) base-containing PCR products. Following ligation, the ligation products are amplified using UniTaq-specific primers (*i.e.*, F1-Bi-Q-Ai, Ci) and detected as described *supra* for Figure 3, or using other suitable means known in the art.

[0151] Another aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample. The method further involves providing one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, and providing one or more first primary oligonucleotide primer(s) that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more methylated or hydroxymethylated residue. The treated sample, the one or more first primary oligonucleotide primers, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. One or more secondary oligonucleotide primer sets are provided. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that is complementary to a portion of the

polymerase extension product formed from the first primary oligonucleotide primer and (b) a second secondary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first secondary oligonucleotide primer. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures, and conditions suitable for carrying out two or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising a 5' primer-specific portion of the first secondary oligonucleotide primer, a DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and a complement of the 5' primer-specific portion of the second secondary oligonucleotide primer. The method further comprises providing one or more tertiary oligonucleotide primer sets. Each tertiary oligonucleotide primer set comprises (a) a first tertiary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first polymerase chain reaction products and (b) a second tertiary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the first polymerase chain reactions product sequence. The first polymerase chain reaction products, the one or more tertiary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU) containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further involves detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in

other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0152] Figures 5, 6, 7, 13 and 14 illustrate various embodiments of this aspect of the present application.

[0153] Figure 5 illustrates an exemplary exPCR-qPCR carryover prevention reaction to detect low-level methylations. Genomic or cfDNA is isolated and is optionally treated with a DNA repair kit (Figure 5, Step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3' OH group which is suitable for polymerase extension (Figure 5, step B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step.

[0154] As shown in Figure 5, step C, following the initial extension reaction, the extension products are aliquoted into separate wells, micro-pores or droplets containing one or more methylation-specific primers comprising 5' primer-specific portions (Ai) (at low concentrations), locus-specific oligonucleotide primers comprising 5' primer-specific portions (Ci) (at low concentrations), as well as matching tag-specific primers Ai and Ci, and methylation-specific Taqman probes (at higher concentrations). These primers combine to amplify the methylation-containing sequence, if present in the sample (Figure 5, step C). In this embodiment, the upstream methylation-specific primer having a sequence specific for detecting the methylation of interest further contains a 5' primer-specific portion (Ai) to facilitate subsequent detection of the nested PCR product. Once again, the presence of blocking LNA or PNA probe comprising TET2 and APOBEC converted unmethylated sequence suppresses extension of TET2 and APOBEC converted unmethylated target sequence if present after the enrichment of methylated sequence during the initial extension step. The reverse locus-specific primer, having a sequence common to both TET2 and APOBEC converted methylated and TET2

and APOBEC converted unmethylated sequences contains a 5' primer-specific portion (Ci) that, together with the 5' primer specific portion (Ai) of the upstream primer having a sequence specific for detecting the methylation region, permit subsequent amplification and detection of only converted methylated PCR products. As illustrated in step C of Figure 5, another layer of specificity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the mutation-specific and locus-specific primers. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to generate a polymerase extension competent 3'OH group (Figure 5, step C). In the initial primer extension (step B) the liberated 3'OH base is a few bases upstream from the methylation position, and thus would extend both TET2 and APOBEC converted unmethylated and methylated sequences if cleaved (although the blocking LNA or PNA should limit cleavage of primer hybridized to TET2 and APOBEC converted unmethylated sequence). In contrast, in the nested PCR (step C), the methylation-specific base of the primer is at the 3'OH base, such that extension on TET2 and APOBEC converted unmethylated sequence would be less likely, since the base is mismatched. The specificity for polymerase extension of TET2 and APOBEC converted methylated over TET2 and APOBEC converted unmethylated sequence may be further improved by: (i) using methylation converted-specific PCR Primers containing a mismatch in the 2nd or 3rd position from the 3'OH base, (ii) using LNA or PNA probes to TET2 and APOBEC converted unmethylated sequence that would reduce hybridization of mutation-specific PCR primers to TET2 and APOBEC converted unmethylated sequences, (iii) using PCR primers to TET2 and APOBEC converted unmethylated sequence that are blocked and do not undergo additional amplification, and (iv) avoiding G:T or T:G mismatches between primer and TET2 and APOBEC converted unmethylated sequence at the 3'OH base. Further, the longer target-specific primers are at a significantly lower concentration than the Taqman probe and tag-specific primers (Ai, Ci), such that the longer mutation-specific primers are depleted, allowing the Taqman probe and tag-specific primers to hybridize and enable target-dependent detection.

[0155] As shown in Figure 5, step D, nested PCR products comprise a 5' primer-specific portion (Ai) target-specific sequence, and a 3' primer-specific portion (Ci') that permits subsequent amplification of the nested PCR product. As shown in Figure 5, steps E and F, detection of the nested PCR products can be carried out using traditional TaqMan™ detection assay, since the tag-specific primer pairs, each pair comprising matched primers Ai and Ci, and probes, are all present in the wells, micro-pores, or droplets (*see* U.S. Patent No. 6,270,967 to Whitcombe et al., and U.S. Patent No. 7,601,821 to Anderson et al., which are hereby incorporated by reference in their entirety). For detection using TaqMan™ an oligonucleotide

probe spanning the mutation-specific region is used in conjunction with primers suitable for hybridization on the primer-specific portions of the nested PCR products for amplification and detection. The TaqMan™ probe contains a fluorescent reporter group on one end (F1) and a quencher molecule (Q) on the other end that are in close enough proximity to each other in the intact probe that the quencher molecule quenches fluorescence of the reporter group. During amplification, the TaqMan™ probe and upstream primer hybridize to their complementary regions of the nested PCR product. The 5' → 3' nuclease activity of the polymerase extends the hybridized primer and liberates the fluorescent group of the TaqMan™ probe to generate a detectable signal (Figure 5, step F). In a preferred embodiment, the Taqman probe contains a second quencher group (ZEN) about 9 bases in from the fluorescent reporter group, and the probe is designed such that the ZEN group is at or adjacent to the mutant base. Use of dUTP during the amplification reaction generates products containing dU, which can subsequently be destroyed using UDG for carryover prevention.

[0156] Alternatively, as shown in Figure 6, step C, nested PCR products comprise a 5' primer-specific portion (Ai) target-specific sequence, and a 3' primer-specific portion (Bi'-Ci') that permits subsequent amplification of the nested PCR product. Following the limited cycle PCR, detection of the nested PCR products can be carried out using the UniTaq method, since the one or more tag-specific primer pairs, each pair comprising matched primers F1-Bi-Q-Ai and Ci, are all present in the wells, micro-pores, or droplets. PCR amplification results in the formation of double stranded products as shown in Figure 6, step D. In this example, a polymerase-blocking unit prevents a polymerase from copying the 5' portion (Bi) of the first universal primer, such that the bottom strand of product cannot form a hairpin when it becomes single-stranded. Formation of such a hairpin would result in the 3' end of the stem annealing to the amplicon such that polymerase extension of this 3' end would terminate the PCR reaction.

[0157] The double stranded PCR products are denatured, and when the temperature is subsequently decreased, the upper strand of product forms a hairpin having a stem between the 5' portion (Bi) of the first oligonucleotide primer and portion Bi' at the opposite end of the strand (Figure 6, step F). Also, during this step, the second oligonucleotide primer anneals to the 5'-primer specific portion (Ci') of the hairpinned product. Upon extension of the second universal primer in step F, 5' nuclease activity of the polymerase cleaves the detectable label D1 or the quencher molecule from the 5' end of the amplicon, thereby increasing the distance between the label and the quencher and permitting detection of the label. Use of dUTP during the amplification reaction generates products containing dU, which can subsequently be destroyed using UDG for carryover prevention

[0158] Alternatively, as shown in Figure 7, regions of interest are selectively extended using locus-specific upstream primers, an optional blocking LNA or PNA probe comprising TET2 and APOBEC converted unmethylated sequence (or its complement), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET1 and APOBEC converted methylated (or hydroxymethylated) target base, and suitable for polymerase extension (Figure 7, step B). An optional blocking LNA or PNA probe comprising the TET2 and APOBEC converted unmethylated sequence (or its complement) that partially overlaps with the upstream PCR primer will preferentially compete for binding to the TET2 and APOBEC converted unmethylated sequence over the upstream primer, thus suppressing amplification of TET2 and APOBEC converted unmethylated sequence DNA during each round of PCR. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step.

[0159] As shown in Figures 5 and 7, step C, TET2 and APOBEC converted methylation base-specific primers (comprising 5' primer-specific portions Ai) and TET2 and APOBEC converted locus-specific primers (comprising 5' primer-specific portions Ci) are added to then perform limited cycle nested PCR to amplify the TET2 and APOBEC converted methylation-containing sequence, if present in the sample. Optionally, blocking LNA or PNA probes comprising the wild-type sequence (or its complement) enables amplification of originally methylated (or hydroxymethylated) but not originally unmethylated alleles. Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products can be detected using pairs of matched primers Ai and Ci, and TaqMan™ probes that span the TET2 and APOBEC -converted methylation target regions as described *supra* for Figure 2 (see Figures 5 and 7, steps D-F), or using other suitable means known in the art.

[0160] Another aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample

are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample. One or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules present in the sample, and one or more primary oligonucleotide primer sets are provided. Each primary oligonucleotide primer set comprises (a) a first primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer, wherein the first or second primary oligonucleotide primer further comprises a 5' primer-specific portion. The treated sample, the one or more first primary oligonucleotide primers of the primer sets, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules in the reaction mixture, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof. One or more secondary oligonucleotide primer sets are then provided. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer having a 3' portion that is

complementary to a portion of a first polymerase chain reaction product formed from the first primary oligonucleotide primer and (b) a second secondary oligonucleotide primer having a 3' portion that comprises a nucleotide sequence that is complementary to a portion of a first polymerase chain reaction product formed from the first secondary oligonucleotide primer. The first polymerase chain reaction products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out two or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further comprises detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0161] Figures 8, 9, 10, 15 and 16 illustrate various embodiments of this aspect of the present application.

[0162] Figure 8 illustrates another exemplary exPCR-qPCR carryover prevention reaction to detect low-level methylation. Genomic or cfDNA is isolated, and optionally treated with a DNA repair kit (Figure 8, Step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', e.g. C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is suitable for polymerase extension (Figure 8, step B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2-APOBEC converted unmethylated sequence. The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Add UDG, which destroys the TET2 and

APOBEC converted DNA (but not the primer extension products). Subsequently, the regions of interest are selectively amplified in a limited cycle PCR (8-20 cycles) using locus-specific upstream primers, an optional blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated sequence (or its complement), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated target region, and suitable for polymerase extension (Figure 8, step C). An optional blocking LNA or PNA probe comprising the TET2-APOBEC converted unmethylated sequence (or its complement) that partially overlaps with the upstream PCR primer will preferentially compete for binding to the TET2-APOBEC converted unmethylated sequence over the upstream primer, thus suppressing amplification of TET2-APOBEC converted unmethylated sequence DNA during each round of PCR.

[0163] Following the limited cycle PCR, the PCR products are aliquoted into separate wells, micro-pores or droplets containing Taqman™ probes, TET2 and APOBEC converted, methylation base-specific, and TET2 and APOBEC converted locus-specific primers, to amplify the TET2 and APOBEC converted methylation-containing sequence, if present in the sample (Figure 8, step D). The TET2 and APOBEC converted methylation-containing products are amplified and detected using TET2 and APOBEC converted methylation base-specific primers, TET2 and APOBEC converted methylation locus-specific primers, and TET2 and APOBEC converted methylation base-specific Taqman™ probes (see Figure 8, steps D-E), or using other suitable means known in the art.

[0164] Figures 9 and 10 illustrate additional exemplary exPCR-qPCR carryover prevention reaction to detect low-level methylation. Genomic or cfDNA is isolated, and optionally treated with a DNA repair kit (Figures 9 and 10, Step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and a deoxynucleotide mix that does NOT include dUTP. . In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is suitable for polymerase extension (Figure 9, step

B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2-APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the regions of interest are selectively amplified in a limited cycle PCR (8-20 cycles) using locus-specific upstream primers, an optional blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated sequence (or its complement), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2 and APOBEC converted methylated (or hydroxymethylated) target base, and suitable for polymerase extension (Figure 9, step C). An optional blocking LNA or PNA probe comprising the TET2-APOBEC converted unmethylated sequence (or its complement) that partially overlaps with the upstream PCR primer will preferentially compete for binding to the TET2-APOBEC converted unmethylated sequence over the upstream primer, thus suppressing amplification of TET2-APOBEC converted unmethylated sequence DNA during each round of PCR.

[0165] Alternatively, as shown in Figure 10, the regions of interest are selectively extended using locus-specific upstream primers, an optional blocking LNA or PNA probe comprising TET2-APOBEC converted unmethylated sequence (or its complement), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated (or hydroxymethylated) target base, and suitable for polymerase extension (Figure 10, step B). An optional blocking LNA or PNA probe comprising the TET2-APOBEC converted unmethylated sequence (or its complement) that partially overlaps with the upstream PCR primer will preferentially compete for binding to the TET2-APOBEC converted unmethylated sequence over the upstream primer, thus suppressing amplification of TET2-APOBEC converted unmethylated sequence DNA during each round of PCR. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the locus-specific downstream primers

comprising a 5' primer-specific portion and a 3' target-specific portion, are added, followed by limited cycle PCR (8 to 12 cycles, Figure 10, step C). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2-APOBEC converted unmethylated sequence.

[0166] For the protocol illustrated in Figures 9 and 10, following the limited cycle PCR, the PCR products are aliquoted into separate wells, micro-pores, or droplets containing Taqman™ probes, TET2-APOBEC converted methylation base-specific primers comprising 5' primer-specific portions (Ai), TET2-APOBEC converted locus-specific primers comprising 5' primer-specific portions (Ci) and matching primers Ai and Ci. These primers combine to amplify the TET2-APOBEC converted methylated or hydroxymethylated-containing sequence, if present in the sample (Figures 9 and 10, step D). Optional blocking LNA or PNA probes comprising the TET2-APOBEC converted unmethylated sequence (or its complement) enables amplification of originally methylated or hydroxymethylated but not originally un-methylated allele. Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products can be detected using pairs of matched primers Ai and Ci, and TaqMan™ probes that span the TET2-APOBEC converted methylation target regions as described *supra* for Figure 5 (see Figure 9, steps E-G), or using other suitable means known in the art.

[0167] Alternatively, following the limited cycle PCR, the PCR products are aliquoted into separate wells, micro-pores or droplets containing Taqman™ probes, TET2 and APOBEC converted methylation base-specific primers comprising 5' primer-specific portions (Ai), TET2 and APOBEC converted locus-specific primers comprising 5' primer-specific portions (Bi-Ci) and matching UniTaq primers F1-Bi-Q-Ai and Ci. Optional blocking LNA or PNA probes comprising the wild-type sequence (or its complement) enables amplification of originally methylated but not originally un-methylated allele. Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products are amplified using UniTaq-specific primers (*i.e.*, F1-Bi-Q-Ai, Ci) and detected as described *supra* for Figure 6, or using other suitable means known in the art.

[0168] Figures 11 and 12 illustrate additional exemplary exPCR-LDR-qPCR carryover prevention reactions to detect low-level methylation. Genomic or cfDNA is isolated and then either treated with: (i) methyl-sensitive restriction endonucleases, *e.g.*, Bsh1236I (CG[^]CG), to completely digest unmethylated DNA and prevent carryover, or (ii) capture and enrich for methylated DNA, (iii) followed by a DNA repair kit (Figures 11 and 12, step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to

convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is suitable for polymerase extension (Figure 11, step B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the regions of interest are selectively amplified in a limited cycle PCR (8-20 cycles) or full cycle PCR (20-40 cycles) using locus-specific upstream primers comprising 5' universal 10-15 base tail sequences and 3' target-specific sequences. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated (or hydroxymethylated) target base, and suitable for polymerase extension (Figure 11, step C). If the locus-specific upstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. The downstream primers contain identical universal primer tails to prevent primer dimers. Further, such tails provide the option for including Universal primer during the PCR step. This may assist in generating more equal amounts of products in a multiplexed PCR reaction. The amplified products contain dU as shown in Figure 11, step D, which allows for subsequent treatment with UDG or a similar enzyme for carryover prevention.

[0169] Alternatively, as shown in Figure 12, the regions of interest are selectively extended using locus-specific upstream primers comprising 5' universal 10-15 base tail sequences and 3' target-specific sequences for TET2-APOBEC converted DNA. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated target

base, and suitable for polymerase extension (Figure 12, step B). If the locus-specific upstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the locus-specific downstream primers comprising 5' universal primer sequences and 3' target-specific sequences are added, followed by limited (8 to 20 cycles) or full (20-40 cycles) PCR using a deoxynucleotide mix that includes dUTP (Figure 12, step C). In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', e.g. C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is suitable for polymerase extension (Figure 12, step C). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. The downstream primers contain identical universal primer tails to prevent primer dimers. Further, such tails provide the option for including Universal primer during the PCR step. This may assist in generating more equal amounts of products in a multiplexed PCR reaction. The amplified products contain dU as shown in Figure 12, step D, which allows for subsequent treatment with UDG or a similar enzyme for carryover prevention.

[0170] For Figures 11 and 12, methylation-specific upstream and locus-specific downstream probes containing tails (Ai, Ci') enable formation of a ligation product in the presence of TET2 and APOBEC converted methylated or hydroxymethylated base-containing PCR products. Following ligation, the ligation products can be detected using pairs of matched primers Ai and Ci, and TaqMan™ probes that span the ligation junction as described *supra* for Figure 2 (see Figure 11, steps E-H), or using other suitable means known in the art.

[0171] Alternatively, methylation-specific upstream and locus-specific downstream probes containing tails (Ai, Bi'-Ci') enable formation of a ligation product in the presence of TET2 and APOBEC converted methylated base-containing PCR products. Following ligation, the ligation products are amplified using UniTaq-specific primers (*i.e.*, F1-Bi-Q-Ai, Ci) and detected as described *supra* for Figure 3, or using other suitable means known in the art.

[0172] Figures 13 and 14 illustrate additional exemplary exPCR-LDR-qPCR carryover prevention reactions to detect low-level methylation. Genomic or cfDNA is isolated, and optionally treated with: (i) methyl-sensitive restriction endonucleases, *e.g.*, Bsh1236I (CG^CG),

to completely digest unmethylated DNA and prevent carryover, or (ii) capture and enrich for methylated DNA, (iii) followed by a DNA repair kit (Figures 13 and 14, step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is suitable for polymerase extension (Figure 13, step B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step.

[0173] Alternatively, as shown in Figure 14, the regions of interest are selectively extended using locus-specific upstream primers comprising 5' universal 10-15 base tail sequences and 3' target-specific sequences for TET2-APOBEC converted DNA, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated (or unmethylated) target base, and suitable for polymerase extension (Figure 14, step B). Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step.

[0174] As shown in Figures 13 and 14, step C, TET2 and APOBEC converted methylation base-specific primers (comprising 5' primer-specific portions Ai) and TET2-APOBEC converted locus-specific primers (comprising 5' primer-specific portions Ci) are added to then perform limited cycle nested PCR to amplify the TET2-APOBEC converted methylation-containing sequence, if present in the sample. Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products can be detected using pairs of matched primers Ai and Ci, and TaqMan™ probes that span the TET2 and APOBEC converted

methylation target regions as described *supra* for Figure 5 (see Figures 13 and 14, steps D-F), or using other suitable means known in the art.

[0175] Alternatively, TET2 and APOBEC converted methylation base-specific primers (comprising 5' primer-specific portions Ai) and TET2-APOBEC converted locus-specific primers (comprising 5' primer-specific portions Bi-Ci) are added to then perform limited cycle nested PCR to amplify the TET2-APOBEC converted methylation-containing sequence, if present in the sample. Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products are amplified using UniTaq-specific primers (*i.e.*, F1-Bi-Q-Ai, Ci) and detected as described *supra* for Figure 6, or using other suitable means known in the art.

[0176] Figure 15 illustrates another exemplary exPCR-qPCR carryover prevention reaction to detect low-level methylation. Genomic or cfDNA is isolated and optionally treated with: (i) methyl-sensitive restriction endonucleases, *e.g.*, Bsh1236I (CG⁺CG), to completely digest unmethylated DNA and prevent carryover, or (ii) capture and enrich for methylated DNA, (iii) followed by a DNA repair kit (Figure 15, step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal 10-15 base tail sequences and 3' target-specific sequences for TET2-APOBEC converted DNA, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group, which is 15 bases or more upstream of the TET2-APOBEC converted methylated or hydroxymethylated target base and suitable for polymerase extension (Figure 15, step B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the regions of interest are selectively amplified in a limited cycle PCR (8-20 cycles) using locus-specific upstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific

hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated target base, and suitable for polymerase extension (Figures 15, step C). If the locus-specific upstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence.

[0177] Following the limited cycle PCR, the PCR products are aliquoted into separate wells, micro-pores or droplets containing Taqman™ probes, TET2-APOBEC converted, methylation base-specific, and TET2-APOBEC converted locus-specific primers, to amplify the TET2-APOBEC converted methylation-containing sequence, if present in the sample (Figure 15, step D). The TET2-APOBEC converted methylation-containing products are amplified and detected as described *supra* for Figure 8 (see Figure 15, steps D-E), or using other suitable means known in the art.

[0178] Figure 16 illustrates an additional exemplary exPCR-qPCR carryover prevention reaction to detect low-level methylation. Genomic or cfDNA is isolated, and optionally treated with: (i) methyl-sensitive restriction endonucleases, *e.g.*, Bsh1236I (CG⁺CG), to completely digest unmethylated DNA and prevent carryover, or (ii) capture and enrich for methylated DNA, (iii) followed by a DNA repair kit (Figure 16, step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using locus-specific downstream primers comprising 5' universal 10-15 base tail sequences and 3' target-specific sequences for TET2-APOBEC converted DNA, and a deoxynucleotide mix that does NOT include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group, which is 15 bases or more upstream of the TET2-APOBEC converted methylated or hydroxymethylated target base, and suitable for polymerase extension (Figure 16, step B). If the locus-specific downstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the regions of interest are selectively amplified in a limited cycle PCR (8-20 cycles) using locus-specific upstream primers comprising 5' universal primer sequences and 3' target-specific sequences, and

a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2 and APOBEC converted methylated target base, and suitable for polymerase extension (Figure 16, step C). If the locus-specific upstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the wild-type unmethylated sequence.

[0179] For the protocol illustrated in Figure 16, following the limited cycle PCR, the PCR products are aliquoted into separate wells, micro-pores, or droplets containing TaqmanTM probes, TET2-APOBEC converted methylation base-specific primers comprising 5' primer-specific portions (Ai), TET2-APOBEC converted locus-specific primers comprising 5' primer-specific portions (Ci) and matching primers Ai and Ci. These primers combine to amplify the TET2-APOBEC converted methylation-containing sequence, if present in the sample (Figure 16, step D). Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products can be detected using pairs of matched primers Ai and Ci, and TaqManTM probes that span the TET2-APOBEC converted methylation target regions as described *supra* for Figure 5 (see Figure 16, steps E-G), or using other suitable means known in the art.

[0180] Alternatively, following the limited cycle PCR, the PCR products are aliquoted into separate wells, micro-pores or droplets containing TaqmanTM probes, TET2-APOBEC converted methylation base-specific primers comprising 5' primer-specific portions (Ai), TET2-APOBEC converted locus-specific primers comprising 5' primer-specific portions (Bi-Ci) and matching UniTaq primers F1-Bi-Q-Ai and Ci. Primers are unblocked with RNaseH2 only when bound to correct target. Following PCR, the products are amplified using UniTaq-specific primers (*i.e.*, F1-Bi-Q-Ai, Ci) and detected as described *supra* for Figure 6, or using other suitable means known in the art.

[0181] Another aspect of the present application is directed to a method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues. The method involves providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues. The nucleic acid molecules in the sample are subjected to a treatment with one or more DNA repair enzymes

under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample. One or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules present in the sample are provided, and one or more primary oligonucleotide primer sets are provided. Each primary oligonucleotide primer set comprises (a) a first primary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer. The treated sample, the one or more first primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more polymerase extension reaction mixtures. The one or more polymerase extension reaction mixtures are subjected to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence. The one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules in the reaction mixture, a deoxynucleotide mix, and a DNA polymerase are blended to form one or more first polymerase chain reaction mixtures. The one or more first polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reactions products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof. One or more secondary oligonucleotide primer sets are then provided. Each secondary oligonucleotide primer set comprises (a) a first secondary oligonucleotide primer comprising the same nucleotide sequence

as the 5' primer-specific portion of the first polymerase chain reaction products or their complements and (b) a second secondary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the first polymerase chain reaction products or their complements. The first polymerase chain reaction products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase are blended to form one or more second polymerase chain reaction mixtures. The one or more second polymerase chain reaction mixtures are subjected to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products. The method further involves detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

[0182] Figure 17 illustrates an embodiment of this aspect of the present application.

[0183] Figure 17 illustrates an additional exemplary exPCR-qPCR carryover prevention reaction to detect low-level methylation. Genomic or cfDNA is isolated, and optionally treated with: (i) methyl-sensitive restriction endonucleases, *e.g.*, Bsh1236I (CG⁺CG), to completely digest unmethylated DNA and prevent carryover, or (ii) capture and enrich for methylated DNA, (iii) followed by a DNA repair kit (Figure 17, step A). The DNA is treated with TET2, for conversion of 5mC and 5hmC to 5caC, and then treated with APOBEC to convert unmethylated-C, but not 5caC (previously 5mC or 5hmC) to dU. The regions of interest are selectively extended using TET2-APOBEC converted locus-specific downstream primers comprising 5' primer-specific portions (Ci for Figure 17), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the downstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is suitable for polymerase extension (Figure 17, step B). In this embodiment, the locus-specific downstream primer covers one or more methylation sites, and another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence. Add UDG, which

destroys the TET2 and APOBEC converted DNA (but not the primer extension products). The sample is optionally aliquoted into 12, 24, 36, 48, or 96 wells prior to the initial extension step. Subsequently, the regions of interest are selectively amplified in a limited cycle PCR (8-20 cycles) using TET2-APOBEC converted methylation base-specific upstream primers comprising 5' primer-specific portions (Ai), and a deoxynucleotide mix that does not include dUTP. In this embodiment, another layer of selectivity can be incorporated into the method by including a 3' cleavable blocking group (Blk 3', *e.g.* C3 spacer), and an RNA base (r), in the upstream primer. Upon target-specific hybridization, RNase H (star symbol) removes the RNA base to liberate a 3'OH group which is a few bases upstream of the TET2-APOBEC converted methylated (or hydroxymethylated) target base, and suitable for polymerase extension (Figure 17, step C). Since the methylation base-specific upstream primer covers one or more methylation sites, another layer of specificity may be added by using blocking primers whose sequence corresponds to the TET2 and APOBEC converted unmethylated sequence.

[0184] As shown in Figure 17 step D, the limited cycle PCR products comprise of Ai tag sequence, methylation-specific sequence, and Ci' tag sequence, and are distributed into wells, micro-pores, or droplets for Taqman™ reactions. Following PCR, the products can be detected using pairs of matched primers Ai and Ci, and TaqMan™ probes that span the TET2 and APOBEC converted methylation target regions as described *supra* for Figure 5 (see Figure 17, steps D-F), or using other suitable means known in the art.

[0185] Alternatively, the limited cycle PCR products comprise of Ai tag sequence, methylation-specific sequence, and Bi'-Ci' tag sequence, and are distributed into wells, micro-pores, or droplets for Taqman™ reactions. Following PCR, the products are amplified using UniTaq-specific primers (*i.e.*, F1-Bi-Q-Ai, Ci) and detected as described *supra* for Figure 6, or using other suitable means known in the art.

[0186] The methods described *supra* may further comprise contacting the sample with at least a first methylation sensitive enzyme to form one or more restriction enzyme reaction mixtures prior to, or concurrent with, said blending to form one or more polymerase extension reaction mixtures. The first methylation sensitive enzyme cleaves nucleic acid molecules in the sample that contain one or more unmethylated residues within at least one methylation sensitive enzyme recognition sequence, and the detecting step involves detection of one or more parent nucleic acid molecules containing the target nucleotide sequence, wherein the parent nucleic acid molecules originally contained one or more methylated or hydroxymethylated residues.

[0187] In accordance with this and all aspects of the present invention, a "methylation sensitive enzyme" is an endonuclease that will not cleave or has reduced cleavage efficiency of

its cognate recognition sequence in a nucleic acid molecule when the recognition sequence contains a methylated residue (*i.e.*, it is sensitive to the presence of a methylated residue within its recognition sequence). A “methylation sensitive enzyme recognition sequence” is the cognate recognition sequence for a methylation sensitive enzyme. In some embodiments, the methylated residue is a 5-methyl-C, within the sequence CpG (*i.e.*, 5-methyl-CpG). A non-limiting list of methylation sensitive restriction endonuclease enzymes that are suitable for use in the methods of the present invention include, without limitation, AciI, HinP1I, Hpy99I, HpyCH4IV, BstUI, HpaII, HhaI, or any combination thereof.

[0188] In certain embodiments, the sample is contacted with an immobilized methylated or hydroxymethylated nucleic acid binding protein or antibody to selectively bind and enrich for methylated or hydroxymethylated nucleic acid in the sample.

[0189] The one or more primary or secondary oligonucleotide primers may comprise a portion that has no or one nucleotide sequence mismatch when hybridized in a base-specific manner to the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof, but have one or more additional nucleotide sequence mismatches that interferes with polymerase extension when said primary or secondary oligonucleotide primers hybridize in a base-specific manner to a corresponding nucleotide sequence portion in DNA repair enzyme and DNA deaminase enzyme-treated unmethylated nucleic acid sequence or complement sequence thereof.

[0190] In certain embodiments, one or both primary oligonucleotide primers of the primary oligonucleotide primer set and/or one or both secondary oligonucleotide primers of the secondary oligonucleotide primer set may have a 3' portion comprising a cleavable nucleotide or nucleotide analogue and a blocking group, such that the 3' end of said primer or primers is unsuitable for polymerase extension. In accordance with this embodiment, the cleavable nucleotide or nucleotide analog of one or both oligonucleotide primers is cleaved during the hybridization treatment, thereby liberating free 3'OH ends on one or both oligonucleotide primers prior to said extension treatment.

[0191] This embodiment may also comprise one or more primary or secondary oligonucleotide primers comprising a sequence that differs from the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof. The difference is located two or three nucleotide bases from the liberated free 3'OH end.

[0192] The cleavable nucleotide may comprise one or more RNA bases.

[0193] The methods of the present application may also further comprise providing one or more blocking oligonucleotide primers comprising one or more mismatched bases at the 3' end or one or more nucleotide analogs and a blocking group at the 3' end, such that the 3' end of the blocking oligonucleotide primer is unsuitable for polymerase extension when hybridized in a base-specific manner to wild-type nucleic acid sequence or complement sequence thereof. The blocking oligonucleotide primer comprises a portion having a nucleotide sequence that is the same as a nucleotide sequence portion in the wild-type nucleic acid sequence or complement sequence thereof to which the blocking oligonucleotide primer hybridizes but has one or more nucleotide sequence mismatches to a corresponding nucleotide sequence portion in the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof. The one or more blocking oligonucleotide primers are blended with the sample or products subsequently produced from the sample prior to a polymerase extension reaction, polymerase chain reaction, or ligation reaction, whereby during the hybridization step the one or more blocking oligonucleotide primers preferentially hybridize in a base-specific manner to a wild-type nucleic acid sequence or complement sequence thereof, thereby interfering with polymerase extension or ligation during reaction of a primer or probes hybridized in a base-specific manner to the DNA repair enzyme and DNA deaminase enzyme-treated unmethylated sequence or complement sequence thereof.

[0194] In one embodiment, the first secondary oligonucleotide primer has a 5' primer-specific portion and the second secondary oligonucleotide primer has a 5' primer-specific portion. The one or more secondary oligonucleotide primer sets further comprise a third secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first secondary oligonucleotide primer and (d) a fourth secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the second secondary oligonucleotide primer.

[0195] In another embodiment, the method involves providing one or more third primary oligonucleotide primers comprising the same nucleotide sequence as the 5' primer-specific portion of the first or second primary oligonucleotide primer, and blending the one or more third primary oligonucleotide primers in the one or more first polymerase chain reaction mixtures.

[0196] In accordance with the methods described herein, the DNA repair enzyme may be the ten-eleven translocation (TET2) dioxygenase and the DNA deaminase enzyme may be an apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC cytidine deaminase).

[0197] In one embodiment, the second oligonucleotide probe of the oligonucleotide probe set further comprises a unitaq detection portion, thereby forming ligated product sequences comprising the 5' primer-specific portion, the target-specific portions, the unitaq detection portion, and the 3' primer-specific portion. In accordance with this embodiment, the method further involves providing one or more unitaq detection probes, wherein each unitaq detection probe hybridizes to a complementary unitaq detection portion and the detection probe comprises a quencher molecule and a detectable label separated from the quencher molecule. The one or more unitaq detection probes are added to the second polymerase chain reaction mixture, and the one or more unitaq detection probes are hybridized to complementary unitaq detection portions on the ligated product sequence or complement thereof during the subjecting the second polymerase chain reaction mixture to conditions suitable for one or more polymerase chain reaction cycles, wherein the quencher molecule and the detectable label are cleaved from the one or more unitaq detection probes during the extension treatment and the detecting involves the detection of the cleaved detectable label.

[0198] In certain embodiments, one primary oligonucleotide primer or one secondary oligonucleotide primer further comprises a unitaq detection portion, thereby forming extension product sequences comprising the 5' primer-specific portion, the target-specific portions, the unitaq detection portion, and the complement of the other 5' primer-specific portion, and complements thereof. In accordance with this embodiment, the method involves providing one or more unitaq detection probes, wherein each unitaq detection probe hybridizes to a complementary unitaq detection portion and the detection probe comprises a quencher molecule and a detectable label separated from the quencher molecule. The one or more unitaq detection probes are added to the one or more polymerase chain reaction mixtures, and the one or more unitaq detection probes are hybridized to complementary unitaq detection portions on the ligated product sequence or complement thereof during polymerase chain reaction cycles after the first polymerase chain reaction, wherein the quencher molecule and the detectable label are cleaved from the one or more unitaq detection probes during the extension treatment and the detecting involves the detection of the cleaved detectable label.

[0199] In another embodiment, one or both oligonucleotide probes of the oligonucleotide probe set comprises a portion that has no or one nucleotide sequence mismatch when hybridized in a base-specific manner to the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof, but have one or more additional nucleotide sequence mismatches that interferes with ligation when said oligonucleotide probe hybridizes in a base-specific

manner to a corresponding nucleotide sequence portion in the DNA repair enzyme and DNA deaminase enzyme-treated unmethylated nucleic acid sequence or complement sequence thereof.

[0200] In one embodiment, the 3' portion of the first oligonucleotide probe of the oligonucleotide probe set comprises a cleavable nucleotide or nucleotide analogue and a blocking group, such that the 3' end is unsuitable for polymerase extension or ligation. In accordance with this embodiment, the cleavable nucleotide or nucleotide analog of the first oligonucleotide probe is cleaved when the probe is hybridized to its complementary target nucleotide sequence of the primary extension product, thereby liberating a 3'OH on the first oligonucleotide probe prior to the ligating step.

[0201] The one or more first oligonucleotide probe of the oligonucleotide probe set may comprise a sequence that differs from the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof. The difference is located two or three nucleotide bases from the liberated free 3'OH end.

[0202] In another embodiment, the second oligonucleotide probe has, at its 5' end, an overlapping identical nucleotide with the 3' end of the first oligonucleotide probe, and, upon hybridization of the first and second oligonucleotide probes of a probe set at adjacent positions on a complementary target nucleotide sequence of a primary extension product to form a junction, the overlapping identical nucleotide of the second oligonucleotide probe forms a flap at the junction with the first oligonucleotide probe. This further involves cleaving the overlapping identical nucleotide of the second oligonucleotide probe with an enzyme having 5' nuclease activity thereby liberating a phosphate at the 5' end of the second oligonucleotide probe prior to the ligating step.

[0203] In one embodiment, the one or more oligonucleotide probe sets further comprise a third oligonucleotide probe having a target-specific portion, wherein the second and third oligonucleotide probes of a probe set are configured to hybridize adjacent to one another on the target nucleotide sequence with a junction between them to allow ligation between the second and third oligonucleotide probes to form a ligated product sequence comprising the first, second, and third oligonucleotide probes of a probe set.

[0204] For the methods described herein, the sample may be, without limitation, tissue, cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, cell-free circulating nucleic acids, cell-free circulating tumor nucleic acids, cell-free circulating fetal nucleic acids in pregnant woman, circulating tumor cells, tumor, tumor biopsy, and exosomes.

[0205] The one or more target nucleotide sequences may be low-abundance nucleic acid molecules comprising one or more nucleotide base mutations, insertions, deletions, translocations, splice variants, mRNA, lncRNA, ncRNA, miRNA variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicing, exon insertions, exon deletions, intron insertions, or other rearrangement at the genome level and/or methylated or hydroxymethylated nucleotide bases.

[0206] As used herein “low abundance nucleic acid molecule” refers to a target nucleic acid molecule that is present at levels as low as 1% to 0.01% of the sample. In other words, a low abundance nucleic acid molecule with one or more nucleotide base mutations, insertions, deletions, translocations, splice variants, miRNA variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicings, exon insertions, exon deletions, intron insertions, other rearrangement at the genome level, and/or methylated nucleotide bases can be distinguished from a 100 to 10,000-fold excess of nucleic acid molecules in the sample (i.e., high abundance nucleic acid molecules) having a similar nucleotide sequence as the low abundance nucleic acid molecules but without the one or more nucleotide base mutations, insertions, deletions, translocations, splice variants, miRNA variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicings, exon insertions, exon deletions, intron insertions, other rearrangement at the genome level, and/or methylated nucleotide bases.

[0207] In some embodiments of the present invention, the copy number of one or more low abundance target nucleotide sequences are quantified relative to the copy number of high abundance nucleic acid molecules in the sample having a similar nucleotide sequence as the low abundance nucleic acid molecules. In other embodiments of the present invention, the one or more target nucleotide sequences are quantified relative to other nucleotide sequences in the sample. In other embodiments of the present invention, the relative copy number of one or more target nucleotide sequences is quantified. Methods of relative and absolute (i.e., copy number) quantitation are well known in the art.

[0208] The low abundance target nucleic acid molecules to be detected can be present in any biological sample, including, without limitation, tissue, cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, cell-free circulating nucleic acids, cell-free circulating tumor nucleic acids, cell-free circulating fetal nucleic acids in pregnant woman, circulating tumor cells, tumor, tumor biopsy, and exosomes.

[0209] The methods of the present invention are suitable for diagnosing or prognosing a disease state and/or distinguishing a genotype or disease predisposition.

[0210] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 6-12 markers, 12-24 markers, 24-36 markers, 36-48 markers, 48-72 markers, 72-96 markers, or > 96 markers. Each marker in a given set is selected by having any one or more of the following criteria: present, or above a cutoff level, in > 50% of biological samples of the disease cells or tissue from individuals diagnosed with the disease state; absent, or below a cutoff level, in > 95% of biological samples of the normal cells or tissue from individuals without the disease state; present, or above a cutoff level, in > 50% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with the disease state; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without the disease state; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with the disease state; and, wherein at least 50% of the markers in a set each comprise one or more methylated or hydroxymethylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated or hydroxymethylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with the disease state. The method involves obtaining the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, and bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. Nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are

carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with the disease state if a minimum of 2 or 3 markers are present or above a cutoff level in a marker set comprising from 6-12 markers; or a minimum of 3, 4, or 5 markers are present or above a cutoff level in a marker set comprising from 12-24 markers; or a minimum of 3, 4, 5, or 6 markers are present or above a cutoff level in a marker set comprising from 24-36 markers; or a minimum of 4, 5, 6, 7, or 8 markers are present or above a cutoff level in a marker set comprising from 36-48 markers; or a minimum of 6, 7, 8, 9, 10, 11, or 12 markers are present or above a cutoff level in a marker set comprising from 48-72 markers, or a minimum of 7, 8, 9, 10, 11, 12 or 13 markers are present or above a cutoff level in a marker set comprising from 72-96 markers, or a minimum of 8, 9, 10, 11, 12, 13 or "n"/12 markers are present or above a cutoff level in a marker set comprising 96 to "n" markers, when "n" > 168 markers.

[0211] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 48-72 total cancer markers, 72-96 total cancer markers or ≥ 96 total cancer markers, wherein on average greater than one quarter such markers in a given set cover each of the aforementioned major cancers being tested. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in > 50% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in > 50% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily

fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining a biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 48-72 total cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 72-96 total cancer markers; or a minimum of 6 or "n"/18 markers are present or are above a cutoff level in a marker set comprising 96 to "n" total cancer markers, when "n" > 96 total cancer markers.

[0212] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of and identifying the most likely specific tissue(s) of origin of a solid tissue cancer in the following groups: Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 36-48 group-specific cancer markers, 48-64 group-specific cancer markers or ≥ 64 group-specific cancer markers, wherein on average greater than one third such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in $> 50\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in $> 50\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the

cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dioxuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 36-48 group-specific cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 48-64 group-specific cancer markers; or a minimum of 6 or "n"/12 markers are present or are above a cutoff level in a marker set comprising 64 to "n" group-specific cancer markers, when "n" > 64 group-specific cancer markers.

[0213] Another aspect of the present application relates to a method of diagnosing or prognosing a disease state of a gastrointestinal cancer including colorectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 6-12 markers, 12-18 markers, 18-24 markers, 24-36 markers, 36-48 markers or ≥ 48 markers. Each marker is selected by having any one or more of the following criteria for gastrointestinal cancer: present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with gastrointestinal cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without gastrointestinal cancer; present, or above a cutoff level, in > 75% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with gastrointestinal cancer; absent, or below a

cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without gastrointestinal cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with gastrointestinal cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with gastrointestinal cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with gastrointestinal cancer if a minimum of 2, 3 or 4 markers are present or are above a cutoff level in a marker set comprising from 6-12 markers; or a minimum of 2, 3, 4, or 5 markers are present or are above a cutoff level in a marker set comprising from 12-18 markers; or a minimum of 3, 4, 5, or 6 markers are present or are above a cutoff level in a marker set comprising from 18-24 markers; or a minimum of 3, 4, 5, 6, 7, or 8 markers are present or are above a cutoff level in a marker set comprising from 24-36 markers; or a minimum of 4, 5, 6, 7, 8, 9, or 10 markers are present or are above a cutoff level in a marker

set comprising from 36-48 markers; or a minimum of 5, 6, 7, 8, 9, 10, 11, 12, or “n”/12 markers are present or are above a cutoff level in a marker set comprising 48 to “n” markers, when “n” > 48 markers.

[0214] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 36-48 total cancer markers, 48-64 total cancer markers, or ≥ 64 total cancer markers. On average greater than half of such markers in a given set cover each of the aforementioned major cancers being tested. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample

including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dioxuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 36-48 total cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 48-64 total cancer markers; or a minimum of 6 or "n"/12 markers are present or are above a cutoff level in a marker set comprising 64 to "n" total cancer markers, when "n" > 96 total cancer markers.

[0215] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state of and identifying the most likely specific tissue(s) of origin of a solid tissue cancer in the following groups: Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half

of such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and

distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 24-36 group-specific cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 36-48 group-specific cancer markers; or a minimum of 6 or " n "/8 markers are present or are above a cutoff level in a marker set comprising 48 to " n " group-specific cancer markers, when " n " > 48 group-specific cancer markers.

[0216] Another aspect of the present application is directed to a method of diagnosing or prognosing a disease state to guide and monitor treatment of a solid tissue cancer in one or more of the following groups; Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in > 75% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions

thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise of one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals with a given tissue-specific cancer will on average have from approximately one-quarter to about one-half or more of the markers scored as present, or are above a cutoff level in the tested marker set, wherein to guide and monitor subsequent treatment, a portion or all of the identified markers scored as present or the identified markers as above a cutoff level in the tested marker set are deemed the “patient-specific marker set”, and retested on a subsequent biological sample from the individual during the treatment protocol, to monitor for loss of marker signal, wherein if a minimum of 3 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 12-24 markers; or if a minimum of 4 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 24-36 markers; or a minimum of 5 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 36-48 markers; or a minimum of 6 or “n”/8 markers remain present or remain above a cutoff level in a patient-specific marker set

comprising 48 to “n” markers, when “n” > 48 markers after the treatment protocol has been administered, then the continuing presence of said markers may guide a decision to change the cancer treatment therapy.

[0217] Another aspect of the present application is directed to a method of diagnosing or prognosing for a disease state recurrence of a solid tissue cancer in one or more of the following groups; Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group. Each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer: present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in > 75% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; and, wherein at least 50% of the markers in a set each comprise of one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions

thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer. The method involves obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. The sample is fractionated into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein. The nucleic acid molecules in one or more fractions are subjected to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues. At least two enrichment steps are carried out for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step. The method further comprises performing one or more assays to detect and distinguish the plurality of cancer -specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals with a given tissue-specific cancer will on average have from approximately one-quarter to about one-half or more of the markers scored as present, or are above a cutoff level in the tested marker set, wherein to monitor for recurrence, a portion or all of of the markers scored as being present, or the markers scored as above a cutoff level in the tested marker set are deemed the "patient-specific marker set", and retested on subsequent biological samples from the individual after a successful treatment, to monitor for gain of marker signal, wherein if a minimum of 3 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 12-24 markers; or if a minimum of 4 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 24-36 markers; or a minimum of 5 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 36-48 markers; or a minimum of 6 or "n"/8 markers reappear or rise above a cutoff level in a patient-specific marker set comprising 48 to "n" markers, when "n" > 48 markers after the treatment protocol has been administered, then the reappearance or rise or rise above a cutoff level in a patient-specific marker set may guide a decision to resume the cancer treatment therapy or change to a new cancer treatment therapy.

[0218] In certain embodiments, each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in > 66% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer; present, or above a cutoff level, in > 66% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer; absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer; present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer.

[0219] In certain embodiments, the at least two enrichment steps comprise of one or more of the following steps: capturing or separating exosomes or extracellular vesicles or markers in other protected states; capturing or separating a platelet fraction; capturing or separating circulating tumor cells; capturing or separating RNA-containing complexes; capturing or separating cfDNA-nucleosome or differentially modified cfDNA-histone complexes; capturing or separating protein targets or protein target complexes; capturing or separating auto-antibodies; capturing or separating cytokines; capturing or separating methylated or hydroxymethylated cfDNA; capturing or separating marker specific DNA, cDNA, miRNA, lncRNA, ncRNA, or mRNA, or amplified complements, by hybridization to complementary capture probes in solution, on magnetic beads, or on a microarray; amplifying miRNA markers, non-coding RNA markers (lncRNA & ncRNA markers), mRNA markers, exon markers, splice-variant markers, translocation markers, or copy number variation markers in a linear or exponential manner via a polymerase extension reaction, polymerase chain reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific polymerase chain reaction, reverse-transcription reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific ligation reaction, and/or ligation reaction, using DNA polymerase, reverse transcriptase, DNA ligase, RNA ligase, DNA repair enzyme, DNA deaminase enzyme, RNase, RNaseH2, endonuclease, restriction endonuclease, exonuclease, CRISPR, DNA glycosylase or combinations thereof; selectively amplifying one or more target regions containing mutation markers or DNA repair enzyme and DNA deaminase enzyme-treated -converted DNA methylation markers, while suppressing amplification of the target regions containing DNA repair enzyme and DNA deaminase enzyme-treated unmethylated sequence or complement

sequence thereof, in a linear or exponential manner via a polymerase extension reaction, polymerase chain reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific polymerase chain reaction, reverse-transcription reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific ligation reaction, and/or ligation reaction, using DNA polymerase, reverse transcriptase, DNA ligase, RNA ligase, DNA repair enzyme, DNA deaminase enzyme, RNase, RNaseH2, endonuclease, restriction endonuclease, exonuclease, CRISPR, DNA glycosylase or combinations thereof; preferentially extending, ligating, or amplifying one or more primers or probes whose 3'-OH end has been liberated in an enzyme and sequence-dependent process; using one or more blocking oligonucleotide primers comprising one or more mismatched bases at the 3' end or comprising one or more nucleotide analogs and a blocking group at the 3' end under conditions that interfere with polymerase extension or ligation during said reaction of target-specific primer or probes hybridized in a base-specific manner to DNA repair enzyme and DNA deaminase enzyme-treated unmethylated sequence or complement sequence thereof.

[0220] In another embodiment, the one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA, or protein markers comprise one or more of the following: a quantitative real-time PCR method (qPCR); a reverse transcriptase-polymerase chain reaction (RTPCR) method; a DNA repair enzyme and DNA deaminase-treated- qPCR method; a digital PCR method (dPCR); a DNA repair enzyme and DNA deaminase-treated- dPCR method; a ligation detection method, a ligase chain reaction, a restriction endonuclease cleavage method; a DNA or RNA nuclease cleavage method; a microarray hybridization method; a peptide-array binding method; an antibody-array method; a Mass spectrometry method; a liquid chromatography-tandem mass spectrometry (LC-MS/MS) method; a capillary or gel electrophoresis method; a chemiluminescence method; a fluorescence method; a DNA sequencing method; a DNA repair enzyme and DNA deaminase-treated -DNA sequencing method; an RNA sequencing method; a proximity ligation method; a proximity PCR method; a method comprising immobilizing an antibody-target complex; a method comprising immobilizing an aptamer-target complex; an immunoassay method; a method comprising a Western blot assay; a method comprising an enzyme linked immunosorbent assay (ELISA); a method comprising a high-throughput microarray-based enzyme-linked immunosorbent assay (ELISA); a method comprising a high-throughput flow-cytometry-based enzyme-linked immunosorbent assay (ELISA).

[0221] In certain embodiments, the one or more cutoff levels of the one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA,

or protein markers comprise one or more of the following calculations, comparisons, or determinations, in the one or more marker assays comparing samples from the disease vs. normal individual: the marker ΔC_t value is > 2 ; the marker ΔC_t value is > 4 ; the ratio of detected marker-specific signal is > 1.5 ; the ratio of detected marker-specific signal is > 3 ; the ratio of marker concentrations is > 1.5 ; the ratio of marker concentrations is > 3 ; the enumerated marker-specific signals differ by $> 20\%$; the enumerated marker-specific signals differ by $> 50\%$; the marker-specific signal from a given disease sample is $> 85\%$; $> 90\%$; $> 95\%$; $> 96\%$; $> 97\%$; or $> 98\%$ of the same marker-specific signals from a set of normal samples; the marker-specific signal from a given disease sample has a z-score of > 1.03 ; > 1.28 ; > 1.65 ; > 1.75 ; > 1.88 ; or > 2.05 compared to the same marker-specific signals from a set of normal samples.

[0222] Another aspect of the present application relates to a two-step method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The method involves obtaining a biological sample that includes exosomes, tumor-associated vesicles, markers within other protected states, cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. A first step is applied to the biological samples with an overall sensitivity of $> 80\%$ and an overall specificity of $> 90\%$ or an overall Z-score of > 1.28 to identify individuals more likely to be diagnosed or prognosed with the disease state. A second step is then applied to biological samples from those individuals identified in the first step with an overall specificity of $> 95\%$ or an overall Z-score of > 1.65 to diagnose or prognose individuals with the disease state. The first step and/or the second step are carried out using a method of the present application.

[0223] Fluorescent labeling. Consider an instrument that can detect 5 fluorescent signals, F1, F2, F3, F4, and F5, respectively. As an example, in the case of colon cancer, the highest frequency mutations will be found for K-ras, p53, APC and BRAF. Mutations in these four genes could be detected with a single fluorescent signal; F1, F2, F3, F4. If the scale is 1000 FU, then primer would be added using ratios of labeled and unlabeled UniTaq primers, such that amplification of LDR products on mutant target of these genes yields about 300 FU at the plateau. For the controls, the F5 would be calibrated to give a signal of 100 FU for a 1:1,000 dilution quantification control, and an additional 300 FU for ligation of mutant probe on wild-type control (should give no or low background signal).

[0224] For the other genes commonly mutated in colon cancer as shown below, (or even lower abundance mutations in the p53 gene,) the following coding system may be used: Two fluorescent signals in equimolar amount at the 5' end of the same UniTaq, with unlabeled primer titrated in, such that both fluorescent signals plateau at 100 FU. If fluorescent signals are F1, F2, F3, F4, then that gives the ability to detect mutations in 4 genes using a single fluorescent signal, and in mutations in 6 genes using combinations of fluorescent signal:

Gene 1 = F1 (300 FU)	(p53, Hot Spots)
Gene 2 = F2 (300 FU)	(KRAS)
Gene 3 = F3 (300 FU)	(APC)
Gene 4 = F4 (300 FU)	(BRAF)
Gene 5 = F1 (100 FU), F2 (100 FU)	(PIK3CA)
Gene 6 = F1 (100 FU), F3 (100 FU)	(FBXW7)
Gene 7 = F1 (100 FU), F4 (100 FU)	(SMAD4)
Gene 8 = F2 (100 FU), F3 (100 FU)	(p53, additional)
Gene 9 = F2 (100 FU), F4 (100 FU)	(CTNNB1)
Gene 10 = F3 (100 FU), F4 (100 FU)	(NRAS)

[0225] Suppose there is a second mutation, combined with a mutation in one of the top genes. This is easy to distinguish, since the top gene will always give more signal, independent if it is overlapping with the other fluorescent signals or not. For example, if the fluorescent signal is F1 100 FU, and F2 400 FU, that would correspond to mutations in Gene 2 and Gene 5.

[0226] If there are two mutations from the less commonly mutated genes (Gene 5- Gene 10) then the results will appear either as an overlap in fluorescent signals, *i.e.* F1 200 FU, F2 100 FU, F4 100 FU, or all 4 fluorescent signals. If the fluorescent signals are in the ratio of 2:1:1, then it is rather straightforward to figure out the 2 mutations: in the above example, F1 200 FU, F2 100 FU, F4 100 FU, would correspond to mutations in Gene 5 and Gene 7. A similar approach for multiplexing different colors has been described by the Kartalov group (Rajagopal et al., "Supercolor Coding Methods for Large-Scale Multiplexing of Biochemical Assays," *Anal. Chem.* 85(16):7629-36 (2013); U.S. Patent Application Publication No. 20140213471A1, which are hereby incorporated by reference in their entirety).

[0227] More recently, digital droplet PCR (ddPCR) has been used to provide accurate quantification of the number of mutant or methylated or hydroxymethylated molecules in a clinical sample. In general, amplification in a droplet implies at least a single molecule of the target was present in that droplet. Thus, when using a sufficient number of droplets that way exceed the number of initial targets, it is assumed that a given droplet only had a single molecule of the target. Thus, end-point PCR is often used to monitor the number of products.

[0228] Consider an instrument that can detect 5 fluorescent signals, F1, F2, F3, F4, and F5, respectively. Methylation in the promoter region of some genes often methylated in colon cancer could be used for the first four channels, for example F1 = VIM, F2 = SEPT9, F3 = CLIP4, and F4 = GSG1L. The last channel, F5 would be used as a control to assure a given droplet contained proper reagents, etc. Once again, combinations of fluorescent signal may be used to simultaneously detect methylation at 10 different promoter regions.

Gene 1 = F1	(VIM)
Gene 2 = F2	(SEPT9)
Gene 3 = F3	(CLIP4)
Gene 4 = F4	(GSG1L)
Gene 5 = F1 + F2	(PP1R16B)
Gene 6 = F1 + F3	(KCNA3)
Gene 7 = F1 + F4	(GDF6)
Gene 8 = F2 + F3	(ZNF677)
Gene 9 = F2 + F4	(CCNA1)
Gene 10 = F3 + F4	(STK32B)

[0229] For simplicity, consider how ddPCR may be used to accurately enumerate the number of original methylated or hydroxymethylated molecules at 4 promoter regions using exPCR-ddPCR (see for example, Figures 5 through 10, and 13 through 17). The approach also works using PCR-LDR-qPCR or exPCR-LDR-qPCR (see Figures 2, 3, 4, 11 and 12). For this illustration, consider a total of 48 methylated regions are being detected, with 4 promoter regions in a single ddPCR reaction comprising 10,000 droplets or micro-pores or micro-wells. Consider a sample with 2, 4, 5, and 1 molecule(s) of methylated promoter regions for VIM, SEPT9, CLIP4, and GSG1L, respectively. Assume the initial one-sided primer extension with blocking primer has an efficiency of 50%, so after 20 cycles, there are = 20; 40; 50; and 10 extension products of methylated promoter regions for VIM, SEPT9, CLIP4, and GSG1L, respectively. Also, with blocking primer for the top strand, again, assuming a general efficiency of 50%, after 10 cycles of PCR, there are (1.5 to the 10th = 57 x number of initial extension products) = 1,140; 2,280; 2,850; and 570 copies of the PCR products for methylated VIM, SEPT9, CLIP4, and GSG1L, respectively. When such products are then diluted into 12 ddPCR reactions, on average, a given chamber will comprise of 95; 190; 237; and 48 copies of the PCR products for methylated VIM, SEPT9, CLIP4, and GSG1L, respectively. This is a total of about 570 of molecules that would be amplified with primers for the total PCR products for methylated VIM, SEPT9, CLIP4, and GSG1L. If the ddPCR comprises 10,000 droplets or micro-pores or micro-wells, on average, only 1 in 20 will actually comprise a PCR reaction; the chances of a given droplet having two amplicons that would compete with each other for resources would be about

1 in 400, or about 25 droplets would comprise 2 amplicons, which would be only 5% of the total number of droplets with only a single amplicon. Since there are 6 combinations of 2 different amplicons, on average, less than 2% of the droplets would contain two amplicons. In other words, the rare droplet comprising 2 or 3 or 4 colors would not need to be de-convoluted, they could simply be ignored as they represent approximately 4-6 droplets compared to about 48 droplets arising from a single molecule in the original sample. While it may be a bit difficult to distinguish 190 from 237 droplets, *i.e.* starting with 4 or 5 molecules of a given methylated target, it should be relatively straightforward to distinguish 95; 190; and 48 copies, corresponding to 2, 4, and 1 target molecules in the original sample.

[0230] For distinguishing and enumerating 10 methylation markers simultaneously in a single ddPCR reaction, consider a total of 50 methylated or hydroxymethylated regions are being detected, with 10 promoter regions in a single ddPCR reaction comprising 10,000 droplets or micro-pores or micro-wells. Consider a sample with 2, 4, 5, 1, 0, 1, 3, 2, 0, and 1 molecule(s) of methylated promoter regions for VIM, SEPT9, CLIP4, GSG1L, PP1R16B, KCNA3, GDF6, ZNF677, CCNA1, and STK32B, respectively. Assume the initial one-sided primer extension with blocking primer has an efficiency of 50%, so after 20 cycles, there are = 20; 40; 50; 10; 0; 10; 30; 20; 0; and 10 extension products of methylated promoter regions for VIM, SEPT9, CLIP4, GSG1L, PP1R16B, KCNA3, GDF6, ZNF677, CCNA1, and STK32B, respectively. Also, with blocking primer for the top strand, again, assuming a general efficiency of 50%, after 6 cycles of PCR, there are $(1.5 \text{ to the } 6^{\text{th}} = 11 \times \text{number of initial extension products}) = 220; 440; 550; 110; 0; 110; 330; 220; 0; \text{ and } 110$ copies of the PCR products for methylated VIM, SEPT9, CLIP4, GSG1L, PP1R16B, KCNA3, GDF6, ZNF677, CCNA1, and STK32B, respectively. When such products are then diluted into 5 ddPCR reactions, on average, a given chamber will comprise of 44; 88; 110; 22; 0; 22; 66; 44; 0; and 22 copies of the PCR products for methylated VIM, SEPT9, CLIP4, GSG1L, PP1R16B, KCNA3, GDF6, ZNF677, CCNA1, and STK32B, respectively. This is a total of about 418 of molecules that would be amplified with primers for the total PCR products for methylated VIM, SEPT9, CLIP4, GSG1L, PP1R16B, KCNA3, GDF6, ZNF677, CCNA1, and STK32B. If the ddPCR comprises 10,000 droplets or micro-pores or micro-wells, on average, only 1 in 25 will actually comprise a PCR reaction; the chances of a given droplet having two amplicons that would compete with each other for resources would be about 1 in 625, or about 16 droplets would comprise 2 amplicons, which would be only 4% of the total number of droplets with only a single amplicon. Since there are 45 combinations of 2 different amplicons, on average, less than 0.1% of the droplets would contain a given two amplicons. In other words, the rare droplet comprising 2 or 3 or 4 colors

would not need to be de-convoluted, they could simply be ignored as they represent one or two droplets compared to about 22 droplets arising from a single molecule in the original sample. While it may be a bit difficult to distinguish 88 from 110 droplets, *i.e.* starting with 4 or 5 molecules of a given methylated or hydroxymethylated target, it should be relatively straightforward to distinguish 44, 88, and 22 copies, corresponding to 2, 4, and 1 target molecules in the original sample.

[0231] The above approach would also work for accurately enumerating mRNA, miRNA, ncRNA or lncRNA target molecules. The sample is used directly for subsequent ddPCR enumeration. For distinguishing and enumerating 10 mRNA, ncRNA, or lncRNA markers simultaneously in a single ddPCR reaction, consider a total of 50 mRNA, ncRNA or lncRNA regions are being detected in a single ddPCR reaction comprising 10,000 droplets or micro-pores or micro-wells. Once again, combinations of fluorescent signal may be used to simultaneously detect 10 mRNA or ncRNA markers.

Gene 1 = F1	(mRNA1)
Gene 2 = F2	(mRNA2)
Gene 3 = F3	(mRNA3)
Gene 4 = F4	(mRNA4)
Gene 5 = F1 + F2	(ncRNA5)
Gene 6 = F1 + F3	(ncRNA6)
Gene 7 = F1 + F4	(ncRNA7)
Gene 8 = F2 + F3	(ncRNA8)
Gene 9 = F2 + F4	(ncRNA9)
Gene 10 = F3 + F4	(ncRNA10)

[0232] Consider a sample with 2, 4, 15, 1, 0, 10, 3, 20, 0, and 1 molecule(s) of mRNA1-4 and ncRNA5-10, respectively. Six cycles of RT-PCR will generate 64 cDNA copies of each transcript generating = 128; 256; 960; 64; 0; 640; 192; 1280; 0; and 64 copies of mRNA1-4 and ncRNA5-10, respectively. When such products are then diluted into 5 ddPCR reactions, on average, a given chamber will comprise of 25; 51; 192; 13; 0; 128; 28; 256; 0; and 13 copies of the PCR products for mRNA1-4 and ncRNA5-10, respectively. This is a total of about 706 of molecules that would be amplified with primers for the total PCR products for methylated mRNA1-4 and ncRNA5-10. If the ddPCR comprises 10,000 droplets or micro-pores or micro-wells, on average, only 1 in 14 will actually comprise a PCR reaction. The two most common RNA's in this example; mRNA 3 and ncRNA5 would be present on average of 1 in 52 and 1 in 39, thus the chances of a given droplet having these two amplicons that would compete with each other for resources would be about 1 in 2028, or about 5 droplets would comprise 2 amplicons, which is still less than for a single molecule after amplification – which will generate

13 copies. In other words, the rare droplet comprising 2 or 3 or 4 colors would not need to be de-convoluted, they could simply be ignored as they represent from 1 to 5 droplets compared to at least 13 droplets arising from a single molecule in the original sample. If some RNA molecules are present in higher amounts, one can still de-convolute multiple signals arising from 2 amplicons in a given droplet, using the same approach of different color probes at different levels of FU (*i.e.* 300 FU for products with a single color; 100 FU each for products using 2 colors) as articulated earlier.

[0233] Another aspect of the present application relates to the ability to distinguish cancer at the earliest stages when analyzing markers within a blood sample. The average body contains about 6 liters (6,000 ml) of blood. A 10 ml sample will then comprise 1/600th of the sample. While some cancers (*i.e.* lung cancer, melanoma) have a high mutational load, other cancers (*i.e.* breast, ovarian) have few mutations, and even fewer at the earliest stages. In contrast, methylation changes in promoter regions (*i.e.* methylation markers) appear to be early events. For the purposes of the calculations below, assume that if a marker is present in the sample, it can be detected down to the single molecule level, independent of the technology that is being used to identify the marker.

[0234] On a practical level, different cancers have different frequencies for different mutational markers. For example, the mutation rate for gene K-ras is ~30% and > 90% for colorectal cancer and pancreatic cancer, respectively. While p53 is found mutated in about 50% of all cancers, more often than not, such a mutation is manifested in late-stage tumors. As a benchmark, a given cancer during its earliest stage, generates at least one detectable mutation. Suppose that at any given time, there are 200 mutated molecules circulating in the plasma of the patient. Given the total volume, if there is a 10 ml sample, taken, then there is about a 1/3rd chance that the sample will contain at least 1 mutated molecule. A more accurate prediction would be based on the Poisson distribution. If there are 200 objects (*i.e.* mutated molecules) distributed into 600 bins (*i.e.* 600 aliquots of 10 ml representing the total blood volume of a patient), Poisson calculation would indicate that: 72% of wells will have 0 objects, 23.7% will have 1 object, 3.9% will have 2 objects, 0.4% will have 3 objects, etc. In other words, 28.1% of the aliquots would have at least one mutated molecule. If the assay is capable of detecting every single mutated molecule, then its sensitivity would be 28.1%. Likewise, if there were 300 objects (*i.e.* mutated molecules) distributed into 600 bins (*i.e.* 600 aliquots of 10 ml), then: 61% of wells will have 0 objects, 30.3% will have 1 object, 7.6% will have 2 objects, 1.3% will have 3 objects, etc. In other words, 39.4% of the aliquots would have at least one mutated molecule. If the assay is capable of detecting every single mutated molecule, its sensitivity is at 39.4%.

Likewise, if there were 400 objects (*i.e.* mutated molecules) distributed into 600 bins (*i.e.* 600 aliquots of 10 ml), then: 51% of wells will have 0 objects, 34.3% will have 1 object, 11.5% will have 2 objects, 2.5% will have 3 objects, etc. In other words, 49% of the aliquots would have at least one mutated molecule. If the assay detects every single mutated molecule, its sensitivity would be 49%. Likewise, if there were 600 objects (*i.e.* mutated molecules) distributed into 600 bins (*i.e.* 600 aliquots of 10 ml), then the Poisson calculation would be: 36.8% of wells will have 0 objects, 36.8% will have 1 object, 18.3% will have 2 objects, 6.1% will have 3 objects, etc. In other words, 63.2% of the aliquots would have at least one mutated molecule. If the assay detects every single mutated molecule, then its sensitivity will be 63.2%. Nevertheless, on a practical level, even with a detectable marker load as high as 600 molecules, the assay would still miss 36.8% of early cancers for that type of tumor. Recent literature results have argued what constitutes “early cancer”, with some groups claiming stage I & II cancers are early cancer, while others claiming that stages I, II, and III cancers are early cancer, both the definition and type varies, but general when scoring for mutations the results have reported sensitivities ranging from around 20% to around 70% -- which translates into missing 30% to 80% of early cancers (Klein et al., “Development of a Comprehensive Cell-free DNA (cfDNA) Assay for Early Detection of Multiple Tumor Types: The Circulating Cell-free Genome Atlas (CCGA) Study,” *Journal of Clinical Oncology* 36(15):12021-12021 (2018); Liu et al., “Breast Cancer Cell-free DNA (cfDNA) Profiles Reflect Underlying Tumor Biology: The Circulating Cell-free Genome Atlas (CCGA) Study,” *Journal of Clinical Oncology* 36(15):536-536 (2018), which are hereby incorporated by reference in their entirety).

[0235] The above calculations are performed based on the assumption that detecting even a single mutation is sufficient to call a patient positive. Initial work identifying mutations in the blood from patients with metastatic disease revealed an average of 5 mutations not only in the patients, but also in age-matched controls (Razavi, et al., “Cell-free DNA (cfDNA) Mutations From Clonal Hematopoiesis: Implications for Interpretation of Liquid Biopsy Tests,” *Journal of Clinical Oncology* 35(15):11526-11526 (2017), which is hereby incorporated by reference in its entirety). This phenomenon, known as clonal hematopoiesis, results from accumulation of mutations in white-blood cells, that then undergo clonal expansion. Once the presence of such mutations are accounted for (by sequencing an aliquot of WBC DNA from the same individual), the accuracy or specificity of these tests has been set at 98%. For some cancers like ovarian cancer, which exhibit low mutation load, an estimated 60% of the disease at its early stage would be missed. To put these number in perspective, there were 20,240 new cases of ovarian cancer in the US in 2018. Thus, about 55 million women (over the age of 50) should be tested for the

disease. Such test would identify 8,096 women with ovarian cancer. However, there would be about 1.1 million false-positives. The positive predictive value of such a test would be around 0.74%. In other words, only one in 136 women who tested positive would actually have ovarian cancer, the rest would be false-positives.

[0236] For a multi-marker test of the present application, two or more markers need to be deemed positive in order the overall screening result to be deemed positive. By increasing the total number of individual markers used, as well as the number of markers required to call the overall screening test positive, both sensitivity and specificity for detecting early cancer may be improved. The overall early cancer detection sensitivity is a function of the average number of each marker in the blood, the average number of markers positive, the minimum number of markers required to call the sample positive, and the total number of markers scored. For example, if the test uses 12 methylation markers, that on average are methylated (or hydroxymethylated) in $> 50\%$ of tumors for that cancer type, then on average, about 6 markers will be methylated for a given sample. If on average there are 600 methylated molecules in the blood for each marker, then on average a total of $600 \times 600 = 3,600$ objects (*i.e.* methylated molecules) are distributed into 600 bins (*i.e.* 600 aliquots of 10 ml). As an approximate calculation based on the Poisson calculation, the distribution would be: 0.2% of wells will have 0 objects, 1.5% will have 1 object, 4.5% will have 2 objects, 8.9% will have 3 objects, 13.3% will have 4 objects, 16.0% will have 5 objects, 16.0% will have 6 objects, 13.8% will have 7 objects, 10.3% will have 8 objects etc. Suppose that at least two markers need to be called positive. In this case, 1.7% ($= 0.2\% + 1.5\%$) of the aliquots with either 0 or 1 object (*i.e.* methylation markers) would be called negative. Thus, if a minimum of two markers are required to call the sample positive, then the sensitivity of the assay would be $100\% - 1.7\% = 98.3\%$ sensitivity. Suppose that at least three markers need to be called positive. In this case, aliquots with either 0, 1 or 2 objects (*i.e.* methylation markers) would be called negative $= 0.2\% + 1.5\% + 4.5\% = 6.2\%$. Thus, if a minimum of two markers are required to call the sample positive, then the sensitivity of the assay would be $100\% - 6.2\% = 93.8\%$ sensitivity. It is understood that a small fraction of aliquots with 3 markers positive will contain 2 molecules of one marker, and one molecule of a second marker, and thus not contain the minimum of three *different* markers positive, nevertheless, this is a small deviation from the approximate calculation above.

[0237] The overall early cancer detection specificity is a function of the average number of markers positive, the false-positive rate for each individual marker, the minimum number of markers required to call the sample positive, and the total number of markers scored. To estimate the overall false-positive rate, a formula is used based on the probability of binning

different color socks into a number of drawers. Consider the percentage of false positives for each marker = “%FP”; the total number of markers “m”, and the minimum number of markers required to call the sample positive “n”. Then the formula for calculating overall false positive would be: $(\%FP)^n \times [m!/(m-n)!n!]$. Suppose that percentage of false positives for each marker = “%FP” is at 4%; the total number of markers “m” is 12, and the minimum number of markers required to call the sample positive “n” is 3. Then the above formula for overall false-positives would be $(4\%)^3 \times [12!/9!3!] = (4\%)^3 \times [12 \times 11 \times 10 / 6] = 1.4\%$. Thus, the overall specificity would be $[100\% - 1.4\%] = 98.6\%$. The actual individual false-positive rate may differ for different markers. Further, it may depend on the source of the false-positive signal. If for example, age-related methylation is due to clonal hematopoiesis, *i.e.* results from accumulation of methylations in white-blood cells, that then undergo clonal expansion. This type of false-positive may be mitigated by also scoring for methylation changes in white blood cells from the same patient. On the other hand, if the source of the false-positive signal is due to release of DNA into the plasma due to tissue inflammation, or for example breakdown of muscle tissue from weight-lifting, then mitigating that signal may require sampling the blood at a different time when the body is rested, or a month later when inflammation has subsided.

[0238] Figures 18, 19, and 20 illustrate results for calculated overall Sensitivity and Specificity for 24, 36, and 48 markers, respectively. These graphs are based on the assumption that the average individual marker sensitivity is 50%, and the average individual marker false-positive rate is from 2% to 5%. The sensitivity curves provide overall sensitivity as a function of the average number of molecules in the blood for each marker, with separate curves for each minimum number of markers needed to call a sample as positive. The specificity curves provide overall specificity as a function of individual marker false-positive rates, again with separate curves for each minimum number of markers needed to call a sample as positive. The calculated numbers for overall Sensitivity and Specificity for 12, 24, 36, 48, 72 and 96 markers, respectively are provided in the tables below.

Table 1.

12 Markers Sensitivity; Avg. Indiv. Mkr,: 50% Sensitivity			
Average Number of Molecules in Blood	Mutation, 1 Positive	12 markers, Minimum 2 Positive	12 markers, Minimum 3 Positive
150	22.1%	44.2%	19.1%
200	28.1%	59.4%	32.3%
240	33.0%	69.2%	43.0%
300	39.4%	80.1%	57.7%
400	48.8%	90.8%	76.2%
480	55.1%	95.2%	85.7%
600	63.2%	98.3%	93.8%

Table 2.

12 Marker Specificity		
Individual marker FP rate	Minimum 2 Markers Positive	Minimum 3 Markers Positive
2%	97.4%	99.9%
3%	94.1%	99.5%
4%	89.4%	98.6%
5%		97.2%

Table 3.

24 Markers Sensitivity; Avg. Indiv. Mkr,: 50% Sensitivity				
Average Number of Molecules in Blood	Mutation, 1 Positive	24 markers, Minimum 3 Positive	24 markers, Minimum 4 Positive	24 markers, Minimum 5 Positive
150	22.1%	57.7%	35.3%	18.5%
200	28.1%	76.2%	56.7%	37.1%
240	33.0%	85.7%	70.6%	52.4%
300	39.4%	93.8%	84.9%	71.5%
400	48.8%	98.6%	95.8%	90.0%
480	55.1%	99.6%	98.6%	96.2%
600	63.2%	99.9%	99.8%	99.2%

Table 4.

24 Marker Specificity			
Individual marker FP rate	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive
2%	98.4%	99.8%	99.9%
3%	94.6%	99.1%	99.9%
4%	87.1%	97.3%	99.6%
5%		93.4%	98.7%

Table 5.

36 Marker Sensitivity; Avg. Indiv. Mkr.; 50% Sensitivity					
Average Number of Molecules in Blood	Mutation, 1 Positive	36 markers, Minimum 3 Positive	36 markers, Minimum 4 Positive	36 markers, Minimum 5 Positive	36 markers, Minimum 6 Positive
150	22.1%	82.6%	65.8%	46.8%	29.7%
200	28.1%	93.8%	84.9%	71.5%	55.4%
240	33.0%	97.5%	92.8%	84.4%	72.4%
300	39.4%	99.4%	97.9%	94.5%	88.4%
400	48.8%	99.9%	99.8%	99.2%	98.0%
480	55.1%	100.0%	100.0%	99.9%	99.6%
600	63.2%	100.0%	100.0%	100.0%	100.0%

Table 6.

36 Marker Specificity				
Individual marker FP rate	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive
2%	94.3%	99.1%	99.9%	100.0%
3%	80.7%	95.2%	99.1%	99.9%
4%		84.9%	96.1%	99.2%
5%			88.2%	97.0%

Table 7.

48 Marker Sensitivity; Avg. Indiv. Mkr.; 50% Sensitivity						
Average Number of Molecules in Blood	Mutation, 1 Positive	48 markers, Minimum 4 Positive	48 markers, Minimum 5 Positive	48 markers, Minimum 6 Positive	48 markers, Minimum 7 Positive	48 markers, Minimum 8 Positive
150	22.1%	84.9%	71.6%	55.6%	39.6%	25.8%
200	28.1%	95.8%	90.1%	80.9%	68.7%	54.8%
240	33.0%	99.1%	97.2%	93.4%	87.1%	78.1%
300	39.4%	99.8%	99.3%	98.1%	95.6%	92.3%
400	48.8%	99.9%	99.9%	99.8%	99.7%	99.1%
480	55.1%	99.9%	99.9%	99.9%	99.9%	99.9%
600	63.2%	99.9%	99.9%	99.9%	99.9%	99.9%

Table 8.

48 Marker Specificity					
Individual marker FP rate	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive
2%	96.9%	99.4%	99.9%	99.9%	99.9%
3%	84.3%	95.8%	99.1%	99.8%	99.9%
4%		82.5%	95.0%	98.8%	99.8%
5%				94.3%	98.6%

Table 9.[illegible]

Table 10.

72 Marker Specificity							
Individual marker FP rate	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive	Minimum 9 Markers Positive	Minimum 10 Markers Positive	Minimum 11 Markers Positive	Minimum 12 Markers Positive
2%	99.0%	99.8%	100.0%	100.0%	100.0%	100.0%	100.0%
3%	88.6%	96.8%	99.2%	99.8%	100.0%	100.0%	100.0%
4%			92.2%	97.8%	99.4%	99.9%	100.0%
5%				83.4%	94.8%	98.5%	99.6%

Table 11.

96 Marker Sensitivity; Avg. Indiv. Mkr.; 50% Sensitivity								
Average Number of Molecules in Blood	Mutation, 1 Positive	96 markers, Minimum 7 Positive	96 markers, Minimum 8 Positive	96 markers, Minimum 9 Positive	96 markers, Minimum 10 Positive	96 markers, Minimum 11 Positive	96 markers, Minimum 12 Positive	96 markers, Minimum 13 Positive
150	22.1%	95.4%	91.0%	84.5%	75.8%	65.3%	53.8%	42.4%
200	28.1%	99.6%	99.0%	97.8%	95.7%	92.3%	87.3%	80.7%
240	33.0%	99.9%	99.9%	99.7%	99.2%	98.3%	96.8%	94.4%
300	39.4%	99.9%	99.9%	99.9%	99.9%	99.9%	99.7%	99.5%
400	48.8%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
480	55.1%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
600	63.2%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%

Table 12.

96 Marker Specificity							
Individual marker FP rate	Minimum 7 Markers Positive	Minimum 8 Markers Positive	Minimum 9 Markers Positive	Minimum 10 Markers Positive	Minimum 11 Markers Positive	Minimum 12 Markers Positive	Minimum 13 Markers Positive
2%	98.5%	99.7%	99.9%	100.0%	100.0%	100.0%	100.0%
3%		91.3%	97.4%	99.3%	99.8%	100.0%	100.0%
4%				88.2%	96.3%	99.0%	99.7%
5%						84.7%	95.1%

[0239] From the above tables, the receiver operating characteristic (ROC) curves may be calculated by plotting Sensitivity vs. 1-Specificity. Since these are theoretical calculations, the curves were generated for different levels of average marker false-positive rates of 2%, 3%, 4%, and 5%. To assist in visualizing the graphs and calculating the AUC (Area under curve), the edges were set at 100% and 0%, respectively. The ROC curves for 24 marker, 3% & 4% average marker false-positives, 36 marker, 3% & 4% average marker false-positives, and 48 marker, 2%, 3%, 4% & 5% average marker false-positives are provided in Table 13 below and for 48 Markers illustrated in Figures 21 and 22, respectively. Generally, the closer the AUC is to

1, the more accurate the test – values of >95% are desirable, and values >99% are superb. Using the benchmark of an average of 300 molecules in the blood for early cancer (Stage I & II), AUC values are at 95% with 24 markers, improve to 99% with 36 markers, and range from 98% to >99% with 48 markers, depending on average marker false-positive values. These graphs provide an illustration of the power of multiple marker assays for achieving good sensitivities and specificities.

Table 13.

24, 36, & 48 Marker AUC Values from ROC Curves; Avg. Indiv. Mkr,: 50% Sensitivity							
Total Markers: Individual marker FP rate	150 Molecules	200 Molecules	240 Molecules	300 Molecules	400 Molecules	480 Molecules	600 Molecules
24 Mkrs: 3%	77%	87%		96%	99%	>99%	>99%
24 Mkrs: 4%	74%	85%		95%	99%	>99%	>99%
36 Mkrs: 3%	87%	95%	98%	99%	>99%	>99%	>99%
36 Mkrs: 4%	78%	89%	95%	98%	99%	>99%	>99%
48 Mkrs: 2%	92%	98%	99%	>99%	>99%	>99%	>99%
48 Mkrs: 3%	89%	97%	99%	>99%	>99%	>99%	>99%
48 Mkrs: 4%	81%	93%	98%	99%	>99%	>99%	>99%
48 Mkrs: 5%	71%	86%	94%	98%	99%	99%	99%

[0240] How would the above markers work in a one-step cancer assay? To illustrate the challenges of developing an early cancer detection screen, consider the challenge of screening 107 million adults in the U.S. over the age of 50 for colorectal cancer – of which there are about 135,000 new cases that are diagnosed a year. In this example, if there is an average of 300 molecules in the blood for early cancer (Stage I & II), and taking the best-case scenario of individual marker FP rate is 2%, then if there is a 3-marker minimum, then overall FP rate is 1.6% for 24 markers, for a specificity of 98.4% (See Figure 18B). At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would miss 6.2%; *i.e.* for Stage I & II cancer the overall sensitivity would be 93.8% (See Figure 18A), *e.g.* the test would correctly identify 93.8% of individuals with disease, which would be 126,630 individuals (out of 135,000 new cases). At a specificity of 98.4%, for 107 million individuals screened, the test would also generate $1.6\% \times 107,000,000 = 1,712,000$ false positives. Thus, the positive predictive value would be $126,630 / (126,630 + 1,712,000) = \text{around } 6.8\%$, in other words, only one in 14 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives.

[0241] However, if the individual marker FP rate is more realistic, say 4%, then more markers will be required to achieve better than 98% specificity, and this will be at the cost of sensitivity. If individual marker FP rate is 4%, then if there is a 5-marker minimum, then overall FP rate is 0.4% for 24 markers, for a specificity of 99.6% (See Figure 18B). At 5 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would miss 28.5%; *i.e.* for Stage I & II cancer the overall sensitivity would be 71.5% (See Figure 18A), *e.g.* the test would correctly identify 71.5% of individuals with disease, which would be 90,540 individuals (out of 135,000 new cases). At a specificity of 99.6%, for 107 million individuals screened, the test would also generate $0.4\% \times 107,000,000 = 428,000$ false positives. Thus, the positive predictive value would be $90,540 / (90,540 + 428,000) =$ around 17.5%, in other words, only one in 5.7 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. A PPV of 17.5% is quite respectable, however, it would be achieved at the cost of missing 28.5% of early cancer.

[0242] As described above, another aspect of the present application relates to a two-step method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual. The method involves obtaining a biological sample that includes exosomes, tumor-associated vesicles, markers within other protected states, cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells. The biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof. A first step is applied to the biological samples with an overall sensitivity of $> 80\%$ and an overall specificity of $> 90\%$ or an overall Z-score of > 1.28 to identify individuals more likely to be diagnosed or prognosed with the disease state. A second step is then applied to biological samples from those individuals identified in the first step with an overall specificity of $> 95\%$ or an overall Z-score of > 1.65 to diagnose or prognose individuals with the disease state. The first step and/or the second step are carried out using a method of the present application.

[0243] The first step and the second step are carried out using a method of the present application. The first step uses markers to cover many cancers, where the aim is to obtain high sensitivity for early cancers where the number of marker molecules in the blood may be limited. The second step then would score for additional markers to verify that the initial result was a true positive, as well as to identify the likely tissue of origin. The second step may include the methods described herein, and/or additional methods such as next-generation sequencing.

[0244] To illustrate one embodiment of how such a two-step cancer test may be designed, consider again the challenge of identifying patients with early colorectal cancer. In 2017, there were an estimated 95,520 new cases of colon cancer and 39,910 cases of rectal cancer diagnosed in the US – for a total of about 135,000 new cases. Consider an initial test using 24 markers. In this example, if there is an average of 300 molecules in the blood for early cancer (Stage I & II), and if that would cover at least one mutation, then the sensitivity for identifying such a cancer by next generation sequencing would be 39.4% (See Figure 18A). If the individual marker FP rate is 3%, then if there is a 3-marker minimum, then overall FP rate is 5.4% for 24 markers, for a specificity of 94.6% (See Figure 18B). At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would miss 6.2%; *i.e.* for Stage I & II cancer the overall sensitivity would be 93.8% (See Figure 18A). Note that these levels of sensitivity and specificity are better than the current tests on the market. However, if the individual marker FP rate is 5%, then if there is a 4-marker minimum, then overall FP rate is 6.6% for 24 markers, for a specificity of 93.4% (See Figure 18B). At 4 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would miss 15.1%; *i.e.* for Stage I & II cancer the sensitivity would be 84.9% (See Figure 18A). These graphs illustrate a basic conflict of most diagnostic tests – improve the sensitivity of a test (*i.e.* less false-negatives), but sacrifice the test specificity (*i.e.* more false-positives), or improve the specificity of a test (less false-positives) at the risk of losing the test sensitivity (*i.e.* more false-negatives).

[0245] By using a two-step cancer test, the parameters may be adjusted to improve BOTH sensitivity and specificity. For example, the aforementioned 24 marker test, using 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the overall sensitivity would be 93.8%. Those samples that are scored as positives in the first step (24-markers specific to GI cancers) – including the false-positives would be retested in the second step with an expanded panel of 48 markers to provide coverage of colorectal cancers. If the individual marker FP rate is 3%, then if there is a 5-marker minimum, then overall FP rate is 4.2% for 48 markers, for a specificity of 95.8% (See Figure 20B). At 5 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would miss 0.7%; *i.e.* for Stage I & 2 cancer the sensitivity would be 99.3% (See Figure 20A). Technically, since the samples were already culled in the first step, the overall sensitivity is $93.8\% \times 99.3\% = 93.1\%$. If the individual marker FP rate is 3%, then if there is a 6-marker minimum, then overall FP rate is <1% for 48 markers, for a specificity of 99.1% (See Figure 18B). At 6 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would

miss 1.9%; *i.e.* for Stage I & II cancer the sensitivity would be 98.1% (See Figure 20A). Since the samples were already culled in the first step, the overall sensitivity is $93.8\% \times 98.1\% = 92.0\%$. If the individual marker FP rate is 3%, then if there is a 7-marker minimum, then overall FP rate is $<0.2\%$ for 48 markers, for a specificity of 99.8% (See Figure 20B). At 7 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), the test would miss 4.4%; *i.e.* for Stage I & II cancer the sensitivity would be 95.6% (See Figure 20A). Since the samples were already culled in the first step, the overall sensitivity is $93.8\% \times 95.6\% = 89.7\%$.

[0246] Returning to the example of colorectal cancer, in particular the cases of microsatellite stable tumors (MSS) where the mutation load is low, for these calculations when relying on NGS sequencing alone (assuming 300 molecules with one mutation in the blood), an estimated 60% of early colorectal cancer would be missed. Again, to put these number in perspective, in the U.S., about 135,000 new cases of colorectal cancer are predicted in 2018. About 107 million individuals in the U.S. are over the age of 50 and should be tested for colorectal cancer. With the assumption of these samples containing at least 300 molecules with one mutation in the blood, such a test would find 54,000 men and women (out of 135,000 new cases) with colorectal cancer. However, with a specificity for sequencing at 98%, there would be about 2.1 million false-positives. The positive predictive value of such a test would be around 2.6%, in other words, only one in 39 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. In contrast, consider the two-step methylation marker test described above, wherein the first step has 24 methylation markers specific to GI cancers, while the second step has 48 methylation markers specific to colorectal cancer. In this example, as above, the calculations are done with the anticipation of an average of 300 methylated (or hydroxymethylated) molecules per positive marker in the blood. Assuming individual marker false-positive rates of 3%, and with the first step requiring a minimum of 3 markers positive, then with an overall specificity of 94.6%, the first step would identify 5,778,000 individuals (out of 107,000,000 total adults over 50 in the U.S.) which would include at 93.8% sensitivity about 126,630 individuals with Stage I & II colorectal cancer (out of 135,000 total). However, those 5,778,000 presumptive positive individuals would be evaluated in the second step of 48 markers requiring a minimum of 6 markers positive, then the two-step test would identify $98.1\% \times 93.8\% = 92.0\% = 124,200$ individuals (out of 135,000 new cases) with colorectal cancer. With a specificity of 99.1%, the second test would also generate $5,778,000 \times 0.9\% = 52,000$ false-positives. The positive predictive value of such a test would be $124,200/176,200 = 70.5\%$, in other words, 2 in 3 individuals who tested positive would actually

have colorectal cancer, an extraordinarily successful screen to focus on those patients who would most benefit from follow-up colonoscopy. The benefit in lives saved would be of incalculable value.

[0247] While the foregoing discussion has focused on methylation markers, with an average sensitivity of 50%, and individual marker false-positives ranging from 2%-5%, there are many other markers of cancer with varying sensitivities and specificities. In general, protein markers (with the exception of PSA and PSMA) have been of limited clinical utility for detection of early cancer because the false-positives are so high, resulting in very low positive predictive value. Cancer markers from bodily fluids (*i.e.* plasma, urine) include, but are not limited to plasma microRNAs (miRNA); mutations or methylation in cfDNA; exosomes with surface cancer-specific protein markers, or internal miRNA, ncRNA, lncRNA, mRNA, DNA; circulating cytokines, circulating proteins, or circulating antibodies against cancer-antigens; or nucleic-acid markers in whole blood (for review, see Nikolaou et al., “Systematic Review of Blood Diagnostic Markers in Colorectal Cancer,” *Techniques in Coloproctology* (2018), which is hereby incorporated by reference in its entirety). Several methods have been reported for detecting cancer-specific miRNAs in the serum or plasma of patients with colorectal (or others) cancers; these miRNAs include, but are not limited to: miR-1290; miR-21; miR-24; miR-320a; miR-423-5p; miR-29a; miR-125b; miR-17-3p; miR-92a; miR-19a; miR-19b; miR-15b; miR-23a; miR-150; miR-223; miR-1229; miR-1246; miR-612; miR-1296; miR-933; miR-937; miR-1207; miR-31; miR-141; miR-224-3p; miR-576-5p; miR-885-5p; miR-200c; miR-203 (Imaoka et al., “Circulating MicroRNA-1290 as a Novel Diagnostic and Prognostic Biomarker in Human Colorectal Cancer,” *Ann. Oncol.* 27(10):1879–1886 (2016); Zhang et al., “Diagnostic and Prognostic Value of MicroRNA-21 in Colorectal Cancer: an Original Study and Individual Participant Data Meta-Analysis,” *Cancer Epidemiol. Biomark. Prev.* 23(12):2783–2792 (2016); Fang et al., “Plasma Levels of MicroRNA-24, MicroRNA-320a, and Micro-RNA-423-5p are Potential Biomarkers for Colorectal Carcinoma,” *J. Exp. Clin. Cancer Res.* 34:86 (2015); Toiyama et al., “MicroRNAs as Potential Liquid Biopsy Biomarkers in Colorectal Cancer: A Systematic Review,” *Biochim. Biophys. Acta*. pii: S0304-419X(18)30067-2 (2018); Nagy et al., “Comparison of Circulating miRNAs Expression Alterations in Matched Tissue and Plasma Samples During Colorectal Cancer Progression,” *Pathol. Oncol. Res.* doi: 10.1007/s12253-017-0308-1 (2017); Wang et al., “Novel Circulating MicroRNAs Expression Profile in Colon Cancer: a Pilot Study,” *Eur. J. Med. Res.* 22(1):51 (2017); U.S. Patent No. 9,689,036 to Getts et al.; U.S. Patent No. 9,708,643 to Dutttagupta et al.; U.S. Patent No. 9,868,992 to Goel et al; U.S. Patent No. 9,926,603 to Sozzi et al., which are hereby incorporated by reference in their entirety).

Additional approaches for detecting low abundance miRNA are described in WO2016057832A2, which is hereby incorporated by reference in its entirety, or using other suitable means known in the art. Figure 23 provides a list of blood-based, colon cancer-specific microRNA markers derived through analysis of TCGA microRNA datasets, which may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood.

[0248] Several methods have been reported for detecting cancer-specific ncRNA or lncRNAs in the serum, plasma, or exosomes of patients with colorectal (and other) cancers; these ncRNAs include, but are not limited to: NEAT_v1; NEAT_v2; CCAT1; HOTAIR; CRNDE-h; H19; MALAT1; 91H; GAS5 (Wu et al., “Nuclear-enriched Abundant Transcript 1 as a Diagnostic and Prognostic Biomarker in Colorectal Cancer,” *Mol. Cancer* 14:191 (2015); Zhao et al., “Combined Identification of Long Non-Coding RNA CCAT1 and HOTAIR in Serum as an Effective Screening for Colorectal Carcinoma,” *Int. J. Clin. Exp. Pathol.* 8(11):14131-40 (2015); Liu et al., “Exosomal Long Noncoding RNA CRNDE-h as a Novel Serum-Based Biomarker for Diagnosis and Prognosis of Colorectal Cancer,” *Oncotarget* 7(51):85551-85563 (2016); Slaby O, “Non-coding RNAs as Biomarkers for Colorectal Cancer Screening and Early Detection,” *Adv Exp Med Biol.* 937:153-70 (2016); Gao et al., “Exosomal lncRNA 91H is Associated With Poor Development in Colorectal Cancer by Modifying HNRNPK Expression,” *Cancer Cell Int.* 23;18:11 (2018); Liu et al., “Prognostic and Predictive Value of Long Non-Coding RNA GAS5 and MicroRNA-221 in Colorectal Cancer and Their Effects on Colorectal Cancer Cell Proliferation, Migration and Invasion,” *Cancer Biomark.* 22(2):283-299 (2018); U.S. Patent No. 9,410,206 to Hoon et al.; U.S. Patent No. 9,921,223 to Kalluri et al., which are hereby incorporated by reference in their entirety). Additional approaches for detecting low abundance lncRNA, ncRNA, mRNA translocations, splice variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicing, exon insertions, exon deletions, and intron insertions are described in WO2016057832A2, which is hereby incorporated by reference in its entirety, or using other suitable means known in the art. Figure 24 provides a list of blood-based, colon cancer-specific ncRNA and lncRNA markers derived through analysis of various publicly available Affymetrix Exon ST CEL data, which were aligned to GENCODE annotations to generate ncRNA and lncRNA transcriptome datasets. Comparative analyses across these datasets (various cancer types, along with normal tissues, and peripheral blood) were conducted to generate the ncRNA and lncRNA markers list. Such lncRNA and ncRNA may be enriched in exosomes or other protected states in the blood. In addition, Figure 25 provides a list of blood-based colon cancer-

specific exon transcripts that may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood.

[0249] The most common protein marker for colorectal cancer is based on detecting hemoglobin from blood in the stool and is known as the FOBT or FIT test. Sensitivity and specificities (Sens.: Spec.) for these tests have been reported as: OC-Light iFOB Test (also called OC Light S FIT), manufactured by Polymedco (78.6%-97.0% : 88.0%-92.8%); QuickVue iFOB, manufactured by Quidel (91.9% : 74.9%); Hemosure One-Step iFOB Test, manufactured by Hemosure, Inc. (54.5% : 90.5%); InSure FIT, manufactured by ClinicalGenomics (75.0% : 96.6%); Hemocult-ICT, manufactured by Beckman Coulter (23.2%-81.8% : 95.8%-96.9%); Cologuard – stool FIT-DNA, manufactured by Exact Sciences (92.3%; 84.4%). The large ranges and differences in sensitivities and specificities may reflect the range from early to late cancer, as well as differences in methodology, number of samples collected, and clinical study size. Cut-off values for FIT tests may range from 10 ug protein/gram stool to 300 ug protein/gram stool (See Robertson et al., “Recommendations on Fecal Immunochemical Testing to Screen for Colorectal Neoplasia: a Consensus Statement by the US Multi-Society Task Force on Colorectal Cancer,” *Gastrointest. Endosc.* 85(1):2-21 (2017), which is hereby incorporated by reference in its entirety).

[0250] A number of tumor-associated antigens elicit an immune response within the patient, and these may be identified as autoantibodies, or indirectly as increased cytokines in the serum. Some tumor antigens may be detected directly within the serum, or on the surface of cancer-associated exosomes or extracellular vesicles, while others may be detected indirectly, for example by an increase in mRNA within cancer-associated exosomes or extracellular vesicles. These cancer-specific protein markers may be identified through, mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product, and these markers include but are not limited to: RPH3AL; RPL36; SLP2; TP53; Survivin; ANAXA4; SEC61B; CCCAP; NYCO16; NMDAR; PLSCR1; HDAC5; MDM2; STOML2; SEC61-beta; IL8; TFF3; CA11-19; IGFBP2; DKK3; PKM2; DC-SIGN; DC-SIGNR; GDF-15; AREG; FasL; Flt3L; IMPDH2; MAGEA4; BAG4; IL6ST; VWF; EGFR; CD44; CEA; NSE; CA 19-9; CA 125; NMMT; PSA; proGRP; DPPIV/seprase complex; TFAP2A; E2F5; CLIC4; CLIC1; TPM1; TPM2; TPM3; TPM4; CTSD-30; PRDX-6; LRG1; TTR; CLU; KLKB1; C1R; KLK3; KLK2; HOXB13; GHRL2; FOXA1 (Fan et al., “Development of a Multiplexed Tumor-Associated Autoantibody-Based Blood Test for the Detection of Colorectal Cancer,” *Clin. Chim. Acta.* 475:157-163 (2017); Xia et al., “Prognostic Value, Clinicopathologic Features and Diagnostic Accuracy of Interleukin-8 in Colorectal Cancer: a Meta-Analysis,” *PLoS One*

10(4):e0123484 (2015); Li et al., “Serum Trefoil Factor 3 as a Protein Biomarker for the Diagnosis of Colorectal Cancer,” *Technol. Cancer. Res. Treat.* 16(4):440-445 (2017); Overholt et al., “CA11-19: a Tumor Marker for the Detection of Colorectal Cancer,” *Gastrointest. Endosc.* 83(3):545-551 (2016); Fung et al., “Blood-based Protein Biomarker Panel for the Detection of Colorectal Cancer,” *PLoS One* 10(3): e0120425 (2015); Jiang et al., “The Clinical Significance of DC-SIGN and DC-SIGNR, Which are Novel Markers Expressed in Human Colon Cancer,” *PLoS One* 9(12):e11474 (2014).; Chen et al., “Development and Validation of a Panel of Five Proteins as Blood Biomarkers for Early Detection of Colorectal Cancer,” *Clin. Epidemiol.* 9:517-526 (2017); Chen et al., “Prospective Evaluation of 64 Serum Autoantibodies as Biomarkers for Early Detection of Colorectal Cancer in a True Screening Setting,” *Oncotarget* 7(13):16420-32 (2016); Rho et al., “Protein and Glycomic Plasma Markers for Early Detection of Adenoma and Colon Cancer,” *Gut* 67(3):473-484 (2018); U.S. Patent No. 9,518,990 to Wild et al.; U.S. Patent No. 9,835,636 to Chan et al.; U.S. Patent No. 9,885,718 to Man et al.; U.S. Patent No. 9,889,135 to Andy Koff et al.; U.S. Patent No. 9,903,870 to Speicher et al.; U.S. Patent No. 9,914,974 to Bajic et al.; U.S. Patent No. 9,983,208 to Choi et al; U.S. Patent No. 10,030,271 to Scher et al., which are hereby incorporated by reference in their entirety). Additional approaches for detecting low abundance mRNA translocations, splice variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicing, exon insertions, exon deletions and intron insertions are described in WO2016057832A2, which is hereby incorporated by references in its entirety, or using other suitable means known in the art. Figure 26 provides a list of cancer protein markers, identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from Colorectal tumors, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma. Figure 27 provides protein markers that can be secreted by Colorectal tumors into the blood. A comparative analysis was performed across various TCGA datasets (tumors, normals), followed by an additional bioinformatics filter (Meinken et al., “Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report,” *Computational Molecular Biology* 2(1):1-7 (2012), which is hereby incorporated by reference in its entirety), which predicts the likelihood that the translated protein is secreted by the cells.

[0251] The distribution of mutations in colorectal cancers are available in the public COSMIC database, with the 20 most commonly altered genes listed as: APC; TP53; KRAS; FAT4; LRP1B; PIK3CA; TGFBR2; ACVR2A; BRAF; ZFH3; KMT2C; KMT2D; FBXW7; SMAD4; ARID1A; TRRAP; RNF43; FAT1; TCF7L2; PREX2 (Forbes et al., “COSMIC:

Exploring the World's Knowledge of Somatic Mutations in Human Cancer,” *Nucleic Acids Res.* 43(Database issue):D805-811 (2015), which is hereby incorporated by reference in its entirety). Analysis of TCGA COADREAD mutational dataset, however indicate the following genes have at least 10% mutation rate among colorectal cancer primary tumors: APC, TP53, KRAS, TTN, SYNE1, PIK3CA, FAT4, MUC16, FBXW7, LRP1B, LRP2, DNAH5, DMD, ANK2, RYR2, FLG, HMCN1, FAT2, TCF7L2, CSMD3, USH2A, SDK1, CSMD1, COL6A3, DNAH2, SMAD4, PKHD1, FAM123B, ATM, ACVR2A, MDN1, DCHS2, ZFX4, CUBN, CSMD2, FREM2, RYR1, TGFB2, RYR3, SACS, DNAH10, ABCA12, BRAF, ODZ1, PCDH9, MACF1, AHNK2. In addition to the approaches described herein, there are several approaches for enriching for and detecting low-abundance mutations either at the DNA or mRNA level (for example, mRNA within exosomes), including but not limited to next generation sequencing, allele-specific PCR, ARMS, primer-extension PCR, using blocking primers, full-COLD-PCR, fast-COLD-PCR, ice-COLD-PCR, E-ice-COLD-PCR, TT-COLD-PCR, etc. (Mauger et al., “COLD-PCR Technologies in the Area of Personalized Medicine: Methodology and Applications,” *Mol. Diagn Ther.* 3:269-283 (2017); Sefrioui et al., “Comparison of the Quantification of KRAS Mutations by Digital PCR and E-ice-COLD-PCR in Circulating-Cell-Free DNA From Metastatic Colorectal Cancer Patients,” *Clin. Chim. Acta.* 465:1-4 (2017); U.S. Patent No. 9,062,350 to Platica; U.S. Patent No. 9,598,735 to Song et al., which are hereby incorporated by reference in their entirety). Additional approaches for detecting low abundance mutations are described in WO2016057832A2, which is hereby incorporated by reference in its entirety, or using other suitable means known in the art.

[0252] The best studied blood-based methylation markers for CRC detection are located in the promoter region of the SEPT9 gene (Church et al., “Prospective Evaluation of Methylated SEPT9 in Plasma for Detection of Asymptomatic Colorectal Cancer,” *Gut* 63(2):317-325 (2014); Lofton-Day et al., “DNA Methylation Biomarkers for Blood-Based Colorectal Cancer Screening,” *Clinical Chemistry* 54(2):414-423 (2008); Potter et al., “Validation of a Real-time PCR-based Qualitative Assay for the Detection of Methylated SEPT9 DNA in Human Plasma,” *Clinical Chemistry* 60(9):1183-1191 (2014); Ravegnini et al., “Simultaneous Analysis of SEPT9 Promoter Methylation Status, Micronuclei Frequency, and Folate-Related Gene Polymorphisms: The Potential for a Novel Blood-Based Colorectal Cancer Biomarker,” *International Journal of Molecular Sciences* 16(12):28486-28497 (2015); Toth et al., “Detection of Methylated SEPT9 in Plasma is a Reliable Screening Method for Both Left- and Right-sided Colon Cancers,” *PLoS One* 7(9):e46000 (2012); Toth et al., “Detection of Methylated Septin 9 in Tissue and Plasma of Colorectal Patients With Neoplasia and the Relationship to the Amount of Circulating Cell-free

DNA,” *PloS One* 9(12):e115415 (2014); Warren et al., “Septin 9 Methylated DNA is a Sensitive and Specific Blood Test for Colorectal Cancer,” *BMC Medicine* 9:133 (2011), which are hereby incorporated by reference in their entirety), and other potential markers for CRC diagnostics include CpG sites on promoter regions of THBD, C9orf50, ZNF154, AGBL4, FLI1, and TWIST1 (Lange et al., “Genome-scale Discovery of DNA-methylation Biomarkers for Blood-based Detection of Colorectal Cancer,” *PloS One* 7(11):e50266 (2012); Margolin et al., “Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development,” *The Journal of Molecular Diagnostics: JMD* 18(2):283-298 (2016); Lin et al., “Clinical Relevance of Plasma DNA Methylation in Colorectal Cancer Patients Identified by Using a Genome-Wide High-Resolution Array,” *Ann. Surg. Oncol.* 22 Suppl 3:S1419-1427 (2015), which are hereby incorporated by reference in their entirety).

[0253] *SEPT9* methylation is the basis for Epi proColon test, a CRC-detection assay by Epigenomics (Lofton-Day et al., “DNA Methylation Biomarkers for Blood-based Colorectal Cancer Screening,” *Clinical Chemistry* 54(2):414-423 (2008), which is hereby incorporated by reference in its entirety). While initial results on smaller sample sets showed promise, large-scale studies with 1,544 plasma samples showed a sensitivity of 64% for stage I-III CRC, and a specificity of 78%-82%, effectively sending 180 to 220 out of 1,000 individuals to unnecessary colonoscopies (Potter et al., “Validation of a Real-time PCR-based Qualitative Assay for the Detection of Methylated SEPT9 DNA in Human Plasma,” *Clinical Chemistry* 60(9):1183-1191 (2014), which is hereby incorporated by reference in its entirety). ClinicalGenomics is currently developing blood-based CRC detection test based on the methylation of the *BCAT1* and *IKZF1* genes (Pedersen et al., “Evaluation of an Assay for Methylated BCAT1 and IKZF1 in Plasma for Detection of Colorectal Neoplasia,” *BMC Cancer* 15:654 (2015), which is hereby incorporated by reference in its entirety). Large-scale studies using 2,105 plasma samples of this two-marker test showed an overall sensitivity of 66%, with 38% for stage I CRC, and an impressive specificity of 94%. Exact Sciences and collaborators have slightly improved the sensitivity of CRC fecal tests (Bosch et al., “Analytical Sensitivity and Stability of DNA Methylation Testing in Stool Samples for Colorectal Cancer Detection,” *Cell Oncol. (Dordr)* 35(4):309-315 (2012); Hong et al., “DNA Methylation Biomarkers of Stool and Blood for Early Detection of Colon Cancer,” *Genet. Test. Mol. Biomarkers* 17(5):401-406 (2013); Imperiale et al., “Multitarget Stool DNA Testing for Colorectal-Cancer Screening,” *N. Engl. J. Med.* 370(14):1287-1297 (2014); Xiao et al., “Validation of Methylation-Sensitive High-resolution Melting (MS-HRM) for the Detection of Stool DNA Methylation in Colorectal Neoplasms,” *Clin. Chim. Acta.* 431:154-163 (2014); Yang et al., “Diagnostic Value of Stool DNA Testing for

Multiple Markers of Colorectal Cancer and Advanced Adenoma: a Meta-Analysis,” *Can. J. Gastroenterol.* 27(8):467-475 (2013), which are hereby incorporated by reference in their entirety), by adding K-ras mutation as well as BMP3 and NDRG4 methylation markers (Lidgard et al., “Clinical Performance of an Automated Stool DNA Assay for Detection of Colorectal Neoplasia,” *Clin. Gastroenterol. Hepatol.* 11(10):1313-1318 (2013), which is hereby incorporated by reference in its entirety). Epigenetic changes may mark not only the DNA (as methylation or hydroxy-methylation of promoter CpG sites) but also by appending methyl or acetyl groups on the histone proteins that bind to these promoters. These different epigenetic marks may be detected in circulating nucleosomes of colorectal cancer patients (Rahier et al., “Circulating Nucleosomes as New Blood-based Biomarkers for Detection of Colorectal Cancer,” *Clin Epigenetics* 9:53 (2017), which is hereby incorporated by reference in its entirety). The identification of blood-based, cancer-specific methylation markers has employed the entire TCGA Illumina 450K methylation datasets (consisting of primary tumors, matching normal for 33 types of cancer including CRC), along with additional methylation datasets (primary tumors, normal tissues, cell lines, peripheral blood, immune cells) from the Gene Expression Omnibus (GEO). In order to identify CRC-specific methylation markers, comparative statistical analyses of these datasets were used to identify candidate methylation markers with the following characteristics: highly methylated (or hydroxymethylated) in CRC tissues and cell lines, unmethylated in normal colon, unmethylated in peripheral blood and immune infiltrates, unmethylated in most other cancer types. Validating the bioinformatic scheme, these methodologies also identified CpG sites previously reported to be hypermethylated in plasma from CRC patients (Church et al., “Prospective Evaluation of Methylated SEPT9 in Plasma for Detection of Asymptomatic Colorectal Cancer,” *Gut* 63(2):317-325 (2014); Lofton-Day et al., “DNA Methylation Biomarkers for Blood-based Colorectal Cancer Screening,” *Clinical Chemistry* 54(2):414-423 (2008); Toth et al., “Detection of Methylated SEPT9 in Plasma is a Reliable Screening Method for Both Left- and Right-sided Colon Cancers,” *PloS One* 7(9):e46000 (2012); Warren et al., “Septin 9 Methylated DNA is a Sensitive and Specific Blood Test for Colorectal Cancer,” *BMC Medicine* 9:133 (2011); Lange et al., “Genome-scale Discovery of DNA-methylation Biomarkers for Blood-based Detection of Colorectal Cancer,” *PloS One* 7(11):e50266 (2012); Margolin et al., “Robust Detection of DNA Hypermethylation of ZNF154 as a Pan-Cancer Locus with in Silico Modeling for Blood-Based Diagnostic Development,” *The Journal of Molecular Diagnostics : JMD* 18(2):283-298 (2016); Lin et al., “Clinical Relevance of Plasma DNA Methylation in Colorectal Cancer Patients Identified by Using a Genome-Wide High-Resolution Array,” *Ann. Surg. Oncol.* 22 Suppl 3:S1419-1427

(2015); Pedersen et al., “Evaluation of an Assay for Methylated BCAT1 and IKZF1 in Plasma for Detection of Colorectal Neoplasia,” *BMC Cancer* 15:654 (2015), which are hereby incorporated by reference in their entirety). To ensure that these methylation sites were specific to CRC and not a result of aging-related methylation (McClay et al., “A Methylome-wide Study of Aging using Massively Parallel Sequencing of the Methyl-CpG-enriched Genomic Fraction from Blood in over 700 subjects,” *Hum. Mol. Genet.* 23(5):1175-1185 (2014), which is hereby incorporated by reference in its entirety), the Pearson correlation was calculated between levels of methylation and patient age. Furthermore, hypermethylation of these sites did not significantly correlate with MSI status, implying that markers have been identified for all CRC subtypes. Overall, ~ 10,000 tissue samples, > 4 billion datapoints (datapoint = CpG percentage methylation per sample) were analyzed to identify an initial list of few hundred CRC-specific markers. CpG markers consistently show up in many types of cancer and are labeled as potential Pan-Oncology markers. Additional approaches for detecting low abundance 5mC (or 5hmC) are described in WO2016057832A2, which is hereby incorporated by reference in its entirety, or using other suitable means known in the art. Figure 28 provides a list of primary CpG sites that are Colorectal cancer and Colon-tissue specific markers, that may be used to identify the presence of Colorectal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 29 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are Colorectal Cancer and Colon-tissue specific markers, that may be used to identify the presence of colorectal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Primer sets for about 60 of these methylation markers are listed in Table 39 in the prophetic experimental section.

[0254] Mutation or methylation status may give a clear analytical cut-off, *i.e.* the assay either records a mutation or CpG methylation event, and false-positives are a consequence of biology, for example from age-related methylation. With other markers there may be a greater overlap between marker level for individuals with cancer and their matched normal controls, especially in attempting to identify cancer at the earliest stages. In such cases, cut-offs may be defined by “Z-score”, 2 standard deviations above normal values, or by setting the false-positive rate at an arbitrary level, *i.e.* 5% when evaluating a suitable set of age-matched normal samples. Generally, the set of age-matched normal should be suitably large enough to set cut-off of the marker-specific signal from a given disease sample at > 85%; > 90%; > 95%; > 96%; > 97%; or > 98% of the same marker-specific signals from the set of normal samples. The “Z-score” may be calculated using the formula: $Z = (X - \mu)/\sigma$; where Z = Z-score, X = each value in the dataset,

μ = mean of all values in the dataset, and σ = standard deviation of a sample. Likewise, when using the Z-score, the cut-off for marker-specific signal from a given disease sample may be set at a z-score of > 1.03 ; > 1.28 ; > 1.65 ; > 1.75 ; > 1.88 ; or > 2.05 compared to the same marker-specific signals from the set of normal samples. In some assays, marker levels, (*i.e.* DNA methylation levels for several gene promoter regions in plasma, or miRNA levels in urine) are quantified in relation to another marker, either internal or externally added in a qPCR reaction, where the cut-off is determined as a ΔC_t value in the assay (Fackler et al., “Novel Methylated Biomarkers and a Robust Assay to Detect Circulating Tumor DNA in Metastatic Breast Cancer,” *Cancer Res.* 74(8):2160-70 (2014); United States Patent No. 9,416,404 to Sukumar et al., which are hereby incorporated by reference in their entirety). Methylation status at defined promoter regions may also be determined using digital bisulfite genomic sequencing and digital MethyLight assays; using bisulfite conversion and preferential amplification of converted methylated sequences by blocking primers that interfere with amplification of converted unmethylated sequences; or depletion of unmethylated DNA using methyl-sensitive restriction endonucleases, followed by PCR (see United States Patent No. 9,290,803 to Laird et al.; U.S. Patent No. 9,476,100 to Frumkin et al.; U.S. Patent No. 9,765,397 to McEvoy et al.; U.S. Patent No. 9,896,732 to Tabori et al.; U.S. Patent No. 9,957,575 to Kottwitz et al., which are hereby incorporated by reference in their entirety). More recently, an elegant technique for bisulfite-free DNA sequencing has been developed based on using TET2 and APOBEC for conversion of 5mC and 5hmC to DHU (Liu et al., “Bisulfite-Free Direct Detection of 5-methylcytosine and 5-hydroxymethylcytosine at Base Resolution,” *Nat Biotechnol.* 37:424-429 (2019), which is hereby incorporated by reference in its entirety).

[0255] The genome-wide methylation profile of cfDNA (known as the methylome) can be determined using next-generation sequencing, and the methylation pattern may be used to identify the presence of fetal, tumor, or other tissue DNA in the plasma (Sun et al., “Plasma DNA Tissue Mapping by Genome-wide Methylation Sequencing for Noninvasive Prenatal, Cancer, and Transplantation Assessments,” *Proc. Natl. Acad. Sci. U S A* 112(40):E5503-12 (2015); Lehmann-Werman et al., “Identification of Tissue-specific Cell Death Using Methylation Patterns of Circulating DNA,” *Proc. Natl. Acad. Sci. U S A* 113(13):E1826-34 (2016); U.S. Patent No. 9,732,390 to Lo et al.; U.S. Patent No. 9,984,201 to Zhang et al., which are hereby incorporated by reference in their entirety).

[0256] While the above calculations are based on increasing the sensitivity of one or two markers, what if the average sensitivity of individual markers was increased from 50% to 66%? Figures 30 through 32 illustrate results for calculated overall Sensitivity and Specificity for 24,

36, and 48 markers, respectively. These graphs are based on the assumption that the average individual marker sensitivity is 66%, and the average individual marker false-positive rate is from 2% to 5%. The sensitivity curves provide overall sensitivity as a function of the average number of molecules in the blood for each marker, with separate curves for each minimum number of markers needed to call a sample as positive. The specificity curves provide overall specificity as a function of individual marker false-positive rates, again with separate curves for each minimum number of markers needed to call a sample as positive. The calculated numbers for overall Sensitivity and Specificity for 24, 36, and 48 markers, respectively, where the average individual marker sensitivity is 50% (as described previously) or 66% are provided in the tables below.

Table 14.

24 Markers Sensitivity; Avg. Indiv. Mkr,: 50% Sensitivity				
Average Number of Molecules in Blood	Mutation, 1 Positive	24 markers, Minimum 3 Positive	24 markers, Minimum 4 Positive	24 markers, Minimum 5 Positive
150	22.1%	57.7%	35.3%	18.5%
200	28.1%	76.2%	56.7%	37.1%
240	33.0%	85.7%	70.6%	52.4%
300	39.4%	93.8%	84.9%	71.5%
400	48.8%	98.6%	95.8%	90.0%
480	55.1%	99.6%	98.6%	96.2%
600	63.2%	99.9%	99.8%	99.2%

Table 15.

24 Markers Sensitivity; Avg. Indiv. Mkr,: 66% Sensitivity				
Average Number of Molecules in Blood	Mutation, 1 Positive	24 markers, Minimum 3 Positive	24 markers, Minimum 4 Positive	24 markers, Minimum 5 Positive
150	22.1%	76.2%	56.7%	37.1%
200	28.1%	89.8%	77.5%	61.0%
240	33.0%	95.4%	88.1%	76.5%
300	39.4%	98.6%	95.8%	90.0%
400	48.8%	99.8%	99.3%	98.0%
480	55.1%	100.0%	99.9%	99.6%
600	63.2%	100.0%	100.0%	100.0%

Table 16.

24 Marker Specificity			
Individual marker FP rate	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive
2%	98.4%	99.8%	99.9%
3%	94.6%	99.1%	99.9%
4%	87.1%	97.3%	99.6%
5%		93.4%	98.7%

Table 17.

36 Marker Sensitivity; Avg. Indiv. Mkr,: 50% Sensitivity					
Average Number of Molecules in Blood	Mutation, 1 Positive	36 markers, Minimum 3 Positive	36 markers, Minimum 4 Positive	36 markers, Minimum 5 Positive	36 markers, Minimum 6 Positive
150	22.1%	82.6%	65.8%	46.8%	29.7%
200	28.1%	93.8%	84.9%	71.5%	55.4%
240	33.0%	97.5%	92.8%	84.4%	72.4%
300	39.4%	99.4%	97.9%	94.5%	88.4%
400	48.8%	99.9%	99.8%	99.2%	98.0%
480	55.1%	100.0%	100.0%	99.9%	99.6%
600	63.2%	100.0%	100.0%	100.0%	100.0%

Table 18.

36 Marker Sensitivity; Avg. Indiv. Mkr,: 66% Sensitivity					
Average Number of Molecules in Blood	Mutation, 1 Positive	36 markers, Minimum 3 Positive	36 markers, Minimum 4 Positive	36 markers, Minimum 5 Positive	36 markers, Minimum 6 Positive
150	22.1%	93.8%	84.9%	71.5%	55.4%
200	28.1%	98.6%	95.8%	90.0%	80.9%
240	33.0%	99.6%	98.6%	96.2%	91.6%
300	39.4%	99.9%	99.8%	99.2%	98.0%
400	48.8%	100.0%	100.0%	100.0%	99.9%
480	55.1%	100.0%	100.0%	100.0%	100.0%
600	63.2%	100.0%	100.0%	100.0%	100.0%

Table 19.

36 Marker Specificity				
Individual marker FP rate	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive
2%	94.3%	99.1%	99.9%	100.0%
3%	80.7%	95.2%	99.1%	99.9%
4%		84.9%	96.1%	99.2%
5%			88.2%	97.0%

Table 20.

48 Marker Sensitivity; Avg. Indiv. Mkr,: 50% Sensitivity						
Average Number of Molecules in Blood	Mutation, 1 Positive	48 markers, Minimum 4 Positive	48 markers, Minimum 5 Positive	48 markers, Minimum 6 Positive	48 markers, Minimum 7 Positive	48 markers, Minimum 8 Positive
150	22.1%	84.9%	71.6%	55.6%	39.6%	25.8%
200	28.1%	95.8%	90.1%	80.9%	68.7%	54.8%
240	33.0%	99.1%	97.2%	93.4%	87.1%	78.1%
300	39.4%	99.8%	99.3%	98.1%	95.6%	92.3%
400	48.8%	99.9%	99.9%	99.8%	99.7%	99.1%
480	55.1%	99.9%	99.9%	99.9%	99.9%	99.9%
600	63.2%	99.9%	99.9%	99.9%	99.9%	99.9%

Table 21.

48 Marker Sensitivity; Avg. Indiv. Mkr,: 66% Sensitivity						
Average Number of Molecules in Blood	Mutation, 1 Positive	48 markers, Minimum 4 Positive	48 markers, Minimum 5 Positive	48 markers, Minimum 6 Positive	48 markers, Minimum 7 Positive	48 markers, Minimum 8 Positive
150	22.1%	95.8%	90.0%	80.9%	68.7%	54.7%
200	28.1%	99.3%	98.0%	95.2%	90.3%	82.9%
240	33.0%	99.9%	99.6%	98.8%	97.1%	94.0%
300	39.4%	100.0%	100.0%	99.9%	99.6%	99.0%
400	48.8%	100.0%	100.0%	100.0%	100.0%	100.0%
480	55.1%	100.0%	100.0%	100.0%	100.0%	100.0%
600	63.2%	100.0%	100.0%	100.0%	100.0%	100.0%

Table 22.

48 Marker Specificity					
Individual marker FP rate	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive
2%	96.9%	99.4%	99.9%	99.9%	99.9%
3%	84.3%	95.8%	99.1%	99.8%	99.9%
4%		82.5%	95.0%	98.8%	99.8%
5%				94.3%	98.6%

[0257] The above tables, and Figures 30 through 32, as well as Figures 18 through 20, allow for a direct comparison in the overall improvement in sensitivity when the average individual marker sensitivity improves from 50% to 66%. In this example, if there is an average of 150 molecules in the blood for the earliest cancer (Stage I), and if that would cover at least one mutation, then the sensitivity for identifying such a cancer by next generation sequencing would be 22.1% (See any of the aforementioned figures). For 24 markers, with a minimum of 3 markers positive and a 3% FP rate, overall sensitivity improves from 57.7% to 76.2%, when the average individual marker sensitivity improves from 50% to 66%, for detecting Stage I cancer (at about 150 molecules of each positive marker in the blood, see Figures 18A and 30A). If the individual marker FP rate is 3%, then if there is a 3-marker minimum, then overall FP rate is 5.4% for 24 markers, for a specificity of 94.6% (See Figures 18B or 30B). However, if the individual marker FP rate is 5%, then if there is a 4-marker minimum, then overall FP rate is 6.6% for 24 markers, for a specificity of 93.4% (See Figure 18B). At 4 markers, for Stage I cancer (at about 150 molecules of each positive marker in the blood), overall sensitivity improves from 35.3% to 56.7%, when the average individual marker sensitivity improves from 50% to 66% (See Figure 18A, and Figure 30A). For 36 markers, with a minimum of 3 markers positive and a 2% FP rate, overall sensitivity improves from 82.6% to 93.8%, when the average individual marker sensitivity improves from 50% to 66%, for detecting Stage I cancer (at about 150 molecules of each positive marker in the blood, see Figures 19A and 31A). If the individual marker FP rate is 2%, then if there is a 3-marker minimum, then overall FP rate is 5.7% for 36 markers, for a specificity of 94.3% (See Figure 19B or 31B). However, if the individual marker FP rate is 3%, then the assay requires a 4-marker minimum, then overall FP rate is 4.8% for 36 markers, for a specificity of 95.2% (See Figure 19B). At 4 markers, for Stage I cancer (at about 150 molecules of each positive marker in the blood), overall sensitivity improves from 65.8% to 84.9%, when the average individual marker sensitivity improves from 50% to 66% (See Figure 19A and Figure 31A). For 48 markers, with a minimum of 4 markers positive and a 2% FP rate,

overall sensitivity improves from 84.9% to 95.8%, when the average individual marker sensitivity improves from 50% to 66%, for detecting Stage I cancer (at about 150 molecules of each positive marker in the blood, see Figures 20A and 32A). If the individual marker FP rate is 2%, then if there is a 4-marker minimum, then overall FP rate is 3.1% for 48 markers, for a specificity of 96.9% (See Figures 20B or 32B). However, if the individual marker FP rate is 3%, then the assay requires a 5-marker minimum, then overall FP rate is 4.2% for 48 markers, for a specificity of 95.8% (See Figure 20B). At 5 markers, for Stage I cancer (at about 150 molecules of each positive marker in the blood), overall sensitivity improves from 71.6% to 90.0%, when the average individual marker sensitivity improves from 50% to 66% (See Figure 20A and Figure 32A).

[0258] From the above charts, the receiver operating characteristic (ROC) curves may be calculated by plotting Sensitivity vs. 1-Specificity. Since these are theoretical calculations, the curves were generated for different levels of average marker false-positive rates of 2%, 3%, 4%, and 5%. The AUC values, calculated for ROC curves for 24 markers, with average individual marker at 66% Sensitivity with 2%-3% FP; 36 markers, with average individual marker at 66% Sensitivity with 2%-3% FP; and 48 markers, with average individual marker at 66% Sensitivity with 2%-3% FP; are provided in Table 23 below. Using the benchmark of an average of 150 molecules in the blood for the earliest cancer (Stage I), and looking only at the 3% individual marker FP rate AUC values are at 77% with 24 markers (average individual marker at 50% Sensitivity), improve to 87% with 24 markers (average individual marker at 66% Sensitivity); AUC values are at 87% with 36 markers (average individual marker at 50% Sensitivity), improve to 95% with 36 markers (average individual marker at 66% Sensitivity); and AUC values are at 89% with 48 markers (average individual marker at 50% Sensitivity), improve to 97% with 48 markers (average individual marker at 66% Sensitivity). These results illustrate that for multiple marker assays achieving good sensitivities and specificities for the earliest cancers is aided when the average individual marker sensitivity improves from 50% to 66%.

Table 23.

24, 36, & 48 Marker AUC Values from ROC Curves; Avg. Indiv. Mkr,: 66% Sensitivity							
Total Markers: Individual marker FP rate	150 Molecules	200 Molecules	240 Molecules	300 Molecules	400 Molecules	480 Molecules	600 Molecules
24 Mkrs: 2%	88%	95%	98%	>99%	>99%	>99%	>99%
24 Mkrs: 3%	87%	94%	97%	99%	>99%	>99%	>99%
36 Mkrs: 2%	96%	99%	>99%	>99%	>99%	>99%	>99%
36 Mkrs: 3%	95%	99%	>99%	>99%	>99%	>99%	>99%
48 Mkrs: 2%	98%	>99%	>99%	>99%	>99%	>99%	>99%
48 Mkrs: 3%	97%	99%	>99%	>99%	>99%	>99%	>99%

[0259] How would increasing the average individual marker sensitivity from 50% sensitivity to 66% sensitivity improve upon a one-step cancer assay? To review: the challenge is to screen 107 million adults in the U.S. over the age of 50 for colorectal cancer – of which there are about 135,000 new cases that are diagnosed a year. In this example, if there is an average of 300 molecules in the blood for early cancer (Stage I & II), and taking the best-case scenario of individual marker FP rate is 2%, then if there is a 3-marker minimum, then overall FP rate is 1.6% for 24 markers, for a specificity of 98.4% (See Figure 18B or 30B). At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), for average marker sensitivity of 50%, the test would miss 6.2%; *i.e.* for Stage I & II cancer the overall sensitivity would be 93.8% (See Figure 18A), *e.g.* the test would correctly identify 93.8% of individuals with disease, which would be 126,630 individuals (out of 135,000 new cases). At a specificity of 98.4%, for 107 million individuals screened, the test would also generate $1.6\% \times 107,000,000 = 1,712,000$ false positives. Thus, the positive predictive value would be $126,630 / (126,630 + 1,712,000) = \text{around } 6.8\%$, in other words, only one in 14 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), for average marker sensitivity of 66%, the test would miss 1.4%; *i.e.* for Stage I & II cancer the overall sensitivity would be 98.6% (See Figure 30A), *e.g.* the test would correctly identify 98.6% of individuals with disease, which would be 133,110 individuals (out of 135,000 new cases). At a specificity of 98.4%, for 107 million individuals screened, the test would also generate $1.6\% \times 107,000,000 = 1,712,000$ false positives. Thus, the positive predictive value would be $133,110 / (133,110 + 1,712,000) = \text{around } 7.2\%$, in other words, only one in 14 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. Thus, if

the FP is low, *i.e.* 2%, then there is marginal benefit in going from an average marker sensitivity of 50% to an average marker sensitivity of 66%.

[0260] However, if the individual marker FP rate is more realistic, say 4%, then more markers will be required to achieve better than 98% specificity, and this will be at the cost of sensitivity. If individual marker FP rate is 4%, then if there is a 5-marker minimum, then overall FP rate is 0.4% for 24 markers, for a specificity of 99.6% (See Figure 18B). At 5 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), at an average marker sensitivity of 50%, the test would miss 28.5%; *i.e.* for Stage I & II cancer the overall sensitivity would be 71.5% (See Figure 18A), *e.g.* the test would correctly identify 71.5% of individuals with disease, which would be 90,540 individuals (out of 135,000 new cases). At a specificity of 99.6%, for 107 million individuals screened, the test would also generate $0.4\% \times 107,000,000 = 428,000$ false positives. Thus, the positive predictive value would be $90,540 / (90,540 + 428,000) = \text{around } 17.5\%$, in other words, one in 5.7 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. A PPV of 17.5% is quite respectable, however, it would be achieved at the cost of missing 28.5% of early cancer. At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), for average marker sensitivity of 66%, the test would miss 10.0%; *i.e.* for Stage I & II cancer the overall sensitivity would be 90.0% (See Figure 30A), *e.g.* the test would correctly identify 90.0% of individuals with disease, which would be 121,500 individuals (out of 135,000 new cases). At a specificity of 99.6%, for 107 million individuals screened, the test would also generate $0.4\% \times 107,000,000 = 428,000$ false positives. Thus, the positive predictive value would be $121,500 / (121,500 + 428,000) = \text{around } 22.1\%$, in other words, one in 4.5 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. A PPV of 22.1% is excellent, and further, it would be achieved at the cost of missing only 10% of early cancer. Thus, if the FP is more realistic *i.e.* 4%, then there is a significant benefit in going from an average marker sensitivity of 50% to an average marker sensitivity of 66%.

[0261] Returning to the example of colorectal cancer, in particular the cases of microsatellite stable tumors (MSS) where the mutation load is low, for these calculations when relying on NGS sequencing alone (assuming 150 molecules with one mutation in the blood), an estimated 78% of early colorectal cancer would be missed. Again, to put these number in perspective, in the U.S., about 135,000 new cases of colorectal cancer were diagnosed in 2018, of which about 60% is late cancer (*i.e.* Stage III & IV). About 107 million individuals in the U.S. are over the age of 50 and should be tested for colorectal cancer. While it cannot be predicted how many individuals have a hidden cancer (*i.e.* Stage I) within them, who are non-

compliant to testing, for the purposes of this calculation, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. With the assumption of these samples containing at least 150 molecules with one mutation in the blood, such a test would find 8,910 individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. However, with a specificity for sequencing at 98%, there would be about 2.1 million false-positives. The positive predictive value of such a test would be around 0.4%, in other words, only one in 236 individuals who tested positive would actually have Stage I colorectal cancer, the rest would be false-positives. In contrast, consider the two-step methylation marker test described above, wherein the first step has 24 methylation markers specific to GI cancers, while the second step has 48 methylation markers specific to colorectal cancer. In this example, the average individual marker sensitivity is set at 66%. In this example, as above, the calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. Assuming individual marker false-positive rates of 3%, and with the first step requiring a minimum of 3 markers positive, then with an overall specificity of 94.6%, the first step would identify 5,778,000 individuals (out of 107,000,000 total adults over 50 in the U.S.) which would include at 76.2% sensitivity or about 30,861 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 5,778,000 presumptive positive individuals would be evaluated in the second step of 48 markers requiring a minimum of 5 markers positive, then the two-step test would identify $76.2\% \times 90.0\% = 68.6\% = 27,775$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $5,778,000 \times 4.2\% = 242,676$ false-positives. The positive predictive value of such a test would be $27,775/270,451 = 10.3\%$, in other words, 1 in 10 individuals who tested positive would actually have Stage I colorectal cancer, an extraordinarily successful screen to focus on those patients who would most benefit from follow-up colonoscopy. Since >90% of individuals identified with Stage I colon cancer have long-term survival after just surgery, the benefit in lives saved would be of incalculable value.

[0262] How would the above numbers change if the initial test in the two-step assay uses 36 markers? In this example, as above, the calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. Assuming individual marker false-positive rates of 3%, and with the first step requiring a minimum of 4 markers positive, then with an overall specificity of 95.2%, the first step would identify 5,136,000 individuals (out of 107,000,000 total adults over 50 in the U.S.) which would include at 84.9% sensitivity or about 34,385 individuals with Stage I colorectal cancer (out of

40,500 individuals with Stage I cancer). However, those 5,136,000 presumptive positive individuals would be evaluated in the second step of 48 markers requiring a minimum of 5 markers positive, then the two-step test would identify $84.9\% \times 90.0\% = 76.4\% = 30,946$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $5,136,000 \times 4.2\% = 215,712$ false-positives. The positive predictive value of such a test would be $30,946/246,658 = 12.5\%$, in other words, 1 in 8 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker, and the higher stages. Also, to be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals with Stage I cancer, 40,500 individuals with Stage II cancer, and the remaining 54,000 individuals have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 95.8% would be identified in the first step, of which $95.8\% \times 98.0\% = 93.9\% = 38,023$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.8% would be identified in the first step, of which $99.8\% \times (100\%) = 53,892$ individuals with late cancer would be identified. This brings the total identified at $30,946 + 38,023 + 53,892 = 122,861$ individuals out of 135,000 with colorectal cancer. Overall, the positive predictive value of such a test would be $122,861/369,519 = 33.2\%$, in other words, 1 in 3 individuals who tested positive would actually have colorectal cancer, and this test would identify $68,969/81,000$ or 85% of those individuals with early cancer – which would be unprecedented in diagnostic approaches to detect early cancer.

[0263] What if the goal is to minimize the total number of markers in an initial high-throughput cancer screen? What if the average sensitivity of individual markers was increased from 66% to 75%? Figures 33 through 38 illustrate results for calculated overall Sensitivity and Specificity for 12, 18, 24, 32, 36, and 48 markers, respectively. These graphs are based on the assumption that the average individual marker sensitivity is 75%, and the average individual marker false-positive rate is from 2% to 5%. The sensitivity curves provide overall sensitivity as a function of the average number of molecules in the blood for each marker, with separate curves for each minimum number of markers needed to call a sample as positive. The specificity curves

provide overall specificity as a function of individual marker false-positive rates, again with separate curves for each minimum number of markers needed to call a sample as positive. The calculated numbers for overall Sensitivity and Specificity for 12, 18, 24, 32, 36, and 48 markers, respectively, where the average individual marker sensitivity is 75% are provided in the tables below.

Table 24.

12 Markers Sensitivity Avg. Indiv. Mkr,: 75% Sensitivity				
Average Number of Molecules in Blood	Mutation, 1 Positive	Minimum 2 Markers Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive
150	22.1%	65.8%	39.1%	19.1%
200	28.1%	80.1%	57.7%	35.3%
240	33.0%	87.4%	69.7%	48.5%
300	39.4%	93.9%	82.6%	65.8%
400	48.8%	98.3%	93.8%	84.9%
480	55.1%	99.4%	97.5%	92.8%
600	63.2%	99.9%	99.4%	97.9%

Table 25.

12 Marker Specificity			
Individual marker FP rate	Minimum 2 Markers Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive
2%	97.4%	99.8%	100.0%
3%	94.1%	99.4%	100.0%
4%	89.4%	98.6%	99.9%
5%	83.5%	97.3%	99.7%

Table 26.

18 Marker Sensitivity Avg. Indiv. Mkr.: 75% Sensitivity					
Average Number of Molecules in Blood	Mutation, 1 Positive	Minimum 2 Markers Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive
150	22.1%	85.0%	65.5%	43.6%	25.1%
200	28.1%	93.9%	82.6%	65.8%	46.8%
240	33.0%	97.1%	90.5%	78.7%	62.7%
300	39.4%	99.1%	96.4%	90.4%	80.3%
400	48.8%	99.9%	99.4%	97.9%	94.5%
480	55.1%	99.9%	99.9%	99.4%	98.3%
600	63.2%	99.9%	99.9%	99.9%	99.7%

Table 27.

18 Marker Specificity				
Individual marker FP rate	Minimum 2 Markers Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive
2%	93.9%	99.3%	100.0%	100.0%
3%	86.2%	97.8%	99.8%	100.0%
4%	75.5%	94.8%	99.2%	99.9%
5%	61.8%	89.8%	98.1%	99.7%

Table 28.

24 Marker Sensitivity Avg. Indiv. Mkr.: 75% Sensitivity					
Average Number of Molecules in Blood	Mutation, 1 Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive
150	22.1%	82.6%	65.8%	46.8%	29.7%
200	28.1%	93.8%	84.9%	71.5%	55.4%
240	33.0%	97.5%	92.8%	84.4%	72.4%
300	39.4%	99.4%	97.9%	94.5%	88.4%
400	48.8%	99.9%	99.8%	99.2%	98.0%
480	55.1%	99.9%	99.9%	99.9%	99.6%
600	63.2%	99.9%	99.9%	99.9%	99.9%

Table 29.

24 Marker Specificity				
Individual marker FP rate	Minimum 2 Markers Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive
2%	89.0%	98.4%	99.8%	100.0%
3%	75.2%	94.5%	99.1%	99.9%
4%	55.8%	87.0%	97.3%	99.6%
5%	31.0%	74.7%	93.4%	98.7%

Table 30.

32 Marker Sensitivity Avg. Indiv. Mkr.; 75% Sensitivity							
Average Number of Molecules in Blood	Mutation, 1 Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive
150	22.1%	93.8%	84.9%	71.5%	55.4%	39.4%	25.6%
200	28.1%	98.6%	95.8%	90.0%	80.9%	68.7%	54.7%
240	33.0%	99.6%	98.6%	96.2%	91.6%	84.3%	74.2%
300	39.4%	99.9%	99.8%	99.2%	98.0%	95.4%	91.0%
400	48.8%	99.9%	99.9%	99.9%	99.9%	99.6%	99.0%
480	55.1%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
600	63.2%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%

Table 31.

32 Marker Specificity						
Individual marker FP rate	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive
2%	96.0%	99.4%	99.9%	100.0%	100.0%	100.0%
3%	86.6%	97.1%	99.5%	99.9%	100.0%	100.0%
4%	68.3%	90.8%	97.9%	99.6%	99.9%	100.0%
5%	38.0%	77.5%	93.7%	98.6%	99.7%	100.0%

Table 32.

36 Marker Sensitivity Avg. Indiv. Mkr.; 75% Sensitivity							
Average Number of Molecules in Blood	Mutation, 1 Positive	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive
150	22.1%	96.4%	90.4%	80.3%	66.6%	51.2%	36.4%
200	28.1%	99.4%	97.9%	94.5%	88.4%	79.3%	67.6%
240	33.0%	99.9%	99.4%	98.3%	95.8%	91.3%	84.3%
300	39.4%	99.9%	99.9%	99.7%	99.2%	98.1%	95.9%
400	48.8%	99.9%	99.9%	99.9%	99.9%	99.9%	99.7%
480	55.1%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
600	63.2%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%

Table 33.

36 Marker Specificity						
Individual marker FP rate	Minimum 3 Markers Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive
2%	94.3%	99.1%	99.9%	100.0%	100.0%	100.0%
3%	80.7%	95.2%	99.1%	99.9%	100.0%	100.0%
4%	54.3%	84.9%	96.1%	99.2%	99.9%	100.0%
5%	10.8%	63.2%	88.2%	97.0%	99.3%	99.9%

Table 34.

48 Marker Sensitivity Avg. Individ. Mkr.: 75% Sensitivity								
Average Number of Molecules in Blood	Mutation, 1 Positive	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive	Minimum 9 Markers Positive	Minimum 10 Markers Positive
150	22.1%	97.9%	94.5%	88.4%	79.3%	67.6%	54.4%	41.3%
200	28.1%	99.8%	99.2%	98.0%	95.4%	91.0%	84.5%	75.8%
240	33.0%	99.9%	99.9%	99.6%	98.9%	97.5%	94.9%	90.8%
300	39.4%	99.9%	99.9%	99.9%	99.9%	99.7%	99.3%	98.5%
400	48.8%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
480	55.1%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%
600	63.2%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%	99.9%

Table 35.

48 Marker Specificity							
Individual marker FP rate	Minimum 4 Markers Positive	Minimum 5 Markers Positive	Minimum 6 Markers Positive	Minimum 7 Markers Positive	Minimum 8 Markers Positive	Minimum 9 Markers Positive	Minimum 10 Markers Positive
2%	96.9%	99.5%	99.9%	100.0%	100.0%	100.0%	100.0%
3%	84.2%	95.8%	99.1%	99.8%	100.0%	100.0%	100.0%
4%	50.2%	82.5%	95.0%	98.8%	99.8%	100.0%	100.0%
5%		46.5%	80.8%	94.2%	98.5%	99.7%	99.9%

[0264] How would increasing the average individual marker sensitivity from 50% sensitivity to 75% sensitivity improve upon a one-step cancer assay? To review: the challenge is to screen 107 million adults in the U.S. over the age of 50 for colorectal cancer – of which there are about 135,000 new cases that are diagnosed a year. In this example, if there is an average of

300 molecules in the blood for early cancer (Stage I & II), and taking the best-case scenario of individual marker FP rate is 2%, then if there is a 3-marker minimum, then overall FP rate is 1.6% for 24 markers, for a specificity of 98.4% (See Figure 18B or 30B). At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), for average marker sensitivity of 50%, the test would miss 6.2%; *i.e.* for Stage I & II cancer the overall sensitivity would be 93.8% (See Figure 18A), *e.g.* the test would correctly identify 93.8% of individuals with disease, which would be 126,630 individuals (out of 135,000 new cases). At a specificity of 98.4%, for 107 million individuals screened, the test would also generate $1.6\% \times 107,000,000 = 1,712,000$ false positives. Thus, the positive predictive value would be $126,630 / (126,630 + 1,712,000) = \text{around } 6.8\%$. In other words, only one in 14 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. At 3 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), with only 18 markers with an average marker sensitivity of 75%, the test would miss 3.6%; *i.e.* for Stage I & II cancer the overall sensitivity would be 96.4% (See Figure 35A), *e.g.* the test would correctly identify 96.4% of individuals with disease, which would be 130,140 individuals (out of 135,000 new cases). At a specificity of 99.3%, for 107 million individuals screened, the test would also generate $0.7\% \times 107,000,000 = 749,000$ false positives. Thus, the positive predictive value would be $130,140 / (130,140 + 749,000) = \text{around } 14.8\%$. In other words, only one in 6.7 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. Thus, if the FP is low, *i.e.* 2%, then there is some benefit in going from an average marker sensitivity of 50% to an average marker sensitivity of 75%.

[0265] However, if the individual marker FP rate is more realistic, say 4%, then more markers will be required to achieve better than 98% specificity, and this will be at the cost of sensitivity. If individual marker FP rate is 4%, then if there is a 5-marker minimum, then overall FP rate is 0.4% for 24 markers, for a specificity of 99.6% (See Figure 18B). At 5 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), at an average marker sensitivity of 50%, the test would miss 28.5%; *i.e.* for Stage I & II cancer the overall sensitivity would be 71.5% (See Figure 18A), *e.g.* the test would correctly identify 71.5% of individuals with disease, which would be 90,540 individuals (out of 135,000 new cases). At a specificity of 99.6%, for 107 million individuals screened, the test would also generate $0.4\% \times 107,000,000 = 428,000$ false positives. Thus, the positive predictive value would be $90,540 / (90,540 + 428,000) = \text{around } 17.5\%$. In other words, one in 5.7 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. A PPV of 17.5% is quite respectable, however, it would be achieved at the cost of missing 28.5% of early cancer.

At 4 markers, for Stage I & II cancer (at about 300 molecules of each positive marker in the blood), for an 18 marker panel with average marker sensitivity of 75%, the test would miss 9.6%; i.e. for Stage I & II cancer the overall sensitivity would be 90.4% (See Figure 35A), *e.g.* the test would correctly identify 90.4% of individuals with disease, which would be 122,040 individuals (out of 135,000 new cases). At a specificity of 99.2%, for 107 million individuals screened, the test would also generate $0.8\% \times 107,000,000 = 856,000$ false positives. Thus, the positive predictive value would be $122,040 / (122,040 + 856,000) = \text{around } 12.5\%$. In other words, one in 8 individuals who tested positive would actually have colorectal cancer, the rest would be false-positives. A PPV of 12.5% is not bad, and further, it would be achieved at the cost of missing only 10% of early cancer. Thus, if the FP is more realistic *i.e.* 4%, then there is a significant benefit in going from an average marker sensitivity of 50% to an average marker sensitivity of 75%.

[0266] Returning to the example of colorectal cancer, in particular the cases of microsatellite stable tumors (MSS) where the mutation load is low, for these calculations when relying on NGS sequencing alone (assuming 150 molecules with one mutation in the blood), an estimated 78% of early colorectal cancer would be missed. Again, to put these numbers in perspective, in the U.S., about 135,000 new cases of colorectal cancer were diagnosed in 2018, of which about 60% is late cancer (*i.e.* Stage III & IV). About 107 million individuals in the U.S. are over the age of 50 and should be tested for colorectal cancer. While it cannot be predicted how many individuals have a hidden cancer (*i.e.* Stage I) within them, who are non-compliant to testing, for the purposes of this calculation, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. With the assumption of these samples containing at least 150 molecules with one mutation in the blood, such a test would find 8,910 individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. However, with a specificity for sequencing at 98%, there would be about 2.1 million false-positives. The positive predictive value of such a test would be around 0.4%, in other words, only one in 236 individuals who tested positive would actually have Stage I colorectal cancer, the rest would be false-positives. In contrast, consider a two-step methylation marker test, wherein the first step has 18 methylation markers with average sensitivity of 75%, specific to GI cancers, while the second step has 36 methylation markers with average sensitivity of 75% specific to colorectal cancer (See Figure 1D). In this example, as above, the calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. Assuming individual marker false-positive rates of 3%, and with the first step requiring a minimum of 3 markers positive,

then with an overall specificity of 97.8%, the first step would identify 2,354,000 individuals (out of 107,000,000 total adults over 50 in the U.S.) which would include, at 65.5% sensitivity, about 26,527 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 2,354,000 presumptive positive individuals would be evaluated in the second step of 36 markers requiring a minimum of 5 markers positive, then the two-step test would identify $65.5\% \times 80.3\% = 52.6\% = 21,302$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 99.1%, the second test would also generate $2,354,000 \times 0.9\% = 21,186$ false-positives. The positive predictive value of such a test would be $21,302 / (21,302 + 21,186) = 50.1\%$. In other words, 1 in 2 individuals who tested positive would actually have Stage I colorectal cancer, an extraordinarily successful screen to focus on those patients who would most benefit from follow-up colonoscopy. Since >90% of individuals identified with Stage I colon cancer have long-term survival after just surgery, the benefit in lives saved would be of incalculable value.

[0267] The ultimate goal is to develop a high-throughput scalable test to detect the majority of cancers that occur worldwide. The solid tumor cancers have been grouped into the following subclasses, as listed below in Tables 36, 37, and 38 for both sexes, for men, and for women.

Table 36

Global cancer incidence; Both Sexes				
(Numbers in thousands; most common cancers have incidence above 100,000 per year)				
	Incidence	% Group	% total All	(Total)
Group 1:				
Colorectal (1,801)	1801	52.9%	12.9%	13981
Stomach (1,033)	1033	30.3%	7.4%	13981
Esophagus (572)	572	16.8%	4.1%	13981
Total, Group 1:	3406			
Group 2:				
Breast (2,089)	2089	62.6%	14.9%	13981
Endometrial & Cervical (570)	570	17.1%	4.1%	13981
Uterine (382)	382	11.5%	2.7%	13981
Ovarian (295)	295	8.8%	2.1%	13981
Total, Group 2:	3336			
Group 3:				
Lung (2,093)	2093	59.9%	15.0%	13981
Head & Neck (832)	832	23.8%	6.0%	13981
Thyroid (567)	567	16.2%	4.1%	13981
Total, Group 3:	3492			
Group 4:				
Prostate (1,276)	1276	57.3%	9.1%	13981
Bladder (549)	549	24.6%	3.9%	13981
Kidney (403)	403	18.1%	2.9%	13981
Total, Group 4:	2228			
Group 5:				
Liver (841)	841	55.4%	6.0%	13981
Pancreas (459)	459	30.2%	3.3%	13981
Gallbladder (219)	219	14.4%	1.6%	13981
Total, Group 5:	1519			
Total	13981			

Table 37

Global cancer incidence; Male				
(Numbers in thousands; most common cancers have incidence above 100,000 per year)				
	Incidence	% Group	% total All	(Total)
Group 1:				
Colorectal (1,801)	1006	48.2%	14.1%	7114
Stomach (1,033)	683	32.7%	9.6%	7114
Esophagus (572)	400	19.1%	5.6%	7114
Total, Group 1:	2089			
Group 2:				
Breast (2,089)				
Endometrial & Cervical (570)				
Uterine (382)				
Ovarian (295)				
Total, Group 2:				
Group 3:				
Lung (2,093)	1368	64.1%	19.2%	7114
Head & Neck (832)	635	29.8%	8.9%	7114
Thyroid (567)	131	6.1%	1.8%	7114
Total, Group 3:	2134			
Group 4:				
Prostate (1,276)	1276	65.3%	17.9%	7114
Bladder (549)	424	21.7%	6.0%	7114
Kidney (403)	254	13.0%	3.6%	7114
Total, Group 4:	1954			
Group 5:				
Liver (841)	597	63.7%	8.4%	7114
Pancreas (459)	243	25.9%	3.4%	7114
Gallbladder (219)	97	10.4%	1.4%	7114
Total, Group 5:	937			
Total	7114			

Table 38

Global cancer incidence; Female				
(Numbers in thousands; most common cancers have incidence above 100,000 per year)				
	Incidence	% Group	% total All	(Total)
Group 1:				
Colorectal (1,801)	795	60.4%	11.5%	6930
Stomach (1,033)	350	26.6%	5.1%	6930
Esophagus (572)	172	13.1%	2.5%	6930
Total, Group 1:	1317			
Group 2:				
Breast (2,089)	2089	62.6%	30.1%	6930
Endometrial & Cervical (570)	570	17.1%	8.2%	6930
Uterine (382)	382	11.5%	5.5%	6930
Ovarian (295)	295	8.8%	4.3%	6930
Total, Group 2:	3336			
Group 3:				
Lung (2,093)	725	53.4%	10.5%	6930
Head & Neck (832)	196	14.4%	2.8%	6930
Thyroid (567)	436	32.1%	6.3%	6930
Total, Group 3:	1357			
Group 4:				
Prostate (1,276)	0	0.0%	0.0%	6930
Bladder (549)	216	63.9%	3.1%	6930
Kidney (403)	122	36.1%	1.8%	6930
Total, Group 4:	338			
Group 5:				
Liver (841)	244	41.9%	3.5%	6930
Pancreas (459)	216	37.1%	3.1%	6930
Gallbladder (219)	122	21.0%	1.8%	6930
Total, Group 5:	582			
Total	6930			

[0268] The above list does not include liquid cancers, nor some of the less common solid tumors. Worldwide incidence (numbers in thousands) of liquid tumors include Non-Hodgkin Lymphoma (225), Leukemia (187), Multiple Myeloma (70), and Hodgkin lymphoma (33). These would be detected in a separate test not discussed herein. Further, the list excludes

Melanoma (287) and Brain tumors (134). Testing for these would be done with separate sets of markers, optimized as described above for colorectal cancer. In addition, while some cancers listed in the tables above are of extreme medical importance (Mesothelioma, Thyroid, the three different subcategories of Kidney cancer), their biology is sufficiently different as to usually merit a separate set of markers, again, optimized as described above for colorectal cancer.

[0269] Thus, for the present application, a Pan-Oncology test was developed to include the following major cancers by the following groupings: Group 1 (colorectal, stomach, and esophagus); Group 2 (breast, endometrial, ovarian, cervical, and uterine); Group 3 (lung adenoma, lung small cell, and head & neck); Group 4 (prostate and bladder); and Group 5 (liver, pancreatic, or gall bladder). that some cancers within Group 3 may be tested as a sputum sample, and while cancers in Group 4 may be tested as a urine sample.

[0270] Careful analysis of the TCGA methylation database revealed a general commonality in methylation patterns among cancers within these 5 separate groups. Further, there are some methylation markers that are common among several cancers, while absent in normal white blood cells. Three different strategies were used to design a multi-step pan-oncology test.

[0271] The first strategy is to identify markers that cover multiple cancers in one or more of the above groups. The markers should be sufficiently diverse as to cover cancers in all 5 groups. For example, a first step of the assay would use a set of 96 markers that on average comprise of at least 36 markers with 50% sensitivity that covers each of the aforementioned 16 types of solid tumors (covered in the 5 Groups; See Figure 1E; for 66% sensitivity, See Figure 1I). If at least 5 markers are positive, the assay would then move to a second step that would be used to verify the initial results and identify the most probable tissue of origin. In most cases, more than 5 markers would be positive, and then pattern of distribution of these methylation markers would guide the choice of which groups to test in the second step. The second step of the assay would test on average 2 or more sets of the group-specific markers. For example, the second step of the assay would use 2 or more sets of 64 group-specific markers that on average comprise of at least 36 markers with 50% sensitivity that covers each of the aforementioned types of solid tumors that may be present in that group (for 66% sensitivity, see Figure 1I). By scoring the markers that are positive and comparing to predicted positives for each cancer type within the group tested, the physician can identify the most probable tissue of origin, and subsequently send the patient to the appropriate imaging.

[0272] The second strategy is to identify markers that cover multiple cancers in one or more of the above groups. The markers should be sufficiently diverse as to cover cancers in all 5

groups. As before, a first step of the assay would use a set of 96 markers that on average comprise of at least 36 markers with 50% sensitivity that covers each of the aforementioned 16 types of solid tumors (covered in the 5 Groups; see Figure 1F; for 66% sensitivity, see Figure 1J). If at least 5 markers are positive, the assay would then move to a second step that would be used to verify the initial results and identify the most probable tissue of origin. In most cases, more than 5 markers would be positive, and then pattern of distribution of these methylation markers would guide the choice of which groups to test in the second step. The second step of the assay would test on average 2 or more sets of the group-specific markers. For example, the second step of the assay would use 2 or more sets of 48 group-specific markers that on average comprise of at least 36 markers with 75% sensitivity that covers each of the aforementioned types of solid tumors that may be present in that group. By scoring the markers that are positive and comparing to predicted positives for each cancer type within the group tested, the physician can most probably verify the group, and probably the tissue of origin, and then subsequently send the patient to the appropriate imaging.

[0273] The third strategy is to identify markers that cover as many cancers as possible, irrespective of group. The markers should be sufficiently diverse as to cover cancers in all 5 groups. For example, a first step of the assay would use a set of 48 markers that on average comprise at least 24 markers with 75% sensitivity that covers each of the aforementioned 16 types of solid tumors (covered in the 5 Groups; see Figure 1G). For even more sensitive detection of early cancer, the first step of the assay would use a set of 64 markers that on average comprise at least 36 markers with 75% sensitivity that covers each of the aforementioned 16 types of solid tumors (covered in the 5 Groups; see Figure 1H). Since these markers would be broadly found in many cancers, the resultant positive markers may not point to which groups to evaluate in a second step to identify the most probable tissue of origin. One approach to do so would be to continue with the first strategy, *i.e.* use the 96-marker set that on average comprise of at least 36 markers with 50% sensitivity for each tumor type to determine the most probable tissue of origin (for 66% sensitivity, see Figures 1K & 1L). Another approach would be to use an alternative technology to identify tissue of origin, such as targeted bisulfite sequencing of 96 or more regions to determine methylation patterns and compare with predicted methylation pattern of different cancer types followed by the appropriate imaging.

[0274] Returning to the first strategy (see Figure 1E), a close evaluation of the TCGA database reveals pan-oncology markers that meet the criteria for use in a set of 96 markers that on average comprise at least 36 markers with 50% sensitivity that covers each of the aforementioned 16 types of solid tumors. These pan-oncology markers include, but are not

limited to, cancer-specific microRNA markers, cancer-specific ncRNA and lncRNA markers, cancer-specific exon transcripts, tumor-associated antigens, cancer protein markers, protein markers that can be secreted by solid tumors into the blood, common mutations, primary CpG sites that are solid tumor and tissue specific markers, chromosomal regions or sub-regions within which are primary CpG sites that are solid tumor and tissue specific markers, and primary and flanking CpG sites that are solid tumor and tissue specific markers.

[0275] Methods for detecting said markers have been discussed earlier in this application, and these markers are listed below and in accompanying figures.

[0276] Blood-based, solid tumor-specific microRNA markers derived through analysis of TCGA microRNA datasets, includes, but is not limited to the following markers: (mir ID , Gene ID); hsa-mir-21, MIR21; hsa-mir-182, MIR182; hsa-mir-454, MIR454; hsa-mir-96, MIR96; hsa-mir-183, MIR183; hsa-mir-549, MIR549; hsa-mir-301^a, MIR301A; hsa-mir-548f-1, MIR548F1; hsa-mir-301b, MIR301B; hsa-mir-103-1, MIR1031; hsa-mir-18^a, MIR18A; hsa-mir-147b , MIR147B; hsa-mir-4326, MIR4326; hsa-mir-573, MIR573. These markers may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood.

[0277] Figure 39 provides a list of blood-based, solid tumor-specific ncRNA and lncRNA markers derived through analysis of various publicly available Affymetrix Exon ST CEL data, which were aligned to GENCODE annotations to generate ncRNA and lncRNA transcriptome datasets. Comparative analyses across these datasets (various cancer types, along with normal tissues, and peripheral blood) were conducted to generate the ncRNA and lncRNA markers list. Such lncRNA and ncRNA may be enriched in exosomes or other protected states in the blood.

[0278] In addition, Figure 40 provides a list of blood-based solid tumor-specific exon transcripts that may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood. Overexpressed oncogene transcripts, or transcripts of mutant oncogenes may be enriched in exosomes, as they may drive spread of the cancer.

[0279] Figure 41 provides a list of cancer protein markers, identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from solid tumors, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0280] Protein markers that can be secreted by solid tumors into the blood include, but are not limited to: (Protein name, UniProt ID); Uncharacterized protein C19orf48, Q6RUI8; Protein FAM72B, Q86X60; Protein FAM72D, Q6L9T8; Hydroxyacylglutathione hydrolase-like

protein, Q6PII5; Putative methyltransferase NSUN5, Q96P11; RNA pseudouridylate synthase domain-containing protein 1, Q9UJJ7; Collagen triple helix repeat-containing protein 1, Q96CG8; Interleukin-11, P20809; Stromelysin-2, P09238; Matrix metalloproteinase-9, P14780; Podocan-like protein 1, Q6PEZ8; Putative peptide YY-2, Q9NRI6; Osteopontin, P10451; Sulphydryl oxidase 2, Q6ZRP7; Glypican-2, Q8N158; Macrophage migration inhibitory factor, P14174; Peptidyl-prolyl cis-trans isomerase A, P62937; and Calreticulin, P27797. A comparative analysis was performed across various TCGA datasets (tumors, normals), followed by an additional bioinformatics filter (Meinken et al., "Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report," *Computational Molecular Biology* 2(1):1-7 (2012), which predicts the likelihood that the translated protein is secreted by the cells.

[0281] Commonly found mutations in solid tumors may be used as plasma-based markers of cancer, and they are available in the COSMIC and/or TCGA datasets: *TP53* (*tumor protein p53*), *TTN* (*titin*), *MUC16* (*mucin 16*), *KRAS* (*K-Ras*). Initial work identifying mutations in the plasma from patients with metastatic disease revealed an average of 5 mutations not only in the patients, but also in age-matched controls. A follow-up study using a 2 Mb, 508-gene panel and sequencing to more than 60,000-fold depth, showed mutations appeared in 93.6 percent of the white blood cells from individuals without cancer and 99.1 percent of those with cancer (Razavi, et al., "Cell-free DNA (cfDNA) Mutations From Clonal Hematopoiesis: Implications for Interpretation of Liquid Biopsy Tests," *Journal of Clinical Oncology* 35(15):11526-11526 (2017); Razavi, et al., "High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants," *Nature Medicine*, Dec;25(12):1928-1937 (2019), which are hereby incorporated by reference in their entirety). This phenomenon, known as age-related clonal hematopoiesis, results from accumulation of mutations in white-blood cells, that then undergo clonal expansion. When screening for mutations markers in the plasma, it is important to always sequence an aliquot of WBC DNA from the same individual, such that a presumptive positive mutation is verified as arising from internal tissue (presumably corresponding to a tumor) and not due to clonal hematopoiesis.

[0282] A deep analysis of the TCGA database of methylation markers that are absent in blood but on average are present in solid tumor types at 50% sensitivity show three general categories of clusters: (i) Markers that are present in colorectal cancers, and related GI cancer (stomach & esophagus), (ii) Markers that are present in colorectal cancers, and related GI cancer (stomach & esophagus), as well as other tumors, and (iii) Markers that are mostly absent in colorectal cancers, but present in other tumors. Second, while for some tumor types one could readily identify markers that were unique to that group, such as Group 2 (breast, endometrial,

ovarian, cervical, and uterine), for other tumor types such as lung cancer or pancreatic cancer, it was difficult to identify methylation markers that were unique to that cancer. Consequently, to assemble a set of 96 markers that satisfied the criteria of at least 36 markers with 50% sensitivity that covers each of the aforementioned 16 types of solid tumors, the first 48 markers comprised of about 12 markers that were strongly represented in Group 2 tumors, about 12 markers that were strongly represented in Group 3 tumors, about 12 markers that were strongly represented in Group 4 tumors, and about 12 markers that were strongly represented in Group 5 tumors. The remaining 48 markers comprised about 12 markers that were strongly represented in Groups 1 & 2 tumors, about 12 markers that were strongly represented in Groups 1 & 3 tumors, about 12 markers that were strongly represented in Groups 1 & 4 tumors, and about 12 markers that were strongly represented in Groups 1 & 5 tumors.

[0283] Figure 42 provides a list of primary CpG sites that are solid tumors and tissue-specific markers, that may be used to identify the presence of solid tumors from cfDNA, DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 43 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are solid tumor and tissue-specific markers, that may be used to identify the presence of solid tumors from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites, as well as alternative markers that are low to no-CRC, and alternative markers that are high is CRC, with or without being high for other cancers as well. Primer sets for these preferred and alternative methylation markers are listed in Table 40 in the prophetic experimental section.

[0284] Table 39 provides simulations of the 96-marker assay, with average sensitivities of 50%, for identifying most probably group for tissue of origin, for both sexes. A set of 96 markers was assembled as above and the percentage of samples positive within each of the cancer patients in the TCGA and GEO databases was assessed. The total number of patients for each cancer analyzed are: Group 1 (colorectal, CRC-Pt = 395; stomach, ST-Pt = 260; esophagus, ES-Pt = 185); Group 2 (breast, BR-Pt = 668; endometrial, END-Pt = 431; ovarian, OV-Pt = 79; cervical, CERV-Pt = 307; uterine, UTCS-Pt = 57); Group 3 (lung adenocarcinoma, LUAD = 450; lung squamous cell carcinoma, LUSC = 372; head & neck, HNSC-Pt = 528); Group 4 (prostate, PROS-Pt = 192; bladder, BLAD-Pt = 412); and Group 5 (liver, LIV-Pt = 377; pancreatic, PANC-Pt = 184; and gall bladder, BILE-Pt = 36). The columns reflect the total percent patients positive for each of the markers divided by the total number of markers used – for the first row of all cancers, that would be 96 markers. Thus, on average, of the 96 markers

chosen, the number of average sensitivity scores are: Group 1 (colorectal = 44, stomach = 45, esophagus = 40); Group 2 (breast = 38, endometrial = 40, ovarian = 22, cervical = 39, uterine = 33); Group 3 (lung adenocarcinoma = 31, lung squamous cell carcinoma = 31, head & neck = 33); Group 4 (prostate = 45, bladder = 36); and Group 5 (liver = 38, pancreatic = 27, gall bladder = 47). This translates into the following number of marker equivalents with average sensitivities of 50% ($= 96 \times \text{score}/50$): (colorectal = 85 marker equivalents; stomach = 86 marker equivalents; esophagus = 78 marker equivalents); Group 2 (breast = 74 marker equivalents; endometrial = 76 marker equivalents; ovarian = 42 marker equivalents; cervical = 75 marker equivalents; uterine = 64 marker equivalents); Group 3 (lung adenocarcinoma = 60 marker equivalents; lung squamous cell carcinoma = 59 marker equivalents; head & neck = 64 marker equivalents); Group 4 (prostate = 86 marker equivalents; bladder = 70 marker equivalents); and Group 5 (liver = 74 marker equivalents; pancreatic = 51 marker equivalents; gall bladder = 91 marker equivalents). Thus, cancers were well represented, ranging from 42 to 91 marker equivalents for the different cancer types, and all well above the minimum of 36 markers with average sensitivities of 50%.

Table 39. Simulation of 96-marker assay, with average sensitivities of 50%, for identifying most probably group for tissue of origin, for both sexes.

	CRC- Pt	ST- Pt	ES- Pt	BR- Pt	END- Pt	OV- Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUUSC- Pt	HNUSC- Pt	PROS- Pt	BLAD- Pt	LIV- Pt	PANC- Pt	BILE- Pt
All																
All Cancer	44	45	40	38	40	22	39	33	31	31	33	45	36	38	27	47
CRC1	66	57	51	39	37	15	48	30	35	34	39	37	42	42	31	51
CRC2	66	53	47	35	43	22	42	31	28	28	35	46	39	35	28	46
ST1	56	55	46	36	34	17	40	26	37	31	33	39	37	46	31	53
ST2	56	54	50	39	41	20	48	30	33	36	42	42	43	37	30	51
ES1	57	54	52	38	47	21	51	38	38	40	42	36	38	28	26	45
ES2	58	56	50	39	34	19	41	23	31	30	39	45	41	48	34	55
BR1	47	49	47	50	47	26	45	35	33	35	41	50	41	38	30	51
BR2	40	39	37	49	50	31	41	41	36	33	32	49	37	34	26	49
ENDO1	44	49	50	50	62	41	48	54	37	43	44	50	43	34	29	53
ENDO2	42	39	39	41	61	30	47	50	31	33	36	39	33	26	24	38
OV1	35	41	43	58	68	58	49	59	38	40	38	49	36	23	27	44
OV2	37	40	40	43	71	56	48	66	31	36	37	33	34	23	21	40
CERV1	40	47	53	41	47	23	57	38	37	49	53	49	43	26	25	44
CERV2	56	52	50	40	53	32	57	43	32	35	44	34	40	32	27	46
UTCS1	25	30	33	47	66	39	42	60	31	34	32	53	34	27	21	37
UTCS2	46	46	47	40	58	38	51	59	28	40	42	38	36	26	24	47
LUAD1	50	55	53	47	43	24	49	31	47	44	44	42	41	37	30	54
LUAD2	49	50	47	44	49	28	49	39	46	39	40	51	41	39	31	52
LUSC1	46	50	55	40	48	30	51	41	40	53	52	40	42	24	24	45
LUSC2	44	49	53	50	52	27	54	43	43	51	52	56	44	33	29	54
HNUSC1	47	54	58	44	49	28	55	39	39	49	53	43	50	34	29	52
HNUSC2	56	51	52	44	51	23	58	40	37	43	52	47	40	30	27	44
PROS1	42	43	41	41	40	21	38	35	36	34	35	63	39	42	29	52
PROS2	42	40	37	42	41	21	39	34	28	30	33	62	39	35	23	45

		CRC- Pt	ST- Pt	ES- Pt	BR- Pt	END- Pt	OV- Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HN5C- Pt	PROS- Pt	BLAD- Pt	LIV- Pt	PANC- Pt	BILE- Pt
BLAD1	Total	48	47	45	38	42	20	47	36	35	39	39	53	49	31	25	48
BLAD2	Total	58	54	51	44	43	21	46	34	35	35	43	42	48	41	31	49
LIV1	Total	46	52	41	37	25	12	30	18	34	27	28	41	36	61	34	62
LIV2	Total	49	48	38	38	33	20	32	27	33	23	27	47	36	60	33	58
PANC1	Total	52	57	47	38	32	16	40	21	34	29	35	46	36	58	41	60
PANC2	Total	53	54	46	39	35	24	36	33	36	33	33	41	45	44	40	57
BILE1	Total	50	51	45	40	37	22	40	27	39	33	35	47	39	44	32	58
BILE2	Total	48	50	43	37	35	17	40	28	29	31	34	42	37	50	30	57

[0285] The above numbers translate into the following number of marker equivalents with average sensitivities of 66% ($= 96 \times \text{score}/66$); (colorectal = 65 marker equivalents; stomach = 65 marker equivalents; esophagus = 59 marker equivalents); Group 2 (breast = 56 marker equivalents; endometrial = 58 marker equivalents; ovarian = 32 marker equivalents; cervical = 57 marker equivalents; uterine = 48 marker equivalents); Group 3 (lung adenocarcinoma = 45 marker equivalents; lung squamous cell carcinoma = 45 marker equivalents; head & neck = 48 marker equivalents); Group 4 (prostate = 65 marker equivalents; bladder = 53 marker equivalents); and Group 5 (liver = 56 marker equivalents; pancreatic = 39 marker equivalents; gall bladder = 69 marker equivalents). Thus, cancers were well represented, ranging from 32 to 69 marker equivalents for the different cancer types, and with the exception of ovarian cancer at 32, the other cancer types are above the minimum of 36 markers with average sensitivities of 66%.

[0286] The aforementioned markers were then re-ordered for each of the above cancer types such that the most prevalent markers were listed first. For example, with CRC, of the 96 markers, 54 markers gave scores above 55 (i.e. were positive in greater than 55% of the 395 patients) and 9 gave scores of between 25 and 54 (i.e. were positive for from 25% to 54% of the 395 patients). Half of the higher, and a third of the lower set, for a total of 30 markers were distributed into two marker test sets, designated “CRC1” and “CRC2” (Table 41, rows 2 & 3). These marker sets would reflect an ideal result if half the markers with the potential to be positive are detected in the assay. This does not account for the chances that earlier stage tumors would have a lower number of marker molecules in the plasma, and thus consequently the actual number of markers positive would be less than the ideal result in this simulation. The percent of patients positive for each of the cancers were recorded and then divided by the total number of markers used for that cancer type. As anticipated, when selecting markers for a given tumor type, those markers should give a higher score than the average, *i.e.* 66 for CRC in each of the two sets of selected 30 markers, compared with a score of 44 for the unselected 96 markers. These markers form a diagonal across Table 41 and are highlighted in bold and light grey background.

[0287] For each column, marker sets that are in the same range or higher than the number of positive markers for that cancer type are also shown with a light grey background. For example, a patient with colorectal, stomach, or esophageal cancer will be scored as potentially positive with stomach cancer. This makes sense as the markers for these three cancers overlap, they all bin to Group 1, so they could be distinguished in step 2 of the assay on the group 1 markers, wherein these markers are more cancer types specific, to tease out the most probable cancer type. Evaluation of the ST-Pt column shows that simulations for one of the two LUAD, BLAD, and both PANC also gave scores that might be interpreted as stomach cancer. Thus, the

first step is not always able to pinpoint what Groups should be tested in the second step of the assay. However, most of the ambiguity is within group members (*i.e.* Group 2), and this makes sense, since the markers were chosen to maximize the ability to chose which groups to test in the second step.

[0288] Tables 40 and 41 take the aforementioned results in the simulations in Table 39 and multiplies them by the percent incidence of the given cancer type for that gender (see Tables 37 & 38 respectively), and the result is adjusted to the same order of magnitude (multiple by 10). The concept is for the physician to take into account that a lower score for a high incidence cancer (such as CRC) may be a more common tissue of origin for a higher score for a low incidence cancer (such as lung squamous cell carcinoma). Tables 40 and 41 show the level of ambiguity in identifying tissue of origin is higher among female patients then among male patients, as indicated by the number of cells highlighted in grey that are not on the diagonal. In all cases, the physician will need to incorporate all data, such as smoking history, not just molecular data to determine the most likely tissue of origin before sending the patient to confirmatory imaging.

Table 40. Simulation of 96-marker assay, with average sensitivities of 50%, for identifying most probably group for tissue of origin, for male cancers.

		CRC- Pt	ST- Pt	ES- Pt	BR- Pt	END -Pt	OV- Pt	CERV -Pt	UTCS -Pt	LUAD- Pt	LUSC -Pt	HNSC -Pt	PROS -Pt	BLAD -Pt	LIV- Pt	PAN C-Pt	BILE- Pt
All	Male Score	63	43	23	0	0	0	0	0	51	9	29	80	22	32	9	7
CRC1	Male Score	94	54	29	0	0	0	0	0	57	10	35	67	25	35	11	7
CRC2	Male Score	93	51	26	0	0	0	0	0	46	8	31	82	23	30	9	6
ST1	Male Score	79	53	26	0	0	0	0	0	61	9	30	70	22	38	10	7
ST2	Male Score	79	52	28	0	0	0	0	0	54	10	38	74	26	31	10	7
ES1	Male Score	80	52	29	0	0	0	0	0	62	11	37	64	23	24	9	6
ES2	Male Score	82	54	28	0	0	0	0	0	50	9	34	80	25	40	12	8
BR1	Male Score	66	47	26	0	0	0	0	0	54	10	36	89	25	32	10	7
BR2	Male Score	56	37	21	0	0	0	0	0	59	10	29	88	22	29	9	7
ENDO1	Male Score	62	47	28	0	0	0	0	0	60	13	40	89	26	28	10	7
ENDO2	Male Score	59	37	22	0	0	0	0	0	50	10	32	70	20	22	8	5
OV1	Male Score	50	39	24	0	0	0	0	0	62	12	34	87	22	19	9	6
OV2	Male Score	52	38	23	0	0	0	0	0	51	10	33	59	20	20	7	6
CERV1	Male Score	57	45	30	0	0	0	0	0	60	14	47	88	26	22	8	6
CERV2	Male Score	79	50	28	0	0	0	0	0	53	10	39	62	24	27	9	6
UTCS1	Male Score	36	29	19	0	0	0	0	0	50	10	29	96	20	23	7	5
UTCS2	Male Score	65	44	26	0	0	0	0	0	46	12	38	69	22	22	8	7
LUAD1	Male Score	71	53	30	0	0	0	0	0	77	13	40	75	25	31	10	8
LUAD2	Male Score	70	48	26	0	0	0	0	0	74	11	36	92	25	33	10	7
LUSC1	Male Score	65	48	31	0	0	0	0	0	65	15	47	71	25	20	8	6
LUSC2	Male Score	61	48	30	0	0	0	0	0	70	15	47	100	26	28	10	8
HNSC1	Male Score	67	52	33	0	0	0	0	0	63	14	47	77	30	28	10	7
HNSC2	Male Score	79	49	29	0	0	0	0	0	60	12	46	85	24	25	9	6
PROS1	Male Score	59	42	23	0	0	0	0	0	59	10	31	113	23	35	10	7
PROS2	Male Score	59	38	21	0	0	0	0	0	46	9	29	111	24	29	8	6
BLAD1	Male Score	67	46	25	0	0	0	0	0	57	11	35	96	29	26	9	7

		CRC- Pt	ST- Pt	ES- Pt	BR- Pt	END -Pt	OV- Pt	CERV -Pt	UTCS -Pt	LUAD- Pt	LUSC -Pt	HNSC -Pt	PROS -Pt	BLAD -Pt	LIV- Pt	PAN C-Pt	BILE- Pt
BLAD2	Male Score	81	51	28	0	0	0	0	0	57	10	38	75	29	35	10	7
LIV1	Male Score	64	50	23	0	0	0	0	0	55	8	25	73	21	51	12	9
LIV2	Male Score	69	46	21	0	0	0	0	0	53	7	24	85	21	50	11	8
PANC1	Male Score	73	55	27	0	0	0	0	0	56	9	31	82	22	48	14	8
PANC2	Male Score	75	52	26	0	0	0	0	0	58	10	29	74	27	37	14	8
BILE1	Male Score	71	49	25	0	0	0	0	0	63	10	31	84	23	37	11	8
BILE2	Male Score	67	48	24	0	0	0	0	0	48	9	30	76	22	42	10	8

Table 41. Simulation of 96-marker assay, with average sensitivities of 50%, for identifying most probably group for tissue of origin, for female cancers.

		CRC- Pt	ST- Pt	ES- Pt	BR- Pt	END -Pt	OV- Pt	CERV -Pt	UTCS -Pt	LUAD- Pt	LUSC -Pt	HNSC -Pt	PROS -Pt	BLAD -Pt	LIV- Pt	PAN C-Pt	BILE- Pt
All	Female Score	46	20	9	103	19	8	29	16	25	4	27	0	20	30	8	6
CRC1	Female Score	68	25	11	105	18	6	36	15	28	5	32	0	23	32	10	7
CRC2	Female Score	68	24	10	94	21	8	31	15	22	4	28	0	22	27	9	6
ST1	Female Score	58	25	10	98	17	7	29	13	29	4	27	0	20	35	9	7
ST2	Female Score	58	24	11	105	20	8	35	15	26	5	35	0	23	28	9	7
ES1	Female Score	58	24	11	101	23	8	38	19	30	6	34	0	21	22	8	6
ES2	Female Score	60	25	11	105	17	7	31	11	24	4	32	0	23	37	11	7
BR1	Female Score	48	22	10	134	23	10	34	17	26	5	33	0	23	30	9	7
BR2	Female Score	41	18	8	132	24	12	30	20	29	5	27	0	20	26	8	6
ENDO1	Female Score	46	22	11	134	30	16	36	27	29	6	36	0	24	26	9	7
ENDO2	Female Score	43	18	9	111	30	11	35	25	24	5	29	0	18	20	7	5
OV1	Female Score	36	18	9	155	34	22	36	29	30	6	32	0	20	18	8	6
OV2	Female Score	38	18	9	115	35	21	36	32	25	5	30	0	19	18	6	5
CERV1	Female Score	42	21	12	111	23	9	42	19	29	7	43	0	24	20	8	6
CERV2	Female Score	58	23	11	108	26	12	42	21	26	5	36	0	22	24	8	6

		CRC- Pt	ST- Pt	ES- Pt	BR- Pt	END -Pt	OV- Pt	CERV -Pt	UTCS -Pt	LUAD- Pt	LUSC -Pt	HNSC -Pt	PROS -Pt	BLAD -Pt	LIV- Pt	PAN C-Pt	BILE- Pt
UTCS1	Female Score	26	14	7	125	33	15	31	30	24	5	27	0	19	21	6	5
UTCS2	Female Score	48	21	10	107	29	15	37	29	22	6	35	0	20	20	7	6
LUAD1	Female Score	52	25	12	127	21	9	36	15	37	6	36	0	23	29	9	7
LUAD2	Female Score	51	23	10	117	24	11	36	19	36	5	33	0	23	30	9	7
LUSC1	Female Score	48	22	12	108	23	11	38	20	32	7	43	0	23	18	7	6
LUSC2	Female Score	45	22	12	136	26	10	40	21	34	7	43	0	24	26	9	7
HNSC1	Female Score	49	25	13	117	24	11	41	19	31	7	43	0	27	26	9	7
HNSC2	Female Score	57	23	12	118	25	9	43	19	29	6	43	0	22	23	8	6
PROS1	Female Score	43	19	9	110	19	8	28	17	29	5	28	0	21	32	9	7
PROS2	Female Score	43	18	8	114	20	8	29	17	22	4	27	0	22	27	7	6
BLAD1	Female Score	49	21	10	103	21	8	35	18	28	5	32	0	27	24	8	6
BLAD2	Female Score	59	24	11	117	21	8	34	17	28	5	35	0	26	32	10	6
LIV1	Female Score	47	23	9	99	12	4	22	9	27	4	23	0	20	47	11	8
LIV2	Female Score	50	21	8	102	16	8	24	13	26	3	22	0	20	46	10	7
PANC1	Female Score	53	26	10	102	16	6	30	10	27	4	29	0	20	44	13	8
PANC2	Female Score	55	24	10	105	17	9	26	16	28	5	27	0	25	34	12	7
BILE1	Female Score	52	23	10	108	18	8	30	13	31	5	29	0	22	33	10	8
BILE2	Female Score	49	23	10	99	17	6	29	14	23	4	28	0	20	38	9	7

[0289] Tables 42, 43, and 44 takes the aforementioned results in the simulations in Tables 39, 40, and 41 and determines the percent deviation from the neutral result by taking the percentage of $(= \text{score specific cancer type simulation} / \text{score all cancer for that type} - 1)$. Thus, the first row of each of these tables should be 0%. Again, those percentages that are higher than, or in the same range as the percentages across the diagonal are highlighted in light gray. While this set of marker selection may be less than ideal for distinguishing esophageal or gall bladder cancers as the tissue of origin, they are nevertheless quite informative for guiding the physician to which groups of the Step 2 assays should be tested. This simple scoring may be augmented by using AI approaches based on a database of results with clinical samples using the aforementioned 96-marker set.

Table 42. Simulation of 96-marker assay, with average sensitivities of 50%, showing percent deviation from neutral result, for identifying most probably group for tissue of origin, for both sexes.

	CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HNSC- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
All	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
CRC1	50%	27%	27%	2%	-6%	-31%	23%	-10%	12%	9%	17%	-16%	16%	9%	16%	8%
CRC2	49%	19%	16%	-9%	9%	2%	8%	-6%	-9%	-8%	5%	3%	8%	-8%	3%	-3%
ST1	26%	23%	14%	-6%	-14%	-22%	2%	-22%	19%	1%	0%	-13%	2%	20%	16%	12%
ST2	26%	21%	24%	2%	3%	-8%	22%	-10%	6%	17%	27%	-6%	19%	-4%	12%	8%
ES1	28%	21%	29%	-1%	18%	-4%	30%	14%	22%	30%	27%	-19%	5%	-27%	-3%	-5%
ES2	31%	26%	25%	2%	-15%	-12%	5%	-30%	-1%	-2%	17%	0%	13%	24%	27%	17%
BR1	6%	10%	16%	30%	18%	19%	15%	5%	6%	14%	24%	12%	13%	-1%	12%	8%
BR2	-11%	-13%	-8%	28%	25%	44%	4%	24%	16%	8%	-2%	11%	1%	-10%	-3%	4%
ENDO1	-1%	10%	24%	30%	56%	88%	22%	63%	19%	40%	33%	12%	19%	-11%	8%	12%
ENDO2	-6%	-13%	-4%	7%	53%	37%	20%	52%	-1%	7%	8%	-13%	-8%	-31%	-12%	-21%
OV1	-21%	-8%	6%	51%	71%	166%	25%	78%	22%	30%	15%	10%	-1%	-40%	1%	-7%
OV2	-17%	-11%	0%	12%	79%	157%	22%	98%	0%	16%	12%	-27%	-7%	-39%	-22%	-15%
CERV1	-9%	5%	32%	8%	18%	7%	46%	14%	18%	61%	59%	10%	18%	-32%	-8%	-7%
CERV2	27%	16%	24%	4%	34%	46%	44%	29%	4%	13%	31%	-23%	9%	-17%	1%	-4%
UTCS1	-44%	-33%	-18%	23%	66%	79%	7%	81%	0%	10%	-3%	19%	-6%	-30%	-22%	-22%
UTCS2	4%	4%	16%	4%	47%	77%	29%	78%	-10%	31%	28%	-14%	0%	-33%	-10%	0%
LUAD1	13%	23%	31%	23%	8%	10%	25%	-7%	51%	43%	33%	-6%	13%	-4%	12%	14%
LUAD2	11%	13%	15%	14%	23%	28%	24%	16%	47%	27%	22%	15%	14%	2%	14%	11%
LUSC1	4%	12%	36%	4%	21%	38%	30%	23%	29%	72%	57%	-10%	16%	-37%	-10%	-5%
LUSC2	-2%	11%	31%	31%	31%	25%	38%	28%	38%	67%	58%	25%	21%	-13%	9%	14%
HNSC1	6%	21%	44%	15%	23%	29%	40%	17%	25%	59%	60%	-4%	38%	-11%	8%	10%
HNSC2	26%	15%	30%	14%	30%	7%	47%	19%	18%	39%	57%	6%	12%	-22%	2%	-8%
PROS1	-6%	-3%	0%	7%	0%	-2%	-3%	5%	16%	12%	5%	42%	8%	9%	7%	9%
PROS2	-7%	-10%	-7%	10%	4%	-5%	0%	1%	-9%	-1%	-2%	39%	8%	-9%	-14%	-5%

		CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HN5C- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
BLAD1	Total	7%	6%	10%	0%	6%	-6%	20%	9%	12%	27%	19%	20%	36%	-18%	-5%	2%
BLAD2	Total	29%	20%	26%	14%	8%	-2%	18%	4%	13%	14%	30%	-6%	32%	8%	15%	3%
LIV1	Total	4%	17%	2%	-4%	-37%	-45%	-24%	-46%	9%	-12%	-16%	-8%	-1%	59%	27%	31%
LIV2	Total	10%	7%	-7%	-1%	-17%	-9%	-19%	-19%	5%	-24%	-18%	6%	-2%	56%	22%	21%
PANC1	Total	16%	29%	18%	-1%	-19%	-27%	3%	-38%	11%	-4%	6%	3%	0%	50%	54%	27%
PANC2	Total	20%	21%	13%	1%	-12%	10%	-9%	-1%	15%	8%	0%	-7%	26%	16%	50%	19%
BILE1	Total	13%	14%	11%	5%	-7%	-1%	2%	-19%	24%	8%	6%	5%	8%	13%	20%	23%
BILE2	Total	8%	12%	7%	-4%	-13%	-24%	1%	-16%	-6%	-1%	2%	-5%	2%	29%	12%	21%

Table 43. Simulation of 96-marker assay, with average sensitivities of 50%, showing percent deviation from neutral result, for identifying most probably group for tissue of origin, for male cancers.

		CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HN5C- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
All	Male Score	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
CRC1	Male Score	50%	26%	28%	0%	0%	0%	0%	0%	13%	12%	19%	-16%	15%	9%	21%	6%
CRC2	Male Score	49%	19%	16%	0%	0%	0%	0%	0%	-9%	-8%	5%	3%	8%	-8%	3%	-3%
ST1	Male Score	26%	24%	15%	0%	0%	0%	0%	0%	20%	0%	1%	-12%	1%	19%	14%	12%
ST2	Male Score	27%	22%	24%	0%	0%	0%	0%	0%	7%	17%	28%	-7%	18%	-4%	13%	8%
ES1	Male Score	27%	21%	29%	0%	0%	0%	0%	0%	22%	29%	26%	-19%	6%	-26%	-2%	-4%
ES2	Male Score	31%	26%	25%	0%	0%	0%	0%	0%	-1%	-2%	17%	0%	13%	24%	27%	17%
BR1	Male Score	6%	11%	16%	0%	0%	0%	0%	0%	7%	13%	23%	12%	14%	0%	12%	7%
BR2	Male Score	-11%	-13%	-8%	0%	0%	0%	0%	0%	16%	8%	-2%	11%	1%	-10%	-3%	4%
ENDO1	Male Score	0%	9%	25%	0%	0%	0%	0%	0%	19%	41%	34%	11%	19%	-12%	10%	12%
ENDO2	Male Score	-6%	-13%	-4%	0%	0%	0%	0%	0%	-1%	7%	8%	-13%	-8%	-31%	-12%	-21%
OV1	Male Score	-21%	-9%	6%	0%	0%	0%	0%	0%	22%	31%	16%	9%	-1%	-40%	0%	-7%
OV2	Male Score	-17%	-11%	0%	0%	0%	0%	0%	0%	0%	16%	12%	-27%	-7%	-39%	-22%	-15%
CERV1	Male Score	-9%	5%	32%	0%	0%	0%	0%	0%	18%	61%	59%	10%	18%	-32%	-8%	-7%

		CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HNSC- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
CERV2	Male Score	27%	16%	24%	0%	0%	0%	0%	0%	4%	13%	31%	-23%	9%	-17%	1%	-4%
UTCS1	Male Score	-43%	-33%	-18%	0%	0%	0%	0%	0%	-2%	10%	-2%	20%	-6%	-28%	-23%	-23%
UTCS2	Male Score	4%	4%	16%	0%	0%	0%	0%	0%	-10%	31%	28%	-14%	0%	-33%	-10%	0%
LUAD1	Male Score	13%	24%	31%	0%	0%	0%	0%	0%	51%	43%	34%	-6%	14%	-2%	10%	14%
LUAD2	Male Score	11%	13%	15%	0%	0%	0%	0%	0%	47%	27%	22%	15%	14%	2%	14%	11%
LUSC1	Male Score	4%	12%	36%	0%	0%	0%	0%	0%	29%	71%	58%	-11%	15%	-38%	-11%	-4%
LUSC2	Male Score	-2%	11%	31%	0%	0%	0%	0%	0%	38%	67%	58%	25%	21%	-13%	9%	14%
HNCS1	Male Score	7%	22%	44%	0%	0%	0%	0%	0%	25%	59%	59%	-4%	38%	-12%	7%	11%
HNCS2	Male Score	26%	15%	30%	0%	0%	0%	0%	0%	18%	39%	57%	6%	12%	-22%	2%	-8%
PROS1	Male Score	-6%	-3%	0%	0%	0%	0%	0%	0%	16%	12%	5%	42%	8%	9%	7%	9%
PROS2	Male Score	-7%	-10%	-7%	0%	0%	0%	0%	0%	-9%	-1%	-2%	39%	8%	-9%	-14%	-5%
BLAD1	Male Score	7%	6%	10%	0%	0%	0%	0%	0%	12%	27%	19%	20%	36%	-18%	-5%	2%
BLAD2	Male Score	29%	20%	26%	0%	0%	0%	0%	0%	13%	14%	30%	-6%	32%	8%	15%	3%
LIV1	Male Score	3%	17%	1%	0%	0%	0%	0%	0%	9%	-11%	-16%	-9%	-1%	59%	27%	30%
LIV2	Male Score	10%	7%	-7%	0%	0%	0%	0%	0%	5%	-24%	-18%	6%	-2%	56%	22%	21%
PANC1	Male Score	16%	29%	18%	0%	0%	0%	0%	0%	11%	-4%	6%	3%	0%	50%	54%	27%
PANC2	Male Score	20%	21%	13%	0%	0%	0%	0%	0%	15%	8%	0%	-7%	26%	16%	50%	19%
BILE1	Male Score	13%	14%	11%	0%	0%	0%	0%	0%	24%	8%	6%	5%	8%	13%	20%	23%
BILE2	Male Score	8%	12%	7%	0%	0%	0%	0%	0%	-6%	-1%	2%	-5%	2%	29%	12%	21%

Table 44. Simulation of 96-marker assay, with average sensitivities of 50%, showing percent deviation from neutral result, for identifying most probably group for tissue of origin, for female cancers.

		CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HNSC- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
All	Female Score	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
CRC1	Female Score	50%	27%	27%	2%	-6%	-31%	23%	-10%	12%	9%	17%	0%	16%	9%	16%	8%
CRC2	Female Score	49%	19%	16%	-9%	9%	2%	8%	-6%	-9%	-8%	5%	0%	8%	-8%	3%	-3%

		CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HNCS- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
ST1	Female Score	26%	24%	15%	-5%	-13%	-20%	1%	-22%	20%	0%	1%	0%	1%	19%	14%	12%
ST2	Female Score	27%	22%	24%	2%	4%	-9%	22%	-9%	7%	17%	28%	0%	18%	-4%	13%	8%
ES1	Female Score	27%	21%	29%	-2%	19%	-3%	31%	14%	22%	29%	26%	0%	6%	-26%	-2%	-4%
ES2	Female Score	31%	26%	25%	2%	-15%	-12%	5%	-30%	-1%	-2%	17%	0%	13%	24%	27%	17%
BR1	Female Score	6%	11%	16%	30%	18%	17%	15%	6%	7%	13%	23%	0%	14%	0%	12%	7%
BR2	Female Score	-11%	-13%	-8%	28%	25%	44%	4%	24%	16%	8%	-2%	0%	1%	-10%	-3%	4%
ENDO1	Female Score	0%	9%	25%	30%	56%	89%	23%	63%	19%	41%	34%	0%	19%	-12%	10%	12%
ENDO2	Female Score	-6%	-13%	-4%	7%	53%	37%	20%	52%	-1%	7%	8%	0%	-8%	-31%	-12%	-21%
OV1	Female Score	-21%	-9%	6%	50%	72%	168%	24%	76%	22%	31%	16%	0%	-1%	-40%	0%	-7%
OV2	Female Score	-17%	-11%	0%	12%	79%	157%	22%	98%	0%	16%	12%	0%	-7%	-39%	-22%	-15%
CERV1	Female Score	-9%	5%	32%	8%	18%	7%	46%	14%	18%	61%	59%	0%	18%	-32%	-8%	-7%
CERV2	Female Score	27%	16%	24%	4%	34%	46%	44%	29%	4%	13%	31%	0%	9%	-17%	1%	-4%
UTCS1	Female Score	-43%	-33%	-18%	21%	67%	80%	7%	82%	-2%	10%	-2%	0%	-6%	-28%	-23%	-23%
UTCS2	Female Score	4%	4%	16%	4%	47%	77%	29%	78%	-10%	31%	28%	0%	0%	-33%	-10%	0%
LUAD1	Female Score	13%	24%	31%	23%	8%	11%	24%	-7%	51%	43%	34%	0%	14%	-2%	10%	14%
LUAD2	Female Score	11%	13%	15%	14%	23%	28%	24%	16%	47%	27%	22%	0%	14%	2%	14%	11%
LUSC1	Female Score	4%	12%	36%	4%	20%	37%	31%	22%	29%	71%	58%	0%	15%	-38%	-11%	-4%
LUSC2	Female Score	-2%	11%	31%	31%	31%	25%	38%	28%	38%	67%	58%	0%	21%	-13%	9%	14%
HNCS1	Female Score	7%	22%	44%	14%	23%	28%	40%	18%	25%	59%	59%	0%	38%	-12%	7%	11%
	Female Score	26%	15%	30%	14%	30%	7%	47%	19%	18%	39%	57%	0%	12%	-22%	2%	-8%
PROS1	Female Score	-6%	-3%	0%	7%	0%	-2%	-3%	5%	16%	12%	5%	0%	8%	9%	7%	9%
PROS2	Female Score	-7%	-10%	-7%	10%	4%	-5%	0%	1%	-9%	-1%	-2%	0%	8%	-9%	-14%	-5%
BLAD1	Female Score	7%	6%	10%	0%	6%	-6%	20%	9%	12%	27%	19%	0%	36%	-18%	-5%	2%
BLAD2	Female Score	29%	20%	26%	14%	8%	-2%	18%	4%	13%	14%	30%	0%	32%	8%	15%	3%
LIV1	Female Score	3%	17%	1%	-4%	-36%	-46%	-24%	-47%	9%	-11%	-16%	0%	-1%	59%	27%	30%
LIV2	Female Score	10%	7%	-7%	-1%	-17%	-9%	-19%	-19%	5%	-24%	-18%	0%	-2%	56%	22%	21%
PANC1	Female Score	16%	29%	18%	-1%	-19%	-27%	3%	-38%	11%	-4%	6%	0%	0%	50%	54%	27%
PANC2	Female Score	20%	21%	13%	1%	-12%	10%	-9%	-1%	15%	8%	0%	0%	26%	16%	50%	19%
BILE1	Female Score	13%	14%	11%	5%	-7%	-1%	2%	-19%	24%	8%	6%	0%	8%	13%	20%	23%

		CRC- Pt	ST-Pt	ES-Pt	BR-Pt	END- Pt	OV-Pt	CERV- Pt	UTCS- Pt	LUAD- Pt	LUSC- Pt	HNSC- Pt	PROS- Pt	BLAD- Pt	LIV-Pt	PANC- Pt	BILE- Pt
BILE2	Female Score	8%	12%	7%	-4%	-13%	-24%	1%	-16%	-6%	-1%	2%	0%	2%	29%	12%	21%

[0290] For the second step of the assay, one or two or more of the following groups will be tested, each group with a set of 64 markers that on average comprise at least 36 markers with 50% sensitivity that covers each of the aforementioned 16 types of solid tumors, in the following groups: Group 1 (colorectal, stomach, and esophagus); Group 2 (breast, endometrial, ovarian, cervical, and uterine); Group 3 (lung and head & neck); Group 4 (prostate and bladder); and Group 5 (liver, pancreatic, or gall bladder). These Group-specific and cancer type-specific markers include, but are not limited to, cancer-specific microRNA markers, cancer-specific ncRNA and lncRNA markers, cancer-specific exon transcripts, tumor-associated antigens, cancer protein markers, and protein markers that can be secreted by solid tumors into the blood, common mutations, primary CpG sites that are solid tumor and tissue specific markers, chromosomal regions or sub-regions within which are primary CpG sites that are solid tumor and tissue specific markers, and primary and flanking CpG sites that are solid tumor and tissue specific markers. Methods for detecting said markers have been discussed earlier in this application, and Figures listing these markers are described for each of the groups below.

[0291] Group 1 (colorectal, stomach, and esophagus): Blood-based, colorectal, stomach, and esophageal cancer-specific microRNA markers that may be used to distinguish group 1 from other groups include, but are not limited to: (mir ID , Gene ID): hsa-mir-624 , MIR624. This miRNA was identified through analysis of TCGA microRNA datasets, and may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood.

[0292] Blood-based, colorectal, stomach, and esophageal cancer-specific ncRNA and lncRNA markers that may be used to distinguish group 1 from other groups include, but are not limited to: [Gene ID, Coordinate (GRCh38)], ENSEMBL ID: LINC01558, chr6:167784537-167796859, ENSG00000146521.8. This ncRNA was identified through comparative analysis of various publicly available Affymetrix Exon ST CEL data, which were aligned to GENCODE annotations to generate ncRNA and lncRNA transcriptome datasets. Such lncRNA and ncRNA may be enriched in exosomes or other protected states in the blood.

[0293] In addition, Figure 44 provides a list of blood-based colorectal, stomach, and esophageal cancer-specific exon transcripts that may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood.

[0294] Colorectal, stomach, and esophageal cancer protein encoding markers that may be used to distinguish group 1 from other groups include, but are not limited to: (Gene Symbol , Chromosome Band Gene Title, UniProt ID): SELE, 1q22-q25, selectin E, P16581; OTUD4 , 4q31.21, OTU domain containing 4, Q01804; BPI, 20q11.23, bactericidal/permeability-increasing protein, P17213; ASB4, 7q21-q22, ankyrin repeat and SOCS box containing 4,

Q9Y574; C6orf123, 6q27, chromosome 6 open reading frame 123, Q9Y6Z2; KPNA3, 13q14.3, karyopherin alpha 3 (importin alpha 4), O00505; NUP98, 11p15, nucleoporin 98kDa, P52948, identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from colorectal, stomach, and esophageal cancers, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0295] Protein markers that can be secreted by colorectal, stomach, and esophageal cancer into the blood, and may be used to distinguish group 1 from other groups include, but are not limited to: (Protein name, UniProt ID); Bactericidal permeability-increasing protein (BPI) (CAP 57), P1721. A comparative analysis was performed across various TCGA datasets (tumors, normals), followed by an additional bioinformatics filter (Meinken et al., "Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report," *Computational Molecular Biology* 2(1):1-7 (2012), which is hereby incorporated by reference in its entirety), which predicts the likelihood that the translated protein is secreted by the cells.

[0296] The distribution of mutations in colorectal, stomach, and esophageal cancer are available in the public COSMIC database, with the most common being: APC (APC regulator of WNT signaling pathway), ATM (ATM serine/threonine kinase), CSMD1 (CUB and Sushi multiple domains 1), DNAH11 (dynein axonemal heavy chain 11), DST (dystonin), EP400 (E1A binding protein p400), FAT3 (FAT atypical cadherin 3), FAT4 (FAT atypical cadherin 4), FLG (filaggrin), GLI3 (GLI family zinc finger 3), KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), LRP1B (LDL receptor related protein 1B), MUC16 (mucin 16, cell surface associated), OBSCN (obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF), PCLO (piccolo presynaptic cytomatrix protein), PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha), RYR2 (ryanodine receptor 2), SYNE1 (spectrin repeat containing nuclear envelope protein 1), TP53 (tumor protein p53), TTN (titin), and UNC13C (unc-13 homolog C).

[0297] Figure 45 provides a list of primary CpG sites that are colorectal, stomach, and esophageal cancer and tissue-specific markers, that may be used to identify the presence of colorectal, stomach, and esophageal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 46 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are colorectal, stomach, and esophageal cancer and tissue-specific markers, that may be used to identify the presence of colorectal, stomach, and esophageal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain

preferred primary CpG sites and their flanking sites, as well as alternative markers that are high in CRC, and alternative markers that are low to no-CRC, but high in stomach and/or esophageal cancers. Primer sets for these preferred and alternative methylation markers are listed in Table 47 in the prophetic experimental section. A selection of 64 of these markers with average sensitivities of 50% gave the following scores for Group 1: (colorectal = 48, stomach = 51, esophagus = 43), which would translate into the following number of markers equivalents with average sensitivities of 50% ($= 64 \times \text{score}/50$); (colorectal = 62 marker equivalents; stomach = 65 marker equivalents; esophagus = 55 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum. The marker equivalents with average sensitivities of 66% ($= 64 \times \text{score}/66$); (colorectal = 47 marker equivalents; stomach = 50 marker equivalents; esophagus = 42 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum.

[0298] Group 2 (breast, endometrial, ovarian, cervical, and uterine): Blood-based, breast, endometrial, ovarian, cervical, and uterine cancer-specific microRNA markers may be used to distinguish group 2 from other groups include, but are not limited to: (mir ID , Gene ID): hsa-mir-1265 , MIR1265. This marker was identified through analysis of TCGA microRNA datasets, which may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood.

[0299] Blood-based breast, endometrial, ovarian, cervical, and uterine cancer-specific exon transcripts may be used to distinguish group 2 from other groups include, but are not limited to: (Exon location, Gene); chr2:179209013-179209087:+ , OSBPL6; chr2:179251788-179251866:+ , OSBPL6; and chr2:179253736-179253880:+ , OSBPL6, and may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood.

[0300] Breast, endometrial, ovarian, cervical, and uterine cancer protein markers, identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from breast, endometrial, ovarian, cervical, and uterine cancer protein markers, may be used to distinguish group 2 from other groups include, but are not limited to: (Gene Symbol , Chromosome Band , Gene Title , UniProt ID): RSPO2 , 8q23.1 , R-spondin 2 , Q6UXX9; KLC4 , 6p21.1 , kinesin light chain 4 , Q9NSK0; GLRX , 5q14 , glutaredoxin (thioltransferase) , P35754. These markers may be identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from breast, endometrial, ovarian, cervical, and uterine cancers, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0301] Protein markers that can be secreted by breast, endometrial, ovarian, cervical, and uterine cancer into the blood may be used to distinguish group 2 from other groups include, but are not limited to: (Protein name , UniProt ID); R-spondin-2 (Roof plate-specific spondin-2) (hRspo2) , Q6UXX9. A comparative analysis was performed across various TCGA datasets (tumors, normals), followed by an additional bioinformatics filter (Meinken et al., “Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report,” *Computational Molecular Biology* 2(1):1-7 (2012), which is hereby incorporated by reference in its entirety), which predicts the likelihood that the translated protein is secreted by the cells.

[0302] The distribution of mutations in breast, endometrial, ovarian, cervical, and uterine cancer are available in the public COSMIC database, with the 20 most commonly altered genes listed as: PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha), and TTN (titin).

[0303] Figure 47 provides a list of primary CpG sites that are breast, endometrial, ovarian, cervical, and uterine cancer and tissue-specific markers, that may be used to identify the presence of breast, endometrial, ovarian, cervical, and uterine cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 48 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are breast, endometrial, ovarian, cervical, and uterine cancer and tissue-specific markers, that may be used to identify the presence of breast, endometrial, ovarian, cervical, and uterine cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites, as well as alternative markers that may be used to distinguish breast, endometrial, ovarian, cervical, and uterine cancers. Primer sets for these preferred and alternative methylation markers are listed in Table 48 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety, in the prophetic experimental section thereof. A selection of 64 of these markers with average sensitivities of 50% gave the following scores for Group 2: (breast = 36, endometrial = 49, ovarian = 32, cervical = 33, uterine = 47), which would translate into the following number of marker equivalents with average sensitivities of 50% ($= 64 \times \text{score}/50$); (breast = 47 marker equivalents; endometrial = 63 marker equivalents; ovarian = 41 marker equivalents; cervical = 42 marker equivalents; uterine = 61 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum. The marker equivalents with average sensitivities of 66% ($= 64 \times \text{score}/66$); (breast = 35 marker equivalents; endometrial = 48 marker equivalents; ovarian = 31 marker equivalents; cervical = 32 marker equivalents; uterine = 46 marker equivalents). Thus, three markers are below and two

markers are above the average 36-marker equivalents minimum. However, such scores may be improved by selection of different markers.

[0304] Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, and head & neck): Blood-based, lung, head, and neck cancer-specific microRNA markers may be used to distinguish group 3 from other groups include, but are not limited to: (mir ID , Gene ID): hsa-mir-28, MIR28. This marker was identified through analysis of TCGA microRNA datasets, and may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood.

[0305] Blood-based lung, head, and neck cancer-specific exon transcripts may be used to distinguish group 3 from other groups include, but are not limited to: (Exon location, Gene); chr2: chr1:93307721-93309752:- , FAM69A; chr1:93312740-93312916:- , FAM69A; chr1:93316405-93316512:- , FAM69A; chr1:93341853-93342152:- , FAM69A; chr1:93426933-93427079:- , FAM69A; chr7:40221554-40221627:+ , C7orf10; chr7:40234539-40234659:+ , C7orf10; chr8:22265823-22266009:+ , SLC39A14; chr8:22272293-22272415:+ , SLC39A14; chr14:39509936-39510091:- , SEC23A; chr14:39511990-39512076:- , SEC23A, and may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood.

[0306] Lung, head, and neck cancer protein encoding markers that may be used to distinguish group 3 from other groups include, but are not limited to: (Gene Symbol , Chromosome Band, Gene Title, UniProt ID): STRN3, 14q13-q21, striatin, calmodulin binding protein 3, Q13033; LRRC17, 7q22.1, leucine rich repeat containing 17, Q8N6Y2; FAM69A, 1p22, family with sequence similarity 69, member A, Q5T7M9; ATF2, 2q32, activating transcription factor 2, P15336; BHMT, 5q14.1, betaine--homocysteine S-methyltransferase, Q93088; ODZ3/TENM3, 4q34.3-q35.1, teneurin transmembrane protein 3, Q9P273; ZFX4, 8q21.11, zinc finger homeobox 4, Q86UP3. These markers may be identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from lung, head, and neck cancers, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0307] Protein markers that can be secreted by lung, head, and neck cancer into the blood may be used to distinguish group 3 from other groups include, but are not limited to: (Protein name , UniProt ID); Leucine-rich repeat-containing protein 17 (p37NB) , Q8N6Y2. A comparative analysis was performed across various TCGA datasets (tumors, normals), followed by an additional bioinformatics filter (Meinken et al., "Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report," *Computational Molecular Biology*

2(1):1-7 (2012), which is hereby incorporated by reference in its entirety), which predicts the likelihood that the translated protein is secreted by the cells.

[0308] The distribution of mutations in lung, head, and neck cancer are available in the public COSMIC database, with the most common being: CSMD3 (CUB and Sushi multiple domains 3), DNAH5 (dynein axonemal heavy chain 5), FAT1 (FAT atypical cadherin 1), FLG (filaggrin), KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), LRP1B (LDL receptor related protein 1B), MUC16 (mucin 16, cell surface associated), PCLO (piccolo presynaptic cytomatrix protein), PKHD1L1 (PKHD1 like 1), RELN (reelin), RYR2 (ryanodine receptor 2), SI (sucrase-isomaltase), SYNE1 (spectrin repeat containing nuclear envelope protein 1), TP53 (tumor protein p53), TTN (titin), USH2A (usherin), and XIRP2 (xin actin binding repeat containing 2).

[0309] Figure 49 provides a list of primary CpG sites that are lung, head, and neck cancer and tissue-specific markers, that may be used to identify the presence of lung, head, and neck cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 50 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are lung, head, and neck cancer and tissue-specific markers, that may be used to identify the presence of lung, head, and neck from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites that may be used to distinguish lung, head, and neck cancers. Primer sets for these preferred methylation markers are listed in Table 49 in the prophetic experimental section. A selection of 64 of these markers with average sensitivities of 50% gave the following scores for Group 3: (lung adenocarcinoma = 41, lung squamous cell carcinoma = 49, head & neck = 53), which would translate into the following number of markers equivalents with average sensitivities of 50% ($= 64 \times \text{score}/50$); (lung adenocarcinoma = 52 marker equivalents; lung squamous cell carcinoma = 62 marker equivalents; head & neck = 67 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum. The marker equivalents with average sensitivities of 66% ($= 64 \times \text{score}/66$); (lung adenocarcinoma = 40 marker equivalents; lung squamous cell carcinoma = 47 marker equivalents; head & neck = 51 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum.

[0310] Group 4 (prostate and bladder): Blood or urine-based, prostate and bladder cancer-specific microRNA markers may be used to distinguish group 4 from other groups include, but are not limited to: (mir ID, Gene ID): hsa-mir-491, MIR491; hsa-mir-1468, MIR1468. These markers were identified through analysis of TCGA microRNA datasets, and

may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood or urine.

[0311] Blood or urine-based, prostate and bladder cancer-specific ncRNA and lncRNA markers may be used to distinguish group 4 from other groups include, but are not limited to: [Gene ID , Coordinate (GRCh38) , ENSEMBL ID]: AC007383.3 , chr2:206084605-206086564 , ENSG00000227946.1; LINC00324 , chr17:8220642-8224043 , ENSG00000178977.3. These markers were identified through comparative analysis of various publicly available Affymetrix Exon ST CEL data, which were aligned to GENCODE annotations to generate ncRNA and lncRNA transcriptome datasets. Such lncRNA and ncRNA may be enriched in exosomes or other protected states in the blood or urine.

[0312] Blood or urine-based prostate and bladder cancer-specific exon transcripts may be used to distinguish group 4 from other groups include, but are not limited to: (Exon location, Gene); chr21:45555942-45556055:+ , C21orf33 and may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood or urine.

[0313] Prostate and bladder cancer protein markers that may be used to distinguish group 4 from other groups include, but are not limited to: (Gene Symbol , Chromosome Band , Gene Title , UniProt ID): PMM1 , 22q13 , phosphomannomutase 1 , Q92871. This marker may be identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from lung, head, and neck cancers, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma, or within the urine.

[0314] The distribution of mutations in prostate and bladder cancer are available in the public COSMIC database, with the most common being: BAGE2 (BAGE family member 2), DNM1P47 (dynamin 1 pseudogene 47), FRG1BP (region gene 1 family member B, pseudogene), KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), RP11-156P1.3, TTN (titin), and TUBB8P7 (tubulin beta 8 class VIII pseudogene 7).

[0315] Figure 51 provides a list of primary CpG sites that are prostate and bladder cancer-specific markers, that may be used to identify the presence of prostate and bladder cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood or urine. Figure 52 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are prostate and bladder cancer specific markers, that may be used to identify the presence of prostate and bladder from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood or urine. These lists contain preferred primary CpG sites and their flanking sites that may be used to distinguish prostate and bladder cancers. Primer sets for these preferred methylation markers are

listed in Table 50 in the prophetic experimental section. A selection of 48 of these markers with average sensitivities of 50% gave the following scores for Group 4: (prostate = 48, bladder = 22), which would translate into the following number of markers equivalents with average sensitivities of 50% ($= 48 \times \text{score}/50$); (prostate = 46 marker equivalents; bladder = 21 marker equivalents). Thus, bladder was below the average 36-marker equivalents minimum. Likewise, the marker equivalents with average sensitivities of 66% ($= 48 \times \text{score}/60$); (prostate = 35 marker equivalents; bladder = 16 marker equivalents). Thus, bladder was well below the average 36-marker equivalents minimum. However, a different selection of markers, for example by increasing from 48 to 64 markers and including markers that were positive for both prostate and bladder, would rectify this situation. The markers were limited to those that were not methylated in normal prostate, bladder, or kidney tissue to minimize false-positive results from urine samples.

[0316] Group 5 (liver, pancreatic and gall-bladder): Blood-based, liver, pancreatic and gall-bladder cancer-specific microRNA markers may be used to distinguish group 5 from other groups include, but are not limited to: (mir ID , Gene ID): hsa-mir-132, MIR132. This marker was identified through analysis of TCGA microRNA datasets, which may be present in exosomes, tumor-associated vesicles, Argonaute complexes, or other protected states in the blood.

[0317] Figure 53 provides a list of blood-based, liver, pancreatic and gall-bladder cancer-specific ncRNA and lncRNA markers derived through comparative analysis of various publicly available Affymetrix Exon ST CEL data, which were aligned to GENCODE annotations to generate ncRNA and lncRNA transcriptome datasets. Such lncRNA and ncRNA may be enriched in exosomes or other protected states in the blood.

[0318] In addition, Figure 54 provides a list of blood-based liver, pancreatic and gall-bladder cancer-specific exon transcripts that may be enriched in exosomes, tumor-associated vesicles, or other protected states in the blood.

[0319] Figure 55 provides a list of liver, pancreatic and gall-bladder cancer protein markers, identified through mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product arising from liver, pancreatic and gall-bladder cancers, which may be identified in the blood, either within exosomes, other protected states, tumor-associated vesicles, or free within the plasma.

[0320] Protein markers that can be secreted by liver, pancreatic and gall-bladder cancer into the blood may be used to distinguish group 5 from other groups include, but are not limited to: (Protein name , UniProt ID); Gelsolin (AGEL) (Actin-depolymerizing factor) (ADF) (Brevin) , P06396; Pro-neuregulin-2 , O14511; CD59 glycoprotein (1F5 antigen) (20 kDa homologous

restriction factor) (HRF-20) (HRF20) (MAC-inhibitory protein) (MAC-IP) (MEM43 antigen) (Membrane attack complex inhibition factor) (MACIF) (Membrane inhibitor of reactive lysis) (MIRL) (Protectin) (CD antigen CD59) , P13987; Divergent protein kinase domain 2B (Deleted in autism-related protein 1) , Q9H7Y0. A comparative analysis was performed across various TCGA datasets (tumors, normals), followed by an additional bioinformatics filter (Meinken et al., "Computational Prediction of Protein Subcellular Locations in Eukaryotes: an Experience Report," *Computational Molecular Biology* 2(1):1-7 (2012), which is hereby incorporated by reference in its entirety), which predicts the likelihood that the translated protein is secreted by the cells.

[0321] The distribution of mutations in liver, pancreatic and gall-bladder cancer are available in the public COSMIC database, with the most common being: KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), MUC16 (mucin 16, cell surface associated), MUC4 (mucin 4, cell surface associated), TP53 (tumor protein p53), and TTN (titin).

[0322] Figure 56 provides a list of primary CpG sites that are liver, pancreatic and gall-bladder cancer and tissue-specific markers, that may be used to identify the presence of lung, head, and neck cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 57 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are liver, pancreatic and gall-bladder cancer and tissue-specific markers, that may be used to identify the presence of liver, pancreatic and gall-bladder from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites, as well as alternative markers that may be used to distinguish liver, pancreatic and gall-bladder cancers. Primer sets for these preferred and alternative methylation markers are listed in Table 51 in the prophetic experimental section. A selection of 64 of these markers with average sensitivities of 50% gave the following scores for Group 5: (liver = 57, pancreatic = 30, gall bladder = 60), which would translate into the following number of marker equivalents with average sensitivities of 50% ($= 64 \times \text{score}/50$); (liver = 73 marker equivalents; pancreatic = 38 marker equivalents; gall bladder = 77 marker equivalents). Thus, all were above the average 36-marker equivalents minimum. The marker equivalents with average sensitivities of 66% ($= 64 \times \text{score}/66$); (liver = 56 marker equivalents; pancreatic = 29 marker equivalents; gall bladder = 58 marker equivalents). Thus, liver and gall bladder were above the average 36-marker equivalents minimum, while pancreatic was below.

[0323] Returning now to the strategy wherein in the first step is to identify markers that cover as many cancers as possible, irrespective of group, and yet are sufficiently diverse as to cover cancers in all 5 groups (Figures 1G, 1H, 1K and 1L), consider the aim is to have a very

sensitive detection of early cancer, where the first step of the assay would use a set of 64 markers that on average comprise of at least 36 markers with 75% sensitivity that covers each of the aforementioned 16 types of solid tumors (covered in the 5 Groups).

[0324] A deep analysis of the TCGA database of methylation markers that are absent in blood but on average are present in many solid tumor types at 75% sensitivity show a paucity of markers for pancreatic, lung adenocarcinoma, lung squamous cell carcinoma, and ovarian cancer. Consequently, to assemble a set of 64 markers that satisfied the criteria of at least 36 markers with 75% sensitivity that covers each of the aforementioned 16 types of solid tumors, the markers were focused on coverage of those cancers first, and bringing up the numbers for the other cancers by choosing markers that were well represented for other cancers as well, with the following average sensitivity scores: Group 1 (colorectal = 75, stomach = 68, esophagus = 72); Group 2 (breast = 66, endometrial = 73, ovarian = 54, cervical = 73, uterine = 67); Group 3 (lung adenocarcinoma = 54, lung squamous cell carcinoma = 58, head & neck = 64); Group 4 (prostate = 72, bladder = 63); and Group 5 (liver = 53, pancreatic = 45, gall bladder = 68). This translates into the following number of marker equivalents with average sensitivities of 75% ($= 64 \times \text{score}/75$): (colorectal = 64 marker equivalents; stomach = 58 marker equivalents; esophagus = 61 marker equivalents); Group 2 (breast = 57 marker equivalents; endometrial = 62 marker equivalents; ovarian = 46 marker equivalents; cervical = 62 marker equivalents; uterine = 57 marker equivalents); Group 3 (lung adenocarcinoma = 46 marker equivalents; lung squamous cell carcinoma = 49 marker equivalents; head & neck = 54 marker equivalents); Group 4 (prostate = 61 marker equivalents; bladder = 53 marker equivalents); and Group 5 (liver = 45 marker equivalents; pancreatic = 39 marker equivalents; gall bladder = 58 marker equivalents). Thus, cancers were well represented, ranging from 39 to 64 marker equivalents for the different cancer types, and all above the minimum of 36 markers with average sensitivities of 75%.

[0325] Figure 58 provides a list of primary CpG sites that are solid tumors and tissue-specific markers, that may be used to identify the presence of solid tumors from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 59 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are solid tumors and tissue-specific markers, that may be used to identify the presence of solid tumors from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites, as well as preferred alternative markers, and additional alternative markers that are high in multiple cancers. Primer sets for these preferred and alternative methylation markers are listed in Table 52 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety in the prophetic experimental section thereof. These

primers are not designed to identify specific types of tissue of origin, but simply determine with reasonable sensitivity and specificity (see below) if the patient has a hidden early cancer within. Those patients with 5 or more markers positive are then automatically subjected to additional tests to determine most probable tissue of origin. As written earlier, one approach is to continue with the set of markers already described, i.e. use the 96-marker set that on average comprise of at least 36 markers with 50% sensitivity for each tumor type (Figure 1G or 1H). Another approach would be to start with the set of 96 markers that on average comprise of at least 36 markers with 50% sensitivity, and then in step 2 continue with 1-2 sets of 48 group-specific markers that on average comprise of at least 36 markers with 75% sensitivity that covers each of the aforementioned types of solid tumors that may be present in that group (Figure 1F). By scoring the markers that are positive and comparing to predicted positives for each cancer type within the group tested, the physician can identify the most probable tissue of origin, and subsequently send the patient to the appropriate imaging.

[0326] For the second step of the assay in Figure 1F, one to two of the following groups will be tested, each group with a set of 48 markers that on average comprise at least 36 markers with 75% sensitivity that covers each of the aforementioned 16 types of solid tumor. These high percentage hit markers are also ideally suited for monitoring treatment efficacy and recurrence (see below). Tumors were in the following groups: Group 1 (colorectal, stomach, and esophagus); Group 2 (breast, endometrial, ovarian, cervical, and uterine); Group 3 (lung and head & neck); Group 4 (prostate and bladder); and Group 5 (liver, pancreatic, or gall bladder). These Group-specific and cancer type-specific markers include, but are not limited to, cancer-specific microRNA markers, cancer-specific ncRNA and lncRNA markers, cancer-specific exon transcripts, tumor-associated antigens, cancer protein markers, protein markers that can be secreted by solid tumors into the blood, common mutations, primary CpG sites that are solid tumor and tissue specific markers, chromosomal regions or sub-regions within which are primary CpG sites that are solid tumor and tissue specific markers, and primary and flanking CpG sites that are solid tumor and tissue specific markers. Methods for detecting said markers have been discussed earlier in this application, and figures listing these markers are described for each of the groups below.

[0327] Figure 60 provides a list of primary CpG sites that are colorectal, stomach, and esophageal cancer and tissue-specific markers, that may be used to identify the presence of colorectal, stomach, and esophageal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 61 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are colorectal, stomach, and esophageal cancer and tissue-specific markers, that may be used to identify the

presence of colorectal, stomach, and esophageal cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites, as well as alternative markers that are high in colorectal, stomach and esophageal cancers. Primer sets for these preferred and alternative methylation markers are listed in Table 53 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety, in the prophetic experimental section thereof. A selection of 48 of these markers with average sensitivities of 75% gave the following scores for Group 1: (colorectal = 87, stomach = 72, esophagus = 75). This translates into the following number of marker equivalents with average sensitivities of 75% ($= 48 \times \text{score}/75$); (colorectal = 56 marker equivalents; stomach = 46 marker equivalents; esophagus = 48 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum.

[0328] Figure 62 provides a list of primary CpG sites that are breast, endometrial, ovarian, cervical, and uterine cancer and tissue-specific markers, that may be used to identify the presence of breast, endometrial, ovarian, cervical, and uterine cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 63 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are breast, endometrial, ovarian, cervical, and uterine cancer and tissue-specific markers, that may be used to identify the presence of breast, endometrial, ovarian, cervical, and uterine cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites that may be used to distinguish breast, endometrial, ovarian, cervical, and uterine cancers. Primer sets for these preferred methylation markers are listed in Table 54 in the prophetic experimental section. A selection of 48 of these markers with average sensitivities of 75% gave the following scores for Group 2: (breast = 70, endometrial = 85, ovarian = 70, cervical = 75, uterine = 83). This would translate into the following number of marker equivalents with average sensitivities of 75% ($= 48 \times \text{score}/75$); (breast = 45 marker equivalents; endometrial = 54 marker equivalents; ovarian = 45 marker equivalents; cervical = 48 marker equivalents; uterine = 53 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum.

[0329] Figure 64 provides a list of primary CpG sites that are lung, head, and neck cancer and tissue-specific markers, that may be used to identify the presence of lung, head, and neck cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 65 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are lung, head, and neck cancer and tissue-specific markers,

that may be used to identify the presence of lung, head, and neck from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites that may be used to distinguish lung, head, and neck cancers. Primer sets for these preferred methylation markers are listed in Table 55 in the prophetic experimental section. A selection of 48 of these markers with average sensitivities of 75% gave the following scores for Group 3: (lung adenocarcinoma = 62, lung squamous cell carcinoma = 67, head & neck = 69). This would translate into the following number of marker equivalents with average sensitivities of 75% ($= 48 \times \text{score}/75$); (lung adenocarcinoma = 39 marker equivalents; lung squamous cell carcinoma = 43 marker equivalents; head & neck = 44 marker equivalents). Thus, all were well above the average 36-marker equivalents minimum.

[0330] Figure 66 provides a list of primary CpG sites that are prostate and bladder cancer-specific markers, that may be used to identify the presence of prostate and bladder cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood or within the urine. Figure 67 provides a list of chromosomal regions or sub-regions within which are primary CpG sites that are prostate and bladder cancer specific markers, that may be used to identify the presence of prostate and bladder from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood or urine. These lists contain preferred primary CpG sites and their flanking sites that may be used to distinguish prostate and bladder cancers. Primer sets for these preferred methylation markers for prostate, bladder and kidney cancer from a blood sample are listed in Table 56A in the prophetic experimental section. Primer sets for these preferred methylation markers for prostate and bladder cancer from a urine sample are listed in Table 56B of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety, in the prophetic experimental section thereof. Most of the kidney-specific methylation markers are found in normal kidney tissue, and thus these would not be suitable for use in a urine test. A selection of 48 of these markers with average sensitivities of 75% gave the following scores for Group 4: (prostate = 70, bladder = 66), which would translate into the following number of markers equivalents with average sensitivities of 75% ($= 48 \times \text{score}/75$); (prostate = 45 marker equivalents; bladder = 42 marker equivalents), Thus, well above the average 36-marker equivalents minimum.

[0331] Figure 68 provides a list of primary CpG sites that are liver, pancreatic and gall-bladder cancer and tissue-specific markers, that may be used to identify the presence of lung, head, and neck cancer from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. Figure 69 provides a list of chromosomal regions or

sub-regions within which are primary CpG sites that are liver, pancreatic and gall-bladder cancer and tissue-specific markers, that may be used to identify the presence of liver, pancreatic and gall-bladder from cfDNA, or DNA within exosomes, or DNA in other protected states (such as within CTCs) within the blood. These lists contain preferred primary CpG sites and their flanking sites that may be used to distinguish liver, pancreatic and gall-bladder cancers. Primer sets for these preferred methylation markers are listed in Table 57 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety, in the prophetic experimental section thereof. A selection of 64 of these markers with average sensitivities of 75% gave the following scores for Group 5: (liver = 68, pancreatic = 58, gall bladder = 74). This would translate into the following number of marker equivalents with average sensitivities of 75% ($= 48 \times \text{score}/75$); (liver = 43 marker equivalents; pancreatic = 37 marker equivalents; gall bladder = 48 marker equivalents). Thus, all were above the average 36-marker equivalents minimum.

[0332] The aforementioned markers with average sensitivities of 75% may also be used to monitor recurrence in Melanoma. Primer sets for exemplary preferred and alternate methylation markers are listed in Table 58 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety, in the prophetic experimental section thereof.

[0333] How would the different Pan-oncology tests compare with each other for detecting either colorectal cancer (both sexes) or ovarian cancer in the U.S.? For the illustrative examples below: (1) The pan-oncology test of 96 markers with average sensitivities of 50% wherein >36 markers/group are positive with each cancer. For colorectal = 84 marker equivalents; for ovarian = 42 marker equivalents. The marker false-positive rates of 3%, for colorectal cancer will be calculated at 48 markers, while for ovarian cancer will be calculated at 36 markers, with a minimum of 5 positives to go to Step 2 or imaging. (2) The pan-oncology test of 64 markers with average sensitivities of 75% wherein >36 markers/group are positive with each cancer. For colorectal = 64 marker equivalents; for ovarian = 46 marker equivalents. The marker false-positive rate of 3%, for colorectal cancer will be calculated at 48 markers, while for ovarian cancer will be calculated at 36 markers, with a minimum of 5 positives to go to Step 2. (3) The group-specific step of 64 markers with average sensitivities of 50% wherein >36 markers/group are positive with each cancer. For colorectal = 61 marker equivalents; for ovarian = 41 marker equivalents. The marker false-positive rates of 3%, for colorectal cancer will be calculated at 48 markers, while for ovarian cancer will be calculated at 36 markers, with a minimum of 5 positives to go to imaging. (4) The group-specific step of 48 markers with average sensitivities of 75% wherein >36 markers/group are positive with each cancer. For

colorectal = 56 marker equivalents; for ovarian = 45 marker equivalents. The marker false-positive rates of 3%, for colorectal cancer will be calculated at 48 markers, while for ovarian cancer will be calculated at 36 markers, with a minimum of 5 positives to go to imaging.

[0334] Consider the first strategy using the 96 pan-oncology markers of detecting early colorectal cancer (Figure 1C). The calculations are done with the anticipation of an average of 150 methylated molecules per positive marker in the blood. As described *supra*, for the example of colorectal cancer, in particular the cases of microsatellite stable tumors (MSS) where the mutation load is low, for these calculations when relying on NGS sequencing alone (assuming 150 molecules with one mutation in the blood), an estimated 78% of early colorectal cancer would be missed. Again, to put these number in perspective, in the U.S., about 135,000 new cases of colorectal cancer were diagnosed in 2018, of which about 60% is late cancer (*i.e.* Stage III & IV). About 107 million individuals in the U.S. are over the age of 50 and should be tested for colorectal cancer. While it cannot be predicted how many individuals have a hidden cancer (*i.e.* Stage I) within them, who are non-compliant to testing, for the purposes of this calculation, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. Assuming individual marker false-positive rates of 3%, and with the first step using 96 markers (48 markers for CRC) with average sensitivities of 50%, requiring a minimum of 5 markers positive, then with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.) which would include at 71.6% sensitivity or about 28,998 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 64 markers (48 markers for CRC) with average sensitivities of 50%, requiring a minimum of 5 markers positive, then the two-step test would identify $71.6\% \times 71.6\% = 51.2\% = 20,762$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $20,762 / (188,748 + 20,762) = 9.9\%$, in other words, 1 in 10 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker, and the higher stages. Also, to be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, than an estimate of 40,500 individuals with Stage I cancer, 40,500 individuals with

Stage II cancer, and the remaining 54,000 individuals have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 90.1% would be identified in the first step, of which $90.1\% \times 90.1\% = 81.0\% = 32,877$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.3 % would be identified in the first step, of which $99.3\% \times 99.3\% = 98.6\% = 53,246$ individuals with late cancer would be identified. This brings the total identified at $20,762 + 32,877 + 53,246 = 106,885$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 79%. Overall, the positive predictive value of such a test would be $106,885 / (188,748 + 106,885) = 36.1\%$, in other words, 1 in 3 individuals who tested positive would actually have colorectal cancer, and this test would identify $53,639 / 81,000$ or 66% of those individuals with early cancer, compared with the current rate of 40%.

[0335] How would these results vary for using this strategy (Figure 1E) for detection of early colorectal cancer using 50% average marker sensitivities, with the anticipation of Stage I CRC has an average of 200 methylated molecules per positive marker in the blood, Stage II CRC has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0336] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average sensitivities of 50%, requiring a minimum of 5 markers positive, and an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.). This would include, at 90.1% sensitivity, or about 36,490 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 64 markers (48 markers for CRC) with average sensitivities of 50%, requiring a minimum of 5 markers positive. The two-step test would identify $90.1\% \times 90.1\% = 81.2\% = 32,877$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $32,877 / (188,748 + 32,877) = 14.8\%$. In other words, 1 in 6.7 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. To be consistent with the idea that, as the test is used repeatedly, more of early and less of late CRC will be detected, an estimated 40,500 individuals with Stage I cancer, 40,500 individuals with Stage II cancer, and 54,000 individuals with late-stage cancer (135,000 total individuals with colorectal cancer) would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For

Stage II cancer, 90.1% would be identified in the first step, of which $97.2\% \times 97.2\% = 94.5\% = 38,263$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.3 % would be identified in the first step, of which $99.3\% \times 99.3\% = 98.6\% = 53,246$ individuals with late cancer would be identified. This brings the total identified to $32,877 + 38,263 + 53,246 = 124,386$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 92.1%. Overall, the positive predictive value of such a test would be $124,386 / (188,748 + 124,386) = 39.7\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test would identify $71,104 / 81,000$ or 87.7% of those individuals with early cancer, compared with the current rate of 40%.

[0337] How would these results vary for using this strategy (Figure 1C) for detection of early ovarian cancer, with the anticipation of an average of 150 methylated molecules per positive marker in the blood? When relying on NGS sequencing alone (assuming 150 molecules with one mutation in the blood), an estimated 78% of early ovarian cancer would be missed. Again, to put these numbers in perspective, in the U.S., about 22,000 new cases of ovarian cancer were diagnosed in 2018, of which about 85% was late cancer (*i.e.* Stage III & IV). About 54 million women in the U.S. are between the ages of 50 and 79 and should be tested for ovarian cancer. While it cannot be predicted how many individuals have a hidden cancer (*i.e.* Stage I), for the purposes of this calculation, assume that the stages are evenly divided. Thus, the number of individuals with Stage I ovarian cancer would be about 5,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for ovarian) with average sensitivities of 50%, requiring a minimum of 5 markers positive, and an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.) with ovarian cancer. This would include, at 46.8% sensitivity, or about 2,574 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 64 markers (36 markers for ovarian cancer) with average sensitivities of 50%, requiring a minimum of 5 markers positive. The two-step test would identify $46.8\% \times 46.8\% = 21.9\% = 1,204$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $1,204 / (4,374 + 1,204) = 21.6\%$. In other words, 1 in 4.6 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. In expanding this example, the calculations are done with the anticipation that Stage I ovarian cancer has an average of 150 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of

200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. To be consistent with the idea that, as the test is used repeatedly, more cancer will be detected and all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S.. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 71.5% would be identified in the first step, of which $71.5\% \times 71.5\% = 51.1\% = 2,810$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 94.5% would be identified in the first step, of which $94.5\% \times 94.5\% = 89.3\% = 9,823$ individuals with late ovarian cancer would be identified. This brings the total identified at $1,204 + 2,810 + 9,823 = 13,837$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 62.9%. Overall, the positive predictive value of such a test would be $13,837 / (13,837 + 4,374) = 76.0\%$. In other words, 3 in 4 women who tested positive would actually have ovarian cancer, and this test would identify $4,014 / 11,000$, or 36.5%, of those individuals with early cancer, compared with the current rate of 15%.

[0338] How would these results vary for using this strategy (Figure 1E) for detection of early ovarian cancer using 50% average marker sensitivities, with the anticipation that Stage I ovarian cancer has an average of 200 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker, and the higher stages?

[0339] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for ovarian) with average sensitivities of 50%, and requiring a minimum of 5 markers positive, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.) with ovarian cancer. This would include at, 71.5% sensitivity, about 3,932 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 64 markers (36 markers for ovarian cancer) with average sensitivities of 50%, requiring a minimum of 5 markers positive. The two-step test would identify $71.5\% \times 71.5\% = 51.1\% = 2,810$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $2,810 / (4,374 + 2,810) = 39.1\%$. In other words, 1 in 2.5 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. As the test is used repeatedly, assume all four stages are at 5,500, and, therefore, $5,500 \times 4 = 22,000$ total individuals with

ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 84.4% would be identified in the first step, of which $84.4\% \times 84.4\% = 71.2\% = 3,916$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 94.5% would be identified in the first step, of which $94.5\% \times 94.5\% = 89.3\% = 9,823$ individuals with late ovarian cancer would be identified. This brings the total identified to $2,810 + 3,916 + 9,823 = 16,549$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 75.2%. Overall, the positive predictive value of such a test would be $16,549 / (16,549 + 4,374) = 79.0\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer. This test would identify 6,006 / 11,000 or 54.6% of those individuals with early cancer, compared with the current rate of 15%.

[0340] Consider the second strategy using the 96 pan-oncology markers of detecting early colorectal cancer (Figure 1F). The calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. As before, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average sensitivities of 50%, requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.). This would include, at 71.6% sensitivity, about 28,998 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 48 markers (48 markers for CRC) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $71.6\% \times 94.5\% = 67.6\% = 27,403$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $27,403 / (188,748 + 27,403) = 12.6\%$. In other words, 1 in 8 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker, and the higher stages. Also, to be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals with Stage I cancer,

40,500 individuals with Stage II cancer, and the remaining 54,000 individuals have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 90.1% would be identified in the first step, of which $90.1\% \times 99.2\% = 89.3\% = 36,198$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.3% would be identified in the first step, of which $99.3\% \times 99.9\% = 99.2\% = 53,568$ individuals with late cancer would be identified. This brings the total identified at $27,403 + 36,198 + 53,568 = 117,169$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 87%. Overall, the positive predictive value of such a test would be $117,169 / (188,748 + 117,169) = 38.3\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test would identify $63,601 / 81,000$ or 78.5% of those individuals with early cancer, compared with the current rate of 40%.

[0341] How would these results vary for using this strategy (Figure 1F) for detection of early colorectal cancer using 50% average marker sensitivities, with the anticipation of Stage I CRC has an average of 200 methylated molecules per positive marker in the blood, Stage II CRC has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0342] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average sensitivities of 50%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.). This would include, at 90.1% sensitivity, about 36,490 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 48 markers (48 markers for CRC) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $90.1\% \times 99.2\% = 89.4\% = 36,198$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $36,198 / (188,748 + 36,198) = 16.1\%$. In other words, 1 in 6.2 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. To be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals with Stage I cancer, 40,500 individuals with Stage II cancer, and the remaining 54,000 individuals have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S.. The above calculation already provided the false-positive

rate for the early cancer. For Stage II cancer, 97.2% would be identified in the first step, of which $97.2\% \times 99.9\% = 97.1\% = 39,325$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.3 % would be identified in the first step, of which $99.3\% \times 99.9\% = 99.2\% = 53,568$ individuals with late cancer would be identified. This brings the total identified at $36,198 + 39,325 + 53,568 = 129,091$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 95.6%. Overall, the positive predictive value of such a test would be $129,091 / (188,748 + 129,091) = 40.6\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test would identify 75,523 / 81,000 or 93.2% of those individuals with early cancer, compared with the current rate of 40%.

[0343] How would these results vary for using the second strategy (Figure 1F) for detection of early ovarian cancer, with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood? Again, assume individuals with Stage I ovarian cancer would be about 5,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for ovarian) with average sensitivities of 50%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.). This would include, at 46.8% sensitivity, about 2,574 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 48 markers (36 markers for ovarian cancer) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $46.8\% \times 80.3\% = 37.6\% = 2,068$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $2,068 / (4,374 + 2,068) = 32.1\%$. In other words, 1 in 3.1 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. In expanding this example, the calculations are done with the anticipation that Stage I ovarian cancer has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II ovarian cancer has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, assuming all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 71.5% would be identified in the first step, of which $71.5\% \times 94.5\% = 67.6\% = 3,718$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and

IV ovarian cancer, 94.5% would be identified in the first step, of which $94.5\% \times 99.7\% = 94.2\%$ = 10,363 individuals with late ovarian cancer would be identified. This brings the total identified at $2,068 + 3,718 + 10,363 = 16,149$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 73.1%. Overall, the positive predictive value of such a test would be $16,149 / (16,149 + 4,374) = 78.7\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify 5,786 / 11,000 or 52.6% of those individuals with early cancer, compared with the current rate of 15%.

[0344] How would these results vary for using this strategy (Figure 1F) for detection of early ovarian cancer using 50% average marker sensitivities, with the anticipation of Stage I ovarian cancer has an average of 200 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0345] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for Ovarian) with average sensitivities of 50%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.). This would include, at 71.5% sensitivity, about 3,932 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 48 markers (36 markers for ovarian cancer) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $71.5\% \times 94.5\% = 67.6\% = 3,718$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $3,718 / (4,374 + 3,718) = 45.9\%$. In other words, 1 in 2.2 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. Assuming all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 84.4% would be identified in the first step, of which $84.4\% \times 98.3\% = 82.9\% = 4,559$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 94.5% would be identified in the first step, of which $94.5\% \times 99.7\% = 94.2\% = 10,363$ individuals with late ovarian cancer would be identified. This brings the total identified at $3,718 + 4,559 + 10,363 = 18,640$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 84.7%. Overall, the positive predictive value of such a test

would be $18,640 / (18,640 + 4,374) = 81.0\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify $8,277 / 11,000$ or 75.2% of those individuals with early cancer, compared with the current rate of 15% .

[0346] Finally, consider the third strategy using the 64 pan-oncology markers of detecting early colorectal cancer (Figure 1H). The calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. As before, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. Assuming individual marker false-positive rates of 3% , the first step using 64 markers (48 markers for CRC) with average sensitivities of 75% , and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8% , the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.). This would include, at 94.5% sensitivity, about 38,272 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 96 markers (48 markers for CRC) with average sensitivities of 50% , requiring a minimum of 5 markers positive. The two-step test would identify $94.5\% \times 71.6\% = 67.6\% = 27,403$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8% , the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $27,403 / (188,748 + 27,403) = 12.6\%$. In other words, 1 in 8 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, to be consistent with the idea that, as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals with Stage I cancer would be identified, 40,500 individuals with Stage II cancer would be identified, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 99.2% would be identified in the first step, of which $99.2\% \times 90.1\% = 89.3\% = 36,198$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9% would be identified in the first step, of which $99.9\% \times 99.3\% = 99.2\% = 53,568$ individuals with late cancer would be identified. This brings the total identified at $27,403 + 36,198 + 53,568 = 117,169$ individuals out of 135,000 with colorectal

cancer, for an overall sensitivity of 87%. Overall, the positive predictive value of such a test would be $117,169 / (188,748 + 117,169) = 38.3\%$. In other words, 2 in 5 individuals who tested positive would actually have colorectal cancer, and this test would identify 63,601 / 81,000 or 78.5% of those individuals with early cancer, compared with the current rate of 40%.

[0347] How would these results vary for using this strategy (Figure 1H) for detection of early colorectal cancer using 75% average marker sensitivities, with the anticipation of Stage I CRC having an average of 200 methylated molecules per positive marker in the blood, Stage II CRC having an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) having at least an average of 300 methylated molecules per positive marker?

[0348] Assuming individual marker false-positive rates of 3%, the first step using 64 markers (48 markers for CRC) with average sensitivities of 75%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.). This would include, at 99.2% sensitivity, about 40,176 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 96 markers (48 markers for CRC) with average sensitivities of 50%, requiring a minimum of 5 markers positive. The two-step test would identify $99.2\% \times 90.1\% = 89.3\% = 36,198$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $36,198 / (188,748 + 36,198) = 16.1\%$. In other words, 1 in 6.2 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. To be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals would have Stage I cancer, 40,500 individuals would have Stage II cancer, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 99.9% would be identified in the first step, of which $99.9\% \times 97.2\% = 97.1\% = 39,325$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9 % would be identified in the first step, of which $99.9\% \times 99.3\% = 99.2\% = 53,568$ individuals with late cancer would be identified. This brings the total identified at $36,198 + 39,325 + 53,568 = 129,091$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 95.6%. Overall, the positive predictive value of such a test would be $129,091 / (188,748 + 129,091) = 40.6\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test

would identify 75,523 / 81,000 or 93.2% of those individuals with early cancer, compared with the current rate of 40%.

[0349] How would these results vary for using the third strategy (Figure 1H) for detection of early ovarian cancer, with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood? Again, assume individuals with Stage I ovarian cancer would be about 4,500 individuals. Assuming individual marker false-positive rates of 3%, and the first step using 64 markers (36 markers for ovarian) with average sensitivities of 75%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.). This would include, at 80.3% sensitivity, about 4,416 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 96 markers (36 markers for ovarian cancer) with average sensitivities of 50%, requiring a minimum of 5 markers positive. The two-step test would identify $80.3\% \times 46.8\% = 37.6\% = 2,068$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $2,068 / (4,374 + 2,068) = 32.1\%$. In other words, 1 in 3.1 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. In expanding this example, the calculations are done with the anticipation that Stage I ovarian cancer has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II ovarian cancer has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 94.5% would be identified in the first step, of which $94.5\% \times 71.5\% = 67.6\% = 3,718$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.7% would be identified in the first step, of which $99.7\% \times 94.5\% = 94.2\% = 10,363$ individuals with late ovarian cancer would be identified. This brings the total identified at $2,068 + 3,718 + 10,363 = 16,149$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 73.1%. Overall, the positive predictive value of such a test would be $16,149 / (16,149 + 4,374) = 78.7\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify 5,786 / 11,000 or 52.6% of those individuals with early cancer, compared with the current rate of 15%.

[0350] How would these results vary for using this strategy (Figure 1H) for detection of early ovarian cancer using 75% average marker sensitivities, with the anticipation of Stage I ovarian cancer having an average of 200 methylated molecules per positive marker in the blood, Stage II ovarian cancer having an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) having at least an average of 300 methylated molecules per positive marker?

[0351] Assuming individual marker false-positive rates of 3%, the first step using 64 markers (36 markers for ovarian) with average sensitivities of 75%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.). This would include, at 94.5% sensitivity, about 5,197 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 96 markers (36 markers for ovarian cancer) with average sensitivities of 50%, requiring a minimum of 5 markers positive. The two-step test would identify $94.5\% \times 71.5\% = 67.6\% = 3,718$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $3,718 / (4,374 + 3,718) = 45.9\%$. In other words, 1 in 2.2 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 84.4% would be identified in the first step, of which $98.3\% \times 84.4\% = 82.9\% = 4,559$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 94.5% would be identified in the first step, of which $99.7\% \times 94.5\% = 94.2\% = 10,363$ individuals with late ovarian cancer would be identified. This brings the total identified at $3,718 + 4,559 + 10,363 = 18,640$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 84.7%. Overall, the positive predictive value of such a test would be $18,640 / (18,640 + 4,374) = 81.0\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify $8,277 / 11,000$ or 75.2% of those individuals with early cancer, compared with the current rate of 15%.

[0352] The above calculations worked under the assumption of limiting at least one set of markers to an average of 50% sensitivities. How would the results improve is the average of 50% sensitivities was improved to 66% sensitivities?

[0353] Consider the first strategy using the 96 pan-oncology markers of detecting early colorectal cancer (Figure 1I). The calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. As previously, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the US). This would include, at 90.0% sensitivity, about 36,450 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 64 markers (48 markers for CRC) with average sensitivities of 66%, requiring a minimum of 5 markers positive. Then the two-step test would identify $90.0\% \times 90.0\% = 81.0\% = 32,805$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $32,805 / (188,748 + 32,805) = 14.8\%$. In other words, 1 in 7 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, to be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals would be identified with Stage I cancer, 40,500 individuals would be identified with Stage II cancer, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the US. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 98.0% would be identified in the first step, of which $98.0\% \times 98.0\% = 96.0\% = 38,896$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.6 % would be identified in the first step, of which $99.6\% \times 99.6\% = 99.2\% = 53,568$ individuals with late cancer would be identified. This brings the total identified at $32,805 + 38,896 + 53,568 = 125,269$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 92.7%. Overall, the positive predictive value of such a test would be $125,269 / (188,748 + 125,269) = 39.9\%$. In other words, 1 in 2.5 individuals who

tested positive would actually have colorectal cancer, and this test would identify 71,701 / 81,000 or 88% of those individuals with early cancer, compared with the current rate of 40%.

[0354] How would these results vary for using this strategy (Figure 1I) for detection of early colorectal cancer using 66% average marker sensitivities, with the anticipation of Stage I CRC has an average of 200 methylated molecules per positive marker in the blood, Stage II CRC has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0355] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S). This would include, at 98.0% sensitivity, about 39,690 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 64 markers (48 markers for CRC) with average sensitivities of 66%, requiring a minimum of 5 markers positive. The two-step test would identify $98.0\% \times 98.0\% = 96.0\% = 38,896$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $38,896 / (188,748 + 38,896) = 17.81\%$. In other words, 1 in 6 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. To be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals would be identified with Stage I cancer, 40,500 individuals would be identified with Stage II cancer, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 98.0% would be identified in the first step, of which $99.6\% \times 99.6\% = 99.2\% = 40,176$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9 % would be identified in the first step, of which $99.9\% \times 99.9\% = 99.8\% = 53,568$ individuals with late cancer would be identified. This brings the total identified at $38,896 + 40,176 + 53,892 = 132,964$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 98.5%. Overall, the positive predictive value of such a test would be $132,964 / (188,748 + 132,964) = 41.3\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test would identify 79,072 / 81,000 or 97.6% of those individuals with early cancer, compared with the current rate of 40%.

[0356] How would these results vary for using the first strategy (Figure 1I) for detection of early ovarian cancer, with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood? Again, assume that the stages are evenly divided, and thus individuals with Stage I ovarian cancer would be about 5,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for ovarian) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S). This would include, at 71.5% sensitivity, about 3,932 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 64 markers (36 markers for ovarian cancer) with average sensitivities of 66%, requiring a minimum of 5 markers positive. The two-step test would identify $71.5\% \times 71.5\% = 51.1\% = 2,810$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $2,810 / (4,374 + 2,810) = 39.1\%$. In other words, 1 in 2.5 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. In expanding this example, the calculations are done with the anticipation that Stage I ovarian cancer has an average of 150 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 90.0% would be identified in the first step, of which $90.0\% \times 90.0\% = 81.0\% = 4,485$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.2% would be identified in the first step, of which $99.2\% \times 99.2\% = 98.4\% = 10,824$ individuals with late ovarian cancer would be identified. This brings the total identified at $2,810 + 4,485 + 10,824 = 18,119$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 82.4%. Overall, the positive predictive value of such a test would be $18,119 / (18,119 + 4,374) = 80.5\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify $7,295 / 11,000$ or 66.3% of those individuals with early cancer, compared with the current rate of 15%.

[0357] How would these results vary for using this strategy (Figure 1I) for detection of early ovarian cancer using 66% average marker sensitivities, with the anticipation that Stage I

ovarian cancer has an average of 200 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker, and the higher stages?

[0358] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for ovarian) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S). This would include, at 90.0% sensitivity, about 4,950 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 64 markers (36 markers for ovarian cancer) with average sensitivities of 66%, requiring a minimum of 5 markers positive. The two-step test would identify $90.0\% \times 90.0\% = 81.0\% = 4,895$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $4,895 / (4,374 + 4,895) = 52.8\%$. In other words, 1 in 2 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 96.2% would be identified in the first step, of which $96.2\% \times 96.2\% = 92.5\% = 5,087$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.2% would be identified in the first step, of which $99.2\% \times 99.2\% = 98.4\% = 10,824$ individuals with late ovarian cancer would be identified. This brings the total identified at $4,895 + 5,087 + 10,824 = 20,806$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 94.6%. Overall, the positive predictive value of such a test would be $20,806 / (20,806 + 4,374) = 87.4\%$. In other words, 7 in 8 women who tested positive would actually have ovarian cancer, and this test would identify 9,982 / 11,000 or 90.1% of those individuals with early cancer, compared with the current rate of 15%.

[0359] Consider the second strategy using the 96 pan-oncology markers of detecting early colorectal cancer (Figure 1J). The calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. As before, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average

sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S). This would include, at 90.0% sensitivity, about 36,450 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 48 markers (48 markers for CRC) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $90.0\% \times 94.5\% = 85.0\% = 34,445$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $34,445 / (188,748 + 34,445) = 15.4\%$. In other words, 1 in 6.5 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, to be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals with Stage I cancer, 40,500 individuals with Stage II cancer, and the remaining 54,000 individuals have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 98.0% would be identified in the first step, of which $98.1\% \times 99.2\% = 97.3\% = 39,412$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9 % would be identified in the first step, of which $99.9\% \times 99.9\% = 99.8\% = 53,892$ individuals with late cancer would be identified. This brings the total identified at $34,445 + 39,412 + 53,892 = 127,749$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 94.6%. Overall, the positive predictive value of such a test would be $127,749 / (188,748 + 127,749) = 40.3\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test would identify $73,857 / 81,000$ or 91.2% of those individuals with early cancer, compared with the current rate of 40%.

[0360] How would these results vary for using this strategy (Figure 1J) for detection of early colorectal cancer using 66% average marker sensitivities, with the anticipation of Stage I CRC has an average of 200 methylated molecules per positive marker in the blood, Stage II CRC has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0361] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (48 markers for CRC) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S). This would include, at 98.0% sensitivity, about 39,690 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 48 markers (48 markers for CRC) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $98.0\% \times 99.2\% = 97.2\% = 39,372$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $39,372 / (188,748 + 39,372) = 17.2\%$. In other words, 1 in 5.8 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. To be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals would be identified with Stage I cancer, 40,500 individuals would be identified with Stage II cancer, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 99.6% would be identified in the first step, of which $99.6\% \times 99.9\% = 99.5\% = 40,297$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9 % would be identified in the first step, of which $99.9\% \times 99.9\% = 99.8\% = 53,892$ individuals with late cancer would be identified. This brings the total identified at $39,372 + 40,297 + 53,892 = 133,561$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 98.9%. Overall, the positive predictive value of such a test would be $133,561 / (188,748 + 133,561) = 41.4\%$. In other words, 1 in 2.4 individuals who tested positive would actually have colorectal cancer, and this test would identify $79,699 / 81,000$ or 98.4% of those individuals with early cancer, compared with the current rate of 40%.

[0362] How would these results vary for using the second strategy (Figure 1J) for detection of early ovarian cancer, with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood? Again, assume individuals with Stage I ovarian cancer would be about 5,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for Ovarian) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total

women ages 50-79 in the U.S). This would include, at 71.5% sensitivity, about 3,932 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 48 markers (36 markers for ovarian cancer) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $71.5\% \times 80.3\% = 57.4\% = 3,157$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $3,157 / (4,374 + 3,157) = 41.9\%$. In other words, 1 in 2.4 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. In expanding this example, the calculations are done with the anticipation that Stage I ovarian cancer has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II ovarian cancer has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 90.0% would be identified in the first step, of which $90.0\% \times 94.5\% = 85\% = 4,675$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.2% would be identified in the first step, of which $99.2\% \times 99.7\% = 98.9\% = 10,879$ individuals with late ovarian cancer would be identified. This brings the total identified at $3,157 + 4,675 + 10,879 = 18,711$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 85.1%. Overall, the positive predictive value of such a test would be $18,711 / (18,711 + 4,374) = 81.1\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify $7,832 / 11,000$ or 71.2% of those individuals with early cancer, compared with the current rate of 15%.

[0363] How would these results vary for using this strategy (Figure 1F) for detection of early ovarian cancer using 66% average marker sensitivities, with the anticipation of Stage I ovarian cancer has an average of 200 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0364] Assuming individual marker false-positive rates of 3%, the first step using 96 markers (36 markers for ovarian) with average sensitivities of 66%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify

486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S). This would include, at 90.0% sensitivity, about 4,950 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 48 markers (36 markers for ovarian cancer) with average sensitivities of 75%, requiring a minimum of 5 markers positive. The two-step test would identify $90.0\% \times 94.5\% = 85\% = 4,675$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $4,675 / (4,374 + 4,675) = 51.6\%$. In other words, 1 in 2 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 96.2% would be identified in the first step, of which $96.2\% \times 98.3\% = 94.5\% = 5,201$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.2% would be identified in the first step, of which $99.2\% \times 99.7\% = 98.9\% = 10,879$ individuals with late ovarian cancer would be identified. This brings the total identified at $4,675 + 5,201 + 10,879 = 20,755$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 94.3%. Overall, the positive predictive value of such a test would be $20,755 / (20,755 + 4,374) = 82.6\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify 9,876 / 11,000 or 89.8% of those individuals with early cancer, compared with the current rate of 15%.

[0365] Finally, consider the third strategy using the 64 pan-oncology markers of detecting early colorectal cancer (Figure 1L). The calculations are done with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood. As before, assume that the average late cancer was once the average early cancer, and thus individuals with Stage I cancer would be about 40,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 64 markers (48 markers for CRC) with average sensitivities of 75%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S). This would include, at 94.5% sensitivity, about 38,272 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 96 markers (48 markers for CRC) with average sensitivities of 66%, requiring a minimum of 5 markers positive. The two-step test would identify $94.5\% \times 90.0\% = 85.0\% = 34,445$ individuals

(out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $34,445 / (188,748 + 34,445) = 15.4\%$. In other words, 1 in 6.5 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. In expanding this example, the calculations are done with the anticipation that Stage I CRC has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II CRC has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, to be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals would be identified with Stage I cancer, 40,500 individuals would be identified with Stage II cancer, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 99.2% would be identified in the first step, of which $99.2\% \times 98.1\% = 97.3\% = 39,412$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9 % would be identified in the first step, of which $99.9\% \times 99.9\% = 99.8\% = 53,892$ individuals with late cancer would be identified. This brings the total identified at $34,445 + 39,412 + 53,892 = 127,749$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 94.6%. Overall, the positive predictive value of such a test would be $127,749 / (188,748 + 127,749) = 40.3\%$. In other words, 1 in 2.5 individuals who tested positive would actually have colorectal cancer, and this test would identify 73,857 / 81,000 or 91.2% of those individuals with early cancer, compared with the current rate of 40%.

[0366] How would these results vary for using this strategy (Figure 1L) for detection of early colorectal cancer using 75% average marker sensitivities, with the anticipation of Stage I CRC has an average of 200 methylated molecules per positive marker in the blood, Stage II CRC has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0367] Assuming individual marker false-positive rates of 3%, the first step using 64 markers (48 markers for CRC) with average sensitivities of 75%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 95.8%, the first step would identify 4,494,000 individuals (out of 107,000,000 total adults over 50 in the U.S.). This would include, at 99.2% sensitivity, about 40,176 individuals with Stage I colorectal cancer (out of 40,500 individuals with Stage I cancer). However, those 4,494,000 presumptive positive individuals would be evaluated in a second step of 96 markers (48 markers for CRC) with average

sensitivities of 66%, requiring a minimum of 5 markers positive. The two-step test would identify $99.2\% \times 98.0\% = 97.2\% = 39,372$ individuals (out of 40,500 individuals with Stage I cancer) with colorectal cancer. With a specificity of 95.8%, the second test would also generate $4,494,000 \times 4.2\% = 188,748$ false-positives. The positive predictive value of such a test would be $39,372 / (188,748 + 39,372) = 17.2\%$. In other words, 1 in 5.8 individuals who tested positive would actually have Stage I colorectal cancer. In reality, one would need to also include the success for identifying Stage 2 and higher cancers. To be consistent with the idea that as the test is used repeatedly, more of early and less of late CRC will be detected, then an estimate of 40,500 individuals would be identified with Stage I cancer, 40,500 individuals would be identified with Stage II cancer, and the remaining 54,000 individuals would have late-stage cancer = 135,000 total individuals with colorectal cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 99.9% would be identified in the first step, of which $99.9\% \times 99.6\% = 99.5\% = 40,297$ individuals with Stage II cancer would be verified in the second step. For Stage III and IV cancer, 99.9% would be identified in the first step, of which $99.9\% \times 99.9\% = 99.8\% = 53,892$ individuals with late cancer would be identified. This brings the total identified at $39,372 + 40,297 + 53,892 = 133,561$ individuals out of 135,000 with colorectal cancer, for an overall sensitivity of 98.9%. Overall, the positive predictive value of such a test would be $133,561 / (188,748 + 133,561) = 41.4\%$. In other words, 1 in 2.4 individuals who tested positive would actually have colorectal cancer, and this test would identify $79,699 / 81,000$ or 98.4% of those individuals with early cancer, compared with the current rate of 40%.

[0368] How would these results vary for using the third strategy (Figure 1L) for detection of early ovarian cancer, with the anticipation of an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood? Again, assume individuals with Stage I ovarian cancer would be about 5,500 individuals. Assuming individual marker false-positive rates of 3%, the first step using 64 markers (36 markers for ovarian) with average sensitivities of 75%, and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1%, the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.). This would include, at 80.3% sensitivity, about 4,416 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 96 markers (36 markers for ovarian cancer) with average sensitivities of 66%, requiring a minimum of 5 markers positive. The two-step test would identify $80.3\% \times 71.5\% = 57.4\% = 3,157$ individuals (out of 4,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1%, the second test would also generate $486,000 \times 0.9\% = 4,374$ false-

positives. The positive predictive value of such a test would be $3,157 / (4,374 + 3,157) = 41.9\%$. In other words, 1 in 2.4 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. In expanding this example, the calculations are done with the anticipation that Stage I ovarian cancer has an average of 150 methylated (or hydroxymethylated) molecules per positive marker in the blood, Stage II ovarian cancer has an average of 200 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 94.5% would be identified in the first step, of which $94.5\% \times 90.0\% = 85\% = 4,675$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.7% would be identified in the first step, of which $99.7\% \times 99.2\% = 98.9\% = 10,879$ individuals with late ovarian cancer would be identified. This brings the total identified at $3,157 + 4,675 + 10,879 = 18,711$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 85.1% . Overall, the positive predictive value of such a test would be $18,711 / (18,711 + 4,374) = 81.1\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify $7,832 / 11,000$ or 71.2% of those individuals with early cancer, compared with the current rate of 15% .

[0369] How would these results vary for using this strategy (Figure 1F) for detection of early ovarian cancer using 75% average marker sensitivities, with the anticipation that Stage I ovarian cancer has an average of 200 methylated molecules per positive marker in the blood, Stage II ovarian cancer has an average of 240 methylated molecules per positive marker, and the higher stages (III & IV) have at least an average of 300 methylated molecules per positive marker?

[0370] Assuming individual marker false-positive rates of 3% , the first step using 64 markers (36 markers for ovarian) with average sensitivities of 75% , and requiring a minimum of 5 markers positive, then, with an overall specificity of 99.1% , the first step would identify 486,000 individuals (out of 54,000,000 total women ages 50-79 in the U.S.). This would include, at 94.5% sensitivity, about 5,197 individuals with Stage I ovarian cancer (out of 5,500 individuals with Stage I ovarian cancer). However, those 486,000 presumptive positive individuals would be evaluated in a second step of 96 markers (36 markers for ovarian cancer) with average sensitivities of 66% , requiring a minimum of 5 markers positive. The two-step test would identify $94.5\% \times 90.0\% = 85\% = 4,675$ individuals (out of 5,500 individuals with Stage I ovarian cancer) with ovarian cancer. With a specificity of 99.1% , the second test would also

generate $486,000 \times 0.9\% = 4,374$ false-positives. The positive predictive value of such a test would be $4,675 / (4,374 + 4,675) = 51.6\%$. In other words, 1 in 2 individuals who tested positive would actually have Stage I ovarian cancer. In reality, one would need to also include the success for identifying Stage 2 and higher ovarian cancers. Also, assume all four stages are at 5,500, then $5,500 \times 4 = 22,000$ total individuals with ovarian cancer would be identified per year in the U.S. The above calculation already provided the false-positive rate for the early cancer. For Stage II cancer, 98.3% would be identified in the first step, of which $98.3\% \times 96.2\% = 94.5\% = 5,201$ individuals with Stage II ovarian cancer would be verified in the second step. For Stage III and IV ovarian cancer, 99.7% would be identified in the first step, of which $99.7\% \times 99.2\% = 98.9\% = 10,879$ individuals with late ovarian cancer would be identified. This brings the total identified at $4,675 + 5,201 + 10,879 = 20,755$ individuals out of 22,000 with ovarian cancer, for an overall sensitivity of 94.3%. Overall, the positive predictive value of such a test would be $20,755 / (20,755 + 4,374) = 82.6\%$. In other words, 4 in 5 women who tested positive would actually have ovarian cancer, and this test would identify $9,876 / 11,000$ or 89.8% of those individuals with early cancer, compared with the current rate of 15%.

[0371] The aforementioned 5 groups of 48 markers, with average sensitivity of 75%, were designed to also be used to monitor treatment (see Figure 1M). Currently, with a newly diagnosed cancer, cancer tissue (or liquid biopsy) is subjected to targeted sequencing to identify mutations or gene rearrangements that may be used to guide therapy. For a given cancer (*i.e.* stomach cancer in Group 1), the cancer tissue or liquid biopsy may be tested with the 48-marker group (1) panel. If the cancer had been identified in the first place using the 2-step screens identified in Figures 1F or 1J, then they will have already undergone the 48-marker group specific test in step 2 of that assay. Of the 48 markers tested, on average 12-24 would be positive. These may then be bundled together in a patient-specific test to monitor treatment efficacy. The plasma of such a patient would be tested post surgery, and during the treatment regimen. The plasma is monitored for loss of the 12-24 marker signal, but if ≥ 3 positive markers remain positive, then this may guide the physician to change therapy. Depending on the cancer type, and how many molecules enter the plasma, 3 markers would be predicted to identify treatment efficacy or failure with an accuracy of 82.6% to 99.4%.

[0372] The aforementioned 5 groups of 48 markers were designed to also be used to monitor for recurrence (see Figure 1N). If the cancer had been identified in the first place using the 2-step screens identified in Figures 1F and 1J, and/or was monitored as described in Figure 1M, then they will have already undergone the 48-marker group specific test, for which on average 12-24 would be positive. These may then be bundled together in a patient-specific test to monitor for recurrence. The plasma of such a patient who recovered from the original cancer

would be monitored for gain of markers from the 12-24 marker panel. Results are scored as follows: 0-2 positive markers are considered cancer-free; ≥ 3 positive markers are directed to go to the second step. The plasma would be subjected to targeted sequencing to identify mutations or gene rearrangements that may be used to guide therapy of the recurrent tumor. Depending on the cancer type, and how many molecules enter the plasma, 3 markers would be predicted to identify early recurrence with an accuracy of 82.6% to 99.4%.

[0373] The biology of each cancer is different, and thus the observed sensitivity and specificity for detecting early cancer, monitoring treatment, and detecting early recurrence may be higher or lower from the idealized calculations described herein.

TABLE 45

Site	Primer Name	Sequence	Length	SEQ ID NO:
Universal Primer VIM	iCDx-2000- Uni_Mul_Pri	GGTGTCTGGAGTTCAACATAAAC/3SpC3/	24	1
Forward PCR Primer	AcDx-5001- VIM-S1-FP	CGAGTCGGTCGAGTTTTAGTCrGGAGC/3SpC3/	27	2
Reverse PCR Primer with long tail	AcDx-5002B- VIM-S1-RP	GGTGTCTGGAGTTCAACATAATCCCGAAAAACGTAACAACTACrGACTG/3SpC3/	54	3
Upstream LDR	AcDx-5003- VIM-S1-Up	TCTCATACCAGACGCGGTAACTCGAGTTTTAGTCGGAGTTACGTGATCACrGTTTCG/3SpC3/	56	4
Downstream LDR	AcDx-5004- VIM-S1-Dn	/5Phos/GTTTATTTCGTATTATAGTTTGGGTAGCGGTTGCGGTCGTGCTGCTGCTTA	55	5
Real-Time Probe	AcDx-5005- VIM-S1-RT- Pb	/56-FAM/AATGATCAC/ZEN/GTTTATTTCGTATTATAGTTTGGGTAGCG/3IABkFQ/	41	6
Tag Forward Primer	AcDx-5006- VIM-S1-RT- FP	TCTCATACCAGACGCGGTAAC	21	7
Tag Reverse Primer	AcDx-5007- VIM-S1-RT- RP	TAAGCACACGCGACACGAAC	19	8
Dwnstrm PCR Primer	AcDx-5008- VIM-S1-PCR- V	TAAGCACACGCGACACGAACCGAAACGTAAAAACTACGACTAATACTAAAAATGrCAACA/3SpC3/	58	9

Site	Primer Name	Sequence	Length	SEQ ID NO:
CLIP4-1				
Forward PCR Primer	AcDx-5021-CLIP4-S1-FP	GGTTGAGGGTTGTGAAGGCAGGTGA/3SpC3/	25	10
Reverse PCR Primer with long tail	AcDx-5022B-CLIP4-S1-RP	GGTGTCTGGAGTTCAACATAATCGTCTACGAAATATCGCAATATTACCTJCCCCCT/3SpC3/	55	11
Upstream LDR	AcDx-5023-CLIP4-S1-Up	TCCAAACAAGCTGATCCGTACAGGTTGTGAAGGCGGTGGGCACrGTATA/3SpC3/	49	12
Downstream LDR	AcDx-5024-CLIP4-S1-Dn	/5Phos/GTACGGCGTGTCTGGAGTCGTTTGGTGTGTCTGGAGCGGTTACTA	43	13
Real-Time Probe	AcDx-5025-CLIP4-S1-RT-Pb	/56-FAM/AATGGGCAC/ZEN/GTACGGCGTGT/3/ABkFQ/	23	14
Tag Forward Primer	AcDx-5026-CLIP4-S1-RT-FP	TCCAAACAAGCTGATCCGTACA	22	15
Tag Reverse Primer	AcDx-5027-CLIP4-S1-RT-RP	TAGTAACCGCTCCGACACA	19	16
Dwnstrm PCR Primer	AcDx-5028-CLIP4-S1-PCR-V	TAGTAACCGCTCCGACACACGCCCGGAAACCAATGrACCCT/3SpC3/	42	17
GSG1L				18
Forward PCR Primer	AcDx-5051-GSG1L-S1-FP	AGTCGGAGTCGAGTTGGTGrGTCGC/3SpC3/	24	19

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer with long tail	AcDx-5052B-GSGIL-S1-RP	GGTGTCTGGAGTTCAACATAATCGTCTACGAAATATCGCAATATTACCUCCCT/3SpC3/	54	20
Upstream LDR	AcDx-5053-GSGIL-S1-Up	TCTGCCAGAACACCCGACACGGAGTCGAGTTGGTCGCTCrGCGTA/3SpC3/	46	21
Downstream LDR	AcDx-5054-GSGIL-S1-Dn	/5Phos/GCGCGTATTTATTAAGTTCGTTGAGTTTTTTTTTCGTACGGTGTGTGGCGTACGGTGA	58	22
Real-Time Probe	AcDx-5055-GSGIL-S1-RT-Pb	/56-FAM/AAGTCGCTC/ZEN/GCGCGTATTTATTAAGTTCGT/3IABkFQ/	30	23
Tag Forward Primer	AcDx-5056-GSGIL-S1-RT-FP	TCTGCCAGAACACCCGACAC	19	24
Tag Reverse Primer	AcDx-5057-GSGIL-S1-RT-RP	TCACCGTACGCCAACACAC	19	25
Dwnstrm PCR Primer	AcDx-5058-GSGIL-S1-PCR-V	TCACCGTACGCCAACACACACACACCATCTAATACTCGTATGrAAAAAG/3SpC3/	49	26
PP1R16B				
Forward PCR Primer	AcDx-5061-PP1R16B-S1-FP	GGGTTTTTATTCGAGAGCGTCGGGAC/3SpC3/	26	27
Reverse PCR Primer with long tail	AcDx-5062B-PP1R16B-S1-RP	GGTGTCTGGAGTTCAACATAATCCCAAAACGAAACCTAACTCCUAAAAG/3SpC3/	50	28
Upstream LDR	AcDx-5063-PP1R16B-S1-Up	TTCTGTGGGCACACAAGCAACGAGAGCGTCGGGATTTTGTCTCrGCGCC/3SpC3/	47	29
Downstream LDR	AcDx-5064-PP1R16B-S1-Dn	/5Phos/GCGTTGTTTTTAAAGTCGGATGGAGTTGAGCTTGCTTGCTTGCTGATCTACCTGA	53	30

Site	Primer Name	Sequence	Length	SEQ ID NO:
Real-Time Probe	AcDx-5065-PPIR16B-S1-RT-Pb	/56-FAM/AATTGTCTC/ZEN/GCGTTGTTTTTAAAGTCGGATG/3IABkFQ/	31	31
Tag Forward Primer	AcDx-5066-PPIR16B-S1-RT-FP	TTTCGTGGGCACACAAGCAA	19	32
Tag Reverse Primer	AcDx-5067-PPIR16B-S1-RT-RP	TCAGGTAGATCAAGCCAAAGCAA	22	33
Dwnstrm PCR Primer	AcDx-5068-PPIR16B-S1-PCR-V	TCAGGTAGATCAAGCCAAAGCAAACCTAAACCTCTAAACATAAAACGTGrCTCAG/3SpC3/	57	34
KCNA3				
Forward PCR Primer	AcDx-5071-KCNA3-S1-FP	GCGCGCGTTTCGTTTTTCrGGGAA/3SpC3/	22	35
Reverse PCR Primer with long tail	AcDx-5072B-KCNA3-S1-RP	GGTGTCTGGAGTTCAACATAATCGCCGAAATACAACATAAAAACTCrUTTCA/3SpC3/	52	36
Upstream LDR	AcDx-5073-KCNA3-S1-Up	TTTCAGGCCCCTAACCAACCGCGTTTCGTTTTTCGGAGGTAATCrGTCAA/3SpC3/	48	37
Downstream LDR	AcDx-5074-KCNA3-S1-Dn	/5Phos/GTCGGGTTTGTAATTTTTTGTAGTTTTTAAGGTTTTTCGGTGTGGGATTAAAGGGCGGATGGA	60	38
Real-Time Probe	AcDx-5075-KCNA3-S1-RT-Pb	/56-FAM/AAGGTAATC/ZEN/GTCGGGTTTGTATTTTTTGTAGTTTTTAAGG/3IABkFQ/	40	39
Tag Forward Primer	AcDx-5076-KCNA3-S1-RT-FP	TTTCAGGCCCCTAACCAACCAC	20	40

Site	Primer Name	Sequence	Length	SEQ ID NO:
Tag Reverse Primer	AcDx-5077-KCNA3-S1-RT-RP	TCCATCGCCCTTAATCCCAC	20	41
Dwnstrm PCR Primer	AcDx-5078-KCNA3-S1-PCR-V	TCCATCGCCCTTAATCCCACCAACATAAAAACTTTTCGCTAACACTGrAAAAAG/3SpC3/	53	42
GDF6				
Forward PCR Primer	AcDx-5081-GDF6-S1-FP	GGTTGCGTTTTTTTAGGAGGGrGGTGA/3SpC3/	26	43
Reverse PCR Primer with long tail	AcDx-5082B-GDF6-S1-RP	GGTGTCTGGAGTTCAACATAATACCCCGACCCGCTATCCrAACCA/3SpC3/	31	44
Upstream LDR	AcDx-5083-GDF6-S1-Up	TCACTATCGGCGTAGTCACCAAGGCGGTGGCAGCtGGCAC/3SpC3/	40	45
Downstream LDR	AcDx-5084-GDF6-S1-Dn	/5Phos/GGCGTAGGACGCGCGGG TGGTGACTTTACCCCGGAGGA	37	46
Real-Time Probe	AcDx-5085-GDF6-S1-RT-Pb	/56-FAM/AATGGCAGC/ZEN/GGCGTAGGACG/3IABkFQ/	20	47
Tag Forward Primer	AcDx-5086-GDF6-S1-RT-FP	TCACTATCGGCGTAGTCACCA	21	48
Tag Reverse Primer	AcDx-5087-GDF6-S1-RT-RP	TCCTCCGGGTAAAGTCACCA	20	49

Site	Primer Name	Sequence	Length	SEQ ID NO:
Dwnstrm PCR Primer	AcDx-5088-GDF6-S1-PCR-V	TCCTCCGGGTAAAGTCACCAAAACCGCTCCGTACCCCTGrCGCGC/3SpC3/	42	50
ADHFE1				
Forward PCR Primer	AcDx-5101-ADHFE1-S1-FP	GGTGGAGCGTCGTTGGGAC/3SpC3/	20	51
Reverse PCR Primer with long tail	AcDx-5102C-ADHFE1-S1-RP	GGTGTCTGGAGTTCAACATAATGCCTACCCACCCGCTUTCGT/3SpC3/	42	52
Upstream LDR	AcDx-5103-ADHFE1-S1-Up	TTGATTGGGATCGTTCGCACGGGTAGTTGGCGTTTGGTTTTATCTCgtGAA/3SpC3/	53	53
Downstream LDR	AcDx-5104-ADHFE1-S1-Dn	/5Phos/GTGGGAAAAATGGTTTTGAGTTCGATTGGTTTGAGGTGGCTCAATAACGGGCAGA	54	54
Real-Time Probe	AcDx-5105-ADHFE1-S1-RT-Pb	/56-FAM/AATTATCTC/ZEN/GTGGGAAAAATGGTTTTGAGTTCGA/3IABkFQ/	33	55
Tag Forward Primer	AcDx-5106-ADHFE1-S1-RT-FP	TTGATTGGGATCGTTCGCAC	20	56
Tag Reverse Primer	AcDx-5107-ADHFE1-S1-RT-RP	TCTGCCCGTTATTGAGCCAC	20	57

Site	Primer Name	Sequence	Length	SEQ ID NO:
Dwnstrm PCR Primer	AcDx-5108-ADHFE1-S1-PCR-V	TCTGCCCGTTATTGAGCCACCCACCCGCTTCGTG rAAATT/3SpC3/	40	58
THBD				
Forward PCR Primer	AcDx-5331-THBD-S1-FP	TATAGGACGTCGATGGCGATArGTTTC/3SpC3/	26	59
Reverse PCR Primer with long tail	AcDx-5332B-THBD-S1-RP	GGTGTCTGGAGTTCAACATAAATCGATCCGCATATCAAAAACTACCUCCGG/3SpC3/	51	60
Upstream LDR	AcDx-5333-THBD-S1-Up	TTCAGAGCACCTGCGTACCACGTCGATGGCGATAGTTTTTTTGCTC-GTTCC/3SpC3/	52	61
Downstream LDR	AcDx-5334-THBD-S1-Dn	/5Phos/GTTTTAGTTTAGATAATTTTTGTCTGTCGCGTAGTTTTTGGGTTCTTCGGCTGGCTCAA	60	62
Real-Time Probe	AcDx-5335-THBD-S1-RT-Pb	/56-FAM/AATTGTGCTC/ZEN/GTTTTAGTTTAGATAATTTTTGTCTGTTGCG/3IABkFQ/	39	63
Tag Forward Primer	AcDx-5336-THBD-S1-RT-FP	TTCAGAGCACCTGCGTACC	19	64
Tag Reverse Primer	AcDx-5337-THBD-S1-RT-RP	TTGAGCCAGCCGGAAGAACC	19	65
Dwnstrm PCR Primer	AcDx-5338-THBD-S1-PCR-V	TTGAGCCAGCCGGAAGAACCATATCAAAAACTACCTCGCAAAAACTATGrCGCAG/3SpC3/	54	66

Site	Primer Name	Sequence	Length	SEQ ID NO:
SEPT9				
Forward PCR Primer	AcDx-5351-SEPT9-S4-FP	GTGGGTGTTGGGTTGGTtUGTTA/3SpC3/	22	67
Reverse PCR Primer with long tail	AcDx-5352B-SEPT9-S4-RP	GGTGTCTGGAGTTCAACATAATCAAAACCCACCCGCAAAATUCCTT/3SpC3/	45	68
Upstream LDR	AcDx-5353-SEPT9-S4-Up	TACACGTGGATATCTCCGACCGGGTGTGGGTGGTGTGTCGCrGGTTA/3SpC3/	47	69
Downstream LDR	AcDx-5354-SEPT9-S4-Dn	/5Phos/GGTCGGGACGTGTGGAGAGGGGTGCTAGTCACACAGTTCCA	43	70
Real-Time Probe	AcDx-5355-SEPT9-S4-RT-Pb	/56-FAM/TATTGTCCG/ZEN/GGTCGGGACG/3IABkFQ/	20	71
Tag Forward Primer	AcDx-5356-SEPT9-S4-RT-FP	TACACGTGGATATCTCCGACC	21	72
Tag Reverse Primer	AcDx-5357-SEPT9-S4-RT-RP	TGGAACCTGTGTGACTAGCACC	21	73
Dwnstrm PCR Primer	AcDx-5358-SEPT9-S4-PCR-V	TGGAACCTGTGTGACTAGCACCCTCTCTCAACATGrUCCGT/3SpC3/	48	74
SEMA3B				

Site	Primer Name	Sequence	Length	SEQ ID NO:
Forward PCR Primer	AcDx-5401-SEMA3B-S1-FP	CGTCGCGTGTAGGGTTCTGGAAA/3SpC3/	23	75
Reverse PCR Primer with long tail	AcDx-5402B-SEMA3B-S1-RP	GGTGTCTGGAGTTCAACATAATCGATACGCTCCTCTACCAACrACCTG/3SpC3/	48	76
Upstream LDR	AcDx-5403-SEMA3B-S1-Up	TCCTGCTCTGAAAAACCTACACCCGTTGTTAGGGTTCGGAAAGTTTGTCTCTGGTCT/3SpC3/	55	77
Downstream LDR	AcDx-5404-SEMA3B-S1-Dn	/5Phos/GGTTCGATATTTTCGTTTACGTTGTTTTTTTGTTCGTAGGGGTTACATAGGCGGCTTAGACA	62	78
Real-Time Probe	AcDx-5405-SEMA3B-S1-RT-Pb	/56-FAM/AATGTTCTC/ZEN/GGTTTCGATATTTTCGTTTACGTTGT/3IABRFQ/	35	79
Tag Forward Primer	AcDx-5406-SEMA3B-S1-RT-FP	TCCTGCTCTGAAAAACCTACACC	22	80
Tag Reverse Primer	AcDx-5407-SEMA3B-S1-RT-RP	TGTCTAAGCCGCCCTATGTAAACC	22	81
Dwnstrm PCR Primer	AcDx-5408-SEMA3B-S1-PCR-V	TGTCTAAGCCGCCCTATGTAAACCCGCTCCTCTACCAACACCTATGrAACAG/3SpC3/	49	82
GATA5				

Site	Primer Name	Sequence	Length	SEQ ID NO:
Forward PCR Primer	AcDx-5421-GATA5-S1-FP	CGCGGTCGTAGGACGTArGGGTC/3SpC3/	22	83
Reverse PCR Primer with long tail	AcDx-5422B-GATA5-S1-RP	GGTGTCTGTGGAGTTCAACATAATTCCAAACCCGAACTACAACCGCGCA/3SpC3/	47	84
Upstream LDR	AcDx-5423-GATA5-S1-Up	TTGTCTCTGCGACCCATCAAGTAGGACGTAGGGTTTGGAGGGCrGGGAC/3SpC3/	48	85
Downstream LDR	AcDx-5424-GATA5-S1-Dn	/5Phos/GGGATTTCGTGCGGTTGGGAGGGTTGGTACACGTTTCGGCACA	42	86
Real-Time Probe	AcDx-5425-GATA5-S1-RT-Pb	/56-FAM/AAGGAGGGC/ZEN/GGGATTTCGTGCGC/3IABkFQ/	22	87
Tag Forward Primer	AcDx-5426-GATA5-S1-RT-FP	TTGTCTCTGCGACCCATCAA	20	88
Tag Reverse Primer	AcDx-5427-GATA5-S1-RT-RP	TGTGCCGAACGTGTATCCAA	19	89
Dwnstrm PCR Primer	AcDx-5428-GATA5-S1-PCR-V	TGTGCCGAACGTGTATCCAAACCCGACCCCTCCCAATGrCGACA/3SpC3/	41	90
ZNF542				
Forward PCR Primer	AcDx-5431-ZNF542-S1-FP	CGTTTTTGTATTTCGGTTATTGGGAAGCGGA/3SpC3/	30	91

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer with long tail	AcDx-5432B-ZNF542-S1-RP	GGTGTCTGGAGTTCAACATAATACGCCCGAATAATTTCTAAAAATAAACGAAAAG/3SpC3/	55	92
Upstream LDR	AcDx-5433-ZNF542-S1-Up	TTTCGCTCGACGCATACACGGTTATTGGGAGCGGGATC GTGAA/3SpC3/	44	93
Downstream LDR	AcDx-5434-ZNF542-S1-Dn	/5Phos/GTGGGAGTTGTATATGCGTATTGCGAGTTTTCTGGCGCGGCTACTGTAAAA	51	94
Real-Time Probe	AcDx-5435-ZNF542-S1-RT-Pb	/56-FAM/TTTCGGGATC/ZEN/GTGGGAGTTGTATATGCG/3IABkFQ/	27	95
Tag Forward Primer	AcDx-5436-ZNF542-S1-RT-FP	TTTCGCTCGACGCATACCA	19	96
Tag Reverse Primer	AcDx-5437-ZNF542-S1-RT-RP	TTTTACAGTAGCCGCGCCA	19	97
Dwnstrm PCR Primer	AcDx-5438-ZNF542-S1-PCR-V	TTTTACAGTAGCCGCGCCACCCCGAATAATTTCTAAAAATAAACGAAAACTTGTCAATG/3SpC3/	57	98
RCN3				
Forward PCR Primer	AcDx-5441-RCN3-S1-FP	CGTGAGGCGTTGTGATTAGAATArGTTGA/3SpC3/	28	99

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer with long tail	AcDx-5442B-RCN3-S1-RP	GGTGTCTGGAGTTCAACATAATTAAACGCGACCCGAAAAAACTACrAACTT/3SpC3/	50	100
Upstream LDR	AcDx-5443-RCN3-S1-Up	TTGCACGTTGTCTCTGCACCCGTTGTGATTAGAATAGTTGGAGGTGAACrGGTGA/3SpC3/	53	101
Downstream LDR	AcDx-5444-RCN3-S1-Dn	/5Phos/GGTAGAGTGTCCGCGACGATTGTTAGGAGTGGTAGTTTCCCATGACGGCA	49	102
Real-Time Probe	AcDx-5445-RCN3-S1-RT-Pb	/56-FAM/TTGGTGAAC/ZEN/GGTAGAGTGTCCGGAC/3IABkFQ/	25	103
Tag Forward Primer	AcDx-5446-RCN3-S1-RT-FP	TTGCACGTTGTCTCTGCACC	19	104
Tag Reverse Primer	AcDx-5447-RCN3-S1-RT-RP	TGCCGTCATGGGAAACTACC	20	105
Dwnstrm PCR Primer	AcDx-5448-RCN3-S1-PCR-V	TGCCGTCATGGGAAACTACCACCCGAAAAAACTACAACCTCCTAACAAATTGrUUGCA/3SpC3/	55	106
MYO15B				
Forward PCR Primer	AcDx-5451-MYO15B-S1-FP	TTTAGGAGTTTTTAATGGAGATACGTCTGGGTGA/3SpC3/	31	107

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer with long tail	AcDx-5452B-MYO15B-S1-RP	GGTGTCTGGAGTTCAACATAATCCGAACTATACCGCGCTAACrUACCA/3SpC3/	48	108
Upstream LDR	AcDx-5453-MYO15B-S1-Up	TTAGCCGCCAAACGTACCATGGGAACGGAGGTAGTTTTTGTCTCTGGACG/3SpC3/	48	109
Downstream LDR	AcDx-5454-MYO15B-S1-Dn	/5Phos/GGATAGCGAAATTTCGCGAGGTTTAGGAGAGTGGGCAGGAACACGATAGTA	50	110
Real-Time Probe	AcDx-5455-MYO15B-S1-RT-Pb	/56-FAM/CCCTTTGCTC/ZEN/GGATAGCGAAATTTCGCGA/3IABkFQ/	27	111
Tag Forward Primer	AcDx-5456-MYO15B-S1-RT-FP	TTAGCCGCCAAACGTACCA	19	112
Tag Reverse Primer	AcDx-5457-MYO15B-S1-RT-RP	TACTATCGTGTTCCTGCCCCA	20	113
Dwnstrm PCR Primer	AcDx-5458-MYO15B-S1-PCR-V	TACTATCGTGTTCCTGCCCCACGAACTATACCGCGCTAACTACTGrCTCTT/3SpC3/	49	114
ANKRD13B				
Forward PCR Primer	AcDx-5461-ANKRD13B-S1-FP	CGAGTAGTTGCGGTTGGCtGATGA/3SpC3/	23	115

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer	AcDx-5462B-ANKRD13B-S1-RP	GGTGTCTGGAGTTCAACATAATCCAACCTCCTCCTCCTAArCGCGT/3SpC3/	43	116
Upstream LDR	AcDx-5463-ANKRD13B-S1-Up	TTTCGTACCTCGGCACACACAGCGGTTGGCGATGGAATTATCtGGCAC/3SpC3/	45	117
Downstream LDR	AcDx-5464-ANKRD13B-S1-Dn	/5Phos/GGCGTAGGAGTAGGAGGAGGCGTGGCTCCGTTACTCTGTCTGA	44	118
Real-Time Probe	AcDx-5465-ANKRD13B-S1-RT-Pb	/56-FAM/CCAAATTATC/ZEN/GGCGTAGGAGTAGGAGGAGAGG/3IABkFQ/	31	119
Tag Forward Primer	AcDx-5466-ANKRD13B-S1-RT-FP	TTTCGTACCTCGGCACACCA	19	120
Tag Reverse Primer	AcDx-5467-ANKRD13B-S1-RT-RP	TCGACAGAGTAACGGAGCCA	20	121
Dwnstrm PCR Primer	AcDx-5468-ANKRD13B-S1-PCR-V	TCGACAGAGTAACGGAGCCACCAACTCCTCCTCCTAArCGCGT/3SpC3/	47	122
FAM155A				
Forward PCR Primer	AcDx-5471-FAM155A-S1-FP	AGGTTGGTGTGGTGTGGCGCA/3SpC3/	23	123

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer	AcDx-5472B-FAM115A-S1-RP	GGTGTCTGGAGTTCAACATAATCGCTAACAATACCTAAATAACCGAAAC+CGCGT/3SpC3/	50	124
Upstream LDR	AcDx-5473-FAM115A-S1-Up	TTTGCCCTCTTGTAGGTGCCAGAGGTTGGGTGTAGGGAGCGATAAA/3SpC3/	44	125
Downstream LDR	AcDx-5474-FAM115A-S1-Dn	/5Phos/GATGGTGGAGGTGATAGGGTGGTGGTGGGCAACGCGGATATTCA	45	126
Real-Time Probe	AcDx-5475-FAM115A-S1-RT-Pb	/56-FAM/AAAGGGAGC/ZEN/GATGGTGGAGGTGA/3IABkFQ/	23	127
Tag Forward Primer	AcDx-5476-FAM115A-S1-RT-FP	TTTGCCCTCTTGTAGGTGCCA	20	128
Tag Reverse Primer	AcDx-5477-FAM115A-S1-RT-RP	TGAATATCCGCGTTGCCCA	19	129
Dwnstrm PCR Primer	AcDx-5478-FAM115A-S1-PCR-V	TGAATATCCGCGTTGCCCATAAACAATACCTAAATAACCGAAACCGTGTCCAAT/3SpC3/	52	130
RGS10				
Forward PCR Primer	AcDx-5491-RGS10-S1-FP	CGTTCTAGCGGAGGC+GGAGG/3SpC3/	21	131

Site	Primer Name	Sequence	Length	SEQ ID NO:
Reverse PCR Primer	AcDx-5492B-RGS10-S1-RP	GGTGTCTGGAGTTCAACATAATAAAAAACGCCCAAAATCTCCrAAACG/3SpC3/	42	132
Upstream LDR	AcDx-5493-RGS10-S1-Up	TCCCTCTGTCATCTCCCTTACCCGGAGGGAGAAAGTTCTGTCGTGC/3SpC3/	45	133
Downstream LDR	AcDx-5494-RGS10-S1-Dn	/5Phos/GTCGTTTCGTTTTCGGAAATTTGGAGTTTTATGTTATTTTGGTCTTGGTGATGGAGCGA	58	134
Real-Time Probe	AcDx-5495-RGS10-S1-RT-Pb	/56-FAM/CCCTTCGTGC/ZEN/GTCGTTTCGTTTTCGGGA/3IABkFQ/	26	135
Tag Forward Primer	AcDx-5496-RGS10-S1-RT-FP	TCCCTCTGTCATCTCCCTTACC	21	136
Tag Reverse Primer	AcDx-5497-RGS10-S1-RT-RP	TCGCTCCATCACCAAGACC	19	137
Dwnstrm PCR Primer	AcDx-5498-RGS10-S1-PCR-V	TCGCTCCATCACCAAGACCCCAAACTTTAAAAATAACATAAAACTCCAAATTTCTGrAAAAAT/3SpC3/	60	138
HCG4				
Forward PCR Primer	AcDx-5501-HCG4-S1-FP	GTCGGAATATTGGGAAGAGGArGATAA/3SpC3/	26	139
Reverse PCR Primer	AcDx-5502B-HCG4-S1-RP	GGTGTCTGGAGTTCAACATAATCCTCACTCTAATTATAATAAACCCTCrAAAAAC/3SpC3/	49	140
Upstream LDR	AcDx-5503-HCG4-S1-Up	TTCTAGATACCAACGGACGCACCGGAATATTGGGAAGAGGAGATAGGGTTCrGTTGG/3SpC3/	55	141

Site	Primer Name	Sequence	Length	SEQ ID NO:
Downstream LDR	AcDx-5504-HCG4-S1-Dn	/5Phos/GTTAAGGTTAAAGTATAGTTTTATCGAGTGAATTTGCCGGAATTTGTGTGGTGTGCAAAAGCTGA	64	142
Real-Time Probe	AcDx-5505-HCG4-S1-RT-Pb	/56-FAM/CCAGGGTTC/ZEN/GTTAAGGTTAAAGTATAGTTTTATCGAGTGA/3IABkFQ/	40	143
Tag Forward Primer	AcDx-5506-HCG4-S1-RT-FP	TTCTAGATACCAACGGACGCAC	21	144
Tag Reverse Primer	AcDx-5507-HCG4-S1-RT-RP	TCAGCTTTGCACACCAACAC	20	145
Dwnstrm PCR Primer	AcDx-5508-HCG4-S1-PCR-V	TCAGCTTTGCACACCAACACCTCACTCTAATTATAATAACCGCTCAAAAATCTGrCAAAAC/3SpC3/	58	146

EXAMPLES

Examples: Multiplex PCR-LDR-qPCR Detection of Cancer-Related Methylation Markers

General Methods for Examples 1-2

[0374] HT-29 colon adenocarcinoma cells were seeded in 60 cm² culture dishes in McCoy's 5A medium containing 4.5 g/l glucose, supplemented with 10% fetal calf serum, and kept in a humidified atmosphere containing 5% CO₂. Once cells reached 80-90% confluence, they were washed in Phosphate Buffered Saline (x3), and collected by centrifugation (500xg). Genomic DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen; Valencia, Calif.), and its concentration measured using Quant-iT Pico green Assay (Life Technologies/ThermoFisher; Waltham, Mass.).

[0375] High molecular weight (>50 kb) genomic DNA (0.2 mg/ml) isolated from human blood (buffy coat) (Roche human genomic DNA) was purchased from Roche (Indianapolis, Ind.). Its concentration was similarly determined using Quant-iT PicoGreen dsDNA Assay Kit.

[0376] Cell free DNA was isolated from 5 ml plasma sample (with K₂EDTA additive as anti-coagulant) using the QIA amp Circulating Nucleic Acid Kit according to manufacturer's instructions, and quantified using Quant-iT Pico Green Assay (Life Technologies/ThermoFisher; Waltham, Mass.).

[0377] 0.5-1.0 µg HT29 cell line genomic DNA was digested with 10 units of the restriction enzyme Bsh1236I in 20 µl of reaction solution containing 1xCutSmart buffer (50 mM Potassium Acetate, 20 mM Tris-Acetate, 10 mM Magnesium Acetate, 100 µg/ml BSA, pH 7.9 at 25°C). The digestion reaction was carried out at 37°C for 1 hour, followed by enzyme inactivation at 80°C for 20 min. Alternatively, genomic DNAs can be fragmented through non random sonication method, using Covaris ultra sonicator (Woburn, Massachusetts). After shearing, the quality of the resulting DNA fragments (length ranged from 50 to 1 kb base pairs) was assessed with Agilent Bioanalyzer system. This is followed by an enrichment step wherein the DNA fragments containing methylated CpGs are then captured by methylation-specific antibodies, using the EpiMark® Methylated DNA Enrichment Kit according to manufacturer's instructions (New England Biolabs; Ipswich, MA).

[0378] PCR primers and LDR probes. All primers to be used in the various categories are listed in the Table 46 above. Primers are purchased from Integrated DNA Technologies Inc. (IDT) (Coralville, Iowa). Alternative primers for use in one or two-step assay to detect colorectal cancer are listed in Table 39 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers designed for use

in Step 1 of the 96-marker assay, with average sensitivities of 50%, detect solid tumors are listed in Table 40 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers designed for use in Step 2 of the Group 1- 64-marker assay, with average sensitivities of 50%, to detect and identify colorectal, stomach, and esophageal cancers are listed in Table 47 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 2- 48-64-marker assay, with average sensitivities of 50%, to detect and identify breast, endometrial, ovarian, cervical, and uterine cancers are listed in Table 48 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 3- 48-64-marker assay, with average sensitivities of 50%, to detect and identify lung adenocarcinomas, lung squamous cell carcinoma, and head & neck cancers are listed in Table 49 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 4- 36-48-marker assay, with average sensitivities of 50%, to detect and identify prostate and bladder cancers are listed in Table 50 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 5- 48-64-marker assay, with average sensitivities of 50%, to detect and identify liver, pancreatic, and gall-bladder cancers are listed in Table 51 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 1 of the 48-64-marker assay, with average sensitivities of 75%, to detect solid tumors are listed in Table 52 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 1- 36-48-marker assay, with average sensitivities of 75%, to detect and identify colorectal, stomach, and esophageal cancers are listed in Table 53 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 2- 32-48-marker assay, with average sensitivities of 75%, to detect and identify breast, endometrial, ovarian, cervical, and uterine cancers are listed in Table 54 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 3- 36-48-marker assay, with average sensitivities of 75%, to detect and identify lung adenocarcinomas, lung squamous cell carcinoma, and head & neck cancers are listed in Table 55 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 4- 36-48-marker assay, with average sensitivities of 75%, to detect and identify prostate, bladder, and kidney cancers from blood samples are listed in Table 56A of U.S. Provisional Patent Application Serial No. 63/019,142, which is

hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 4- 36-48-marker assay, with average sensitivities of 75%, to detect and identify prostate and bladder cancers from urine samples are listed in Table 56B of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Step 2 of the Group 5- 36-48-marker assay, with average sensitivities of 75%, to detect and identify liver, pancreatic, and gall-bladder cancers are listed in Table 57 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety. Primers for use in Group 7- 36-48-marker assay, with average sensitivities of 75%, to detect recurrence in melanoma are listed in Table 58 of U.S. Provisional Patent Application Serial No. 63/019,142, which is hereby incorporated by reference in its entirety.

Example 1: Using TET2_APOBEC converted DNA templates in Universal Primer based multiplex amplification of 20 plex PCR-LDR-qPCR for Colon Cancer-Related Methylation Marker detection.

[0379] DNA preparation: HT-29 colon adenocarcinoma cells were seeded in 60 cm² culture dishes in McCoy's 5A medium containing 4.5 g/l glucose, supplemented with 10% fetal calf serum, and kept in a humidified atmosphere containing 5% CO₂. Once cells reached 80-90% confluence, they were washed in Phosphate Buffered Saline (x3), and cells collected by centrifugation (500xg). Colon Cancer cell line genomic DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA.), and its concentration measured using Quant-iT Pico green dsDNA Assay kit (Thermo-Fisher, Waltham, MA.). High molecular weight (>50 kb) genomic DNA (0.2 mg/ml) isolated from human blood (buffy coat) (Roche human genomic DNA) was purchased from Roche (Indianapolis, IN.). Its concentration was similarly determined using Quant-iT PicoGreen dsDNA Assay Kit. 0.5-1.0 µg HT29 cell line genomic DNA was fragmented with a non-random sonication method, using a Covaris ultra sonicator E220 (Covaris, Woburn, MA). After shearing, the quality of the resulting DNA fragments (length ranged from 50 to 1 kb base pairs) was assessed with an Agilent Bioanalyzer system 2100 (Agilent, Santa Clara, CA).

[0380] Enrichment of methylated DNA: The DNA fragments containing methylated CpGs was captured by binding to methylation-specific antibodies, using the EpiMark® Methylated DNA Enrichment Kit from New England Biolabs, according to manufacturer's instructions (New England Biolabs, Ipswich, MA). DNA fragments containing methylated CpG sites were enriched by binding to the antibodies containing methyl-CpG binding domain. After a series of wash steps followed by magnetic capture, the enriched methylated DNA sample was eluted in a small volume of water by incubation at 65°C.

[0381] Primers and Probes: All primers used are listed in Table 45 above. All primers were purchased from Integrated DNA Technologies Inc. (IDT) (Coralville, IA). The PCR reverse primers which are used as primers for linear amplification, have 23 bp long tails comprising a universal primer sequence. After the linear amplification step, in the first PCR step, the universal primer was added to enhance amplification.

[0382] TET2- APOBEC conversion of DNA: New England Biolabs developed a two-enzyme protocol to mimic the equivalent of a bisulfite conversion step to determine the presence or absence of methylated or hydroxymethylated cytosines. The approach uses TET2 for conversion of 5mC (5-methyl cytosine) and 5hmC (5-hydroxy-methyl cytosine) through a cascade reaction into 5-carboxycytosine [i.e. 5-methylcytosine (5mC) → 5-hydroxymethylcytosine (5hmC) → 5-formylcytosine (5fC) → 5-carboxycytosine (5caC)], thus protecting 5mC and 5hmC, but not unmethylated C from deamination by APOBEC, (see Technical Report and Protocol with New England Biolabs product: NEBNext Enzymatic Methyl-seq Kit E7120, which is hereby incorporated by reference in its entirety). The description below is copied directly from this protocol.

[0383] Step1. Oxidation of 5-methylcytosine: The reaction mixtures are prepared as following: 28 ul methylated DNA (enriched or not enriched), 10 ul of TET2 Reaction buffer, 1 ul of Oxidation supplement, 1 ul of DTT, 1 ul of Oxidation enhancer, 4 ul of TET2. After thoroughly mixing, add 1 ul of 1249 fold diluted 500 mM Fe(II) solution, and incubate at 37 °C for 1 hour. Add 1 ul of stop reagent to the whole reaction and incubate at 37 °C for 30 min to stop the reaction.

[0384] Step 2 Clean up the TET2 converted DNA: Add 90 µl of resuspended NEBNext Sample Purification Beads to each sample. Incubate samples on bench top for at least 5 minutes at room temperature. Place the tubes against an appropriate magnetic stand to separate the beads from the supernatant. After 5 minutes (or when the solution is clear), carefully remove and discard the supernatant. Use 200 µl of 80% freshly prepared ethanol to wash the beads twice when the tubes is in the magnetic stand. Air dry the beads and add 17 µl of Elution Buffer to elute the DNA from the beads. Place the tube on the magnetic stand. After 3 minutes, transfer 16 µl of the supernatant to a new PCR tube.

[0385] Step 3, using Formamide to denature oxidized DNA: Add 4 µl Formamide to the 16 µl of oxidized DNA. Incubate at 85°C for 10 minutes in the pre-heated thermocycler. Immediately place on ice.

[0386] Step 4, Deamination of Cytosines: The deamination reaction mixture (prepared on ice) contains 20 µl of denatured DNA, 68 ul of water, 10 ul of APOBEC Reaction buffer, 1 ul

of BSA, 1 μ l of APOBEC enzyme. The mixture was incubated at 37 °C for 3 hours and then 4°C in a thermocycler.

[0387] Step 5 Clean up of Deaminated DNA: Add 100 μ l of re-suspended NEBNext Sample Purification Beads to each deaminated DNA sample. Mix well. Incubate samples on bench top for at least 5 minutes at room temperature. Place the tubes against an appropriate magnetic stand to separate the beads from the supernatant. After 5 minutes (or when the solution is clear), carefully remove and discard the supernatant. Using 200 μ l of 80% freshly prepared ethanol to wash the beads twice when the tubes in the magnetic stand. Air dry the beads for up to 90 seconds while the tubes are on the magnetic stand with the lid open. Remove the tubes from the magnetic stand. Elute the DNA target from the beads by adding 52 μ l of Elution Buffer and mix well, Place the tube on the magnetic stand. After 3 minutes transfer 50 μ l of the supernatant to a new PCR tube.

[0388] Linear Amplification Step. In a 25 μ l reaction volume, the linear amplification step was performed by mixing: 5 μ l of 5x Gotaq Flexi buffer (no Magnesium) (Promega, Madison, Wis.), 3.5 μ l of 25 mM MgCl₂ (Promega, Madison, Wis.), 0.5 μ l of 10 mM dNTPs (dATP, dCTP, dGTP and dTTP) (Promega, Madison, Wis.), 1.25 μ l of 20 plex gene specific reverse primers that have a long 23 bp universal sequence tail (concentration of each primer is 1 μ M in condition A, and 0.5 μ M in condition B), 0.5 μ l of tween 20 (5%), 0.9 μ l of 20 mU/ μ l RNaseH2 (diluted in RNaseH2 dilution buffer from IDT) (IDT), and 0.5 μ l of KlenTaq polymerase (DNA Polymerase Technology, St. Louis, Mo.) mixed with Platinum Taq Antibody (Invitrogen/Thermo Fisher, Waltham, Mass.) (the mixture is prepared by adding 1 μ l of KlenTaq polymerase at 50 U/ μ l to 10 μ l of Platinum Taq Antibody), and 5.0 μ l of DNA templates. There are four templates. The template A was 1 μ l of TET2-APOBEC deaminated HT29 cell line genomic DNA, the starting DNA amount is 1 μ g, which was sonicated but was not methylation enriched by antibody method. Template B was 1 μ l of TET2-APOBEC deaminated normal DNA, DNA starting amount is 1 μ g, which was sonicated but was not methylation enriched by antibody method. Template C was 1 μ l of TET2-APOBEC deaminated HT29 cell line genomic DNA, the starting DNA amount is 1 μ g, which was sonicated and was not methylation enriched by antibody method. Template D was 1 μ l of TET2-APOBEC deaminated normal DNA, DNA starting amount is 1 μ g, which was sonicated and was methylation enriched by antibody method. The reactions were run in a ProFlex PCR system thermocycler (Applied Biosystems/ThermoFisher, Waltham, Mass.) using the following program: 2 min at 94 °C, 40 cycles of (20 sec at 94°C, 40 sec at 60 °C, and 30 sec at 72 °C.), and a final hold at 4° C. After the reaction, 0.5 μ l of platinum Taq Antibodies were added in the reaction mixture to inhibit the KlenTaq DNA polymerase.

[0389] PCR reaction. Two each of 10-plex PCR reactions were carried out using the linear amplification products divided equally into two 10 µl parts. The multiplex reaction consisted of two 10-plex reactions with the addition of marker-specific forward primers, universal primer (which base pairing with the long tail of linear amplification primer.) and other PCR reagents. The PCR reaction was performed in a 20 µl of mixture prepared with 2 µl of Gotaq Flexi buffer 5x without Magnesium (Promega, Madison, Wis.), 1.4 µl of MgCl₂ at 25 mM (Promega, Madison, Wis.), 0.4 µl of dNTPs (with dATP, dCTP, dGTP and dUTP, 10 mM each) (Promega, Madison, Wis.), 0.2 µl of tween20(5%), 1 µl total of each corresponding 10 plex marker-specific forward primers at 0.5 µM each for condition A, and 0.25 µM for condition B, 0.4 µl of Antarctic Thermolabile UDG (1u/µl)(New England Biolab, Ipswich, MA), 0.29 µl of RNaseH2 (IDT) at 20 mU/µl, 1.6 µl of KlenTaq polymerase (DNA Polymerase Technology, St. Louis, Mo.) that was mixed with Platinum Taq Antibody (Invitrogen/Thermo Fisher, Waltham, Mass.) (The mixture is prepared by adding 1 µl of KlenTaq polymerase at 50 U/µl with 10 µl of Platinum Taq Antibody at 5 U/µl), and 9 µl of corresponding linear amplification products, and 1 µl of universal primer 2000 (20 uM). PCR reactions were carried out in a ProFlex PCR system thermocycler (Applied Biosystems/ ThermoFisher, Waltham, Mass.) using the following program : 10 min at 37 °C, 40 cycles of (20 sec at 94 °C, 40 sec at 60 °C. and 30 sec at 72 °C), 10 min at 99.5 °C., and a final hold at 4 °C.

[0390] LDR step. The LDR reaction was performed in a 10 µl of mixture that was prepared by combining: 4.82 µl of nuclease free water (IDT), 1 µl of 10X AK16D ligase reaction buffer [1x buffer contains 20 mM Tris-HCl pH 8.5 (Bio-Rad, Hercules, Calif.), 5 mM MgCl₂ (Sigma-Aldrich, St. Louis, Mo.), 50 mM KCl (Sigma-Aldrich, St. Louis, Mo.), 10 mM DTT (Sigma-Aldrich, St. Louis, Mo.) and 20 µg/ml of BSA (Sigma Aldrich, St. Louis, Mo.)], 0.25 µl of 40 mM DTT (Sigma-Aldrich, St. Louis, Mo.), 0.2 µl of 50 mM NAD⁺ (Sigma-Aldrich, St. Louis, Mo.), 0.25 µl of 20 mU/µl RNaseH2 (IDT), 0.2 µl of corresponding 10 plex specific LDR upstream and downstream probes at 500 nM each, 0.28 µl of purified AK16D ligase (at 0.88 µM), and 3 µl of the corresponding PCR amplification products from second steps. LDR reactions were run in a ProFlex PCR system thermocycler (Applied Biosystems/Thermo-Fisher; Waltham, Mass.) using the following program: 20 cycles of (10 sec at 94 °C, and 4 min at 60 °C) followed by a final hold at 4° C.

[0391] qPCR step. The qPCR is a uni-plex reaction, carried out for each marker. The qPCR step was performed in 10 µl of reaction volume by combining: 3 µl of nuclease free water (IDT), 5 µl of 2x TaqMan® Fast Universal PCR Master Mix (Fast Amplitaq, UDG and dUTP) from Applied Biosystems (Applied Biosystems/ThermoFisher; Waltham, Mass.), 0.5 µl of TaqMan™ Assay forward primer and reverse primer (concentration is 5 µM each primer), 0.5 µl

of 5 μ M Taqman™ probe, and 1 μ l of LDR reaction products. qPCR reactions are run in a ViiA7 real-time thermo-cycler from Applied Biosystems (Applied Biosystems/Thermo-Fisher; Waltham, Mass.), using MicroAmp® Fast-96-Well Reaction 0.1 ml plates sealed with MicroAmp™ Optical adhesive film (Applied Biosystems/ThermoFisher; Waltham, Mass.), with the following setting: fast block, Standard curve as experiment type, ROX as passive reference, Ct as quantification method (automatic threshold, but adjusted to 0.05 when needed), TAMRA as reporter, and NFQ-MGB as quencher. The program employed was: 2 min at 50 °C, and 45 cycles of (1 sec at 95 °C, and 20 sec at 60 °C.). The results are shown in Figures 70A-B and 71A-B and Table 46 below.

[illegible]

[illegible]

Example 2: Multiplexed Detection of 20 CRCM Markers Using ex-PCR-LDR-qPCR on Bisulfite Converted HT29 Cell Line DNA using long-tail primers and Universal Primer.

[0393] HT-29 colon adenocarcinoma cells were seeded in 60 cm² culture dishes in McCoy's 5A medium containing 4.5 g/l glucose, supplemented with 10% fetal calf serum, and kept in a humidified atmosphere containing 5% CO₂. Once cells reached 80-90% confluence, they were washed in Phosphate Buffered Saline (x3), and cells collected by centrifugation (500xg). Colon Cancer cell line genomic DNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA.), and its concentration measured using Quant-iT Pico green dsDNA Assay kit (Thermo-Fisher, Waltham, MA.). High molecular weight (>50 kb) genomic DNA (0.2 mg/ml) isolated from human blood (buffy coat) (Roche human genomic DNA) was purchased from Roche (Indianapolis, IN.). Its concentration was similarly determined using Quant-iT PicoGreen dsDNA Assay Kit. 0.5-1.0 µg HT29 cell line genomic DNA was fragmented with a non-random sonication method, using a Covaris ultra sonicator E220 (Covaris, Woburn, MA). After shearing, the quality of the resulting DNA fragments (length ranged from 50 to 1 kb base pairs) was assessed with an Agilent Bioanalyzer system 2100 (Agilent, Santa Clara, CA).

[0394] Enrichment of methylated DNA: The DNA fragments containing methylated CpGs was captured by binding to methylation-specific antibodies, using the EpiMark® Methylated DNA Enrichment Kit from New England Biolabs, according to manufacturer's instructions (New England Biolabs, Ipswich, MA). DNA fragments containing methylated CpG sites were enriched by binding to the antibodies containing methyl-CpG binding domain. After a series of wash steps followed by magnetic capture, the enriched methylated DNA sample was eluted in a small volume of water by incubation at 65°C.

[0395] Bisulfite conversion of DNA: Bisulfite conversion of cytosine bases in DNA was then carried out using the Cells-to-CpG Bisulfite Conversion kit from Applied Biosystem division of Thermo Fisher (Carlsbad, Calif.). 5 µl of Denaturation Reagent was added to 45 µl of methyl enriched genomic DNA, followed by the mixture's incubation at 50 °C for 10 min. This is followed by addition of 100 µl of Conversion Reagent, and incubated in a thermal cycler with the following program: 65 °C 30 min, 90°C 30 sec, 65°C 30 min, 90°C 30 sec, 65°C 30 min. 150 µl of converted DNA mixture was mixed with 600 µl of binding buffer in the binding column. The column was centrifuged at 10,000 rpm for 1 min, followed by discarding the flow through. The column was washed with 600 µl of washing buffer. 200 µl of Desulfonation Reagent was added to the column, followed by incubation at room temperature for 15 min. After spinning, the column was washed again with 400 µl of washing buffer. 50 µl of Elution Buffer was then added to the column to elute the bound DNA. The mostly single stranded, bisulfite

converted DNA was quantified with both Quant-iT Oli Green and Pico Green kit (Life Technologies/ThermoFisher, Waltham, Mass.).

[0396] Primers and Probes: All primers used are listed in Table 45 above. All primers were purchased from Integrated DNA Technologies Inc. (IDT) (Coralville, IA). The PCR reverse primers which are used as primers for linear amplification, have 23 bp long tails comprising a universal primer sequence. After the linear amplification step, in the first PCR step, the universal primer was added to enhance amplification.

[0397] Linear amplification step. In a 25 µl reaction volume, the linear amplification step was performed by mixing: 5 µl of 5x Gotaq Flexi buffer (no Magnesium) (Promega, Madison, Wis.), 3.5 µl of 25 mM MgCl₂ (Promega, Madison, Wis.), 0.5 µl of 10 mM dNTPs (dATP, dCTP, dGTP and dTTP) (Promega, Madison, Wis.), 1.25 µl of 20 plex gene specific reverse primers that have a long 23 bp universal sequence tail (concentration of each primer is 1 µM in condition A, and 0.5 µM in condition B), 0.5 µl of tween 20 (5%), 0.9 µl of 20 mU/µl RNaseH2 (diluted in RNaseH2 dilution buffer from IDT) (IDT), and 0.5 µl of KlenTaq polymerase (DNA Polymerase Technology, St. Louis, Mo.) mixed with Platinum Taq Antibody (Invitrogen/Thermo Fisher, Waltham, Mass.) (the mixture is prepared by adding 1 µl of KlenTaq polymerase at 50 U/µl to 10 µl of Platinum Taq Antibody), and 5.0 µl of DNA templates. The template was 5 µl of DNA mixture containing 200 GE of HT29 DNA and 7,500 GE of Roche DNA. (GE: Genome Equivalent). The DNA is methylation specific enriched and converted by bisulfite reaction. The reactions were run in a ProFlex PCR system thermocycler (Applied Biosystems/ ThermoFisher, Waltham, Mass.) using the following program: 2 min at 94 °C, 40 cycles of (20 sec at 94°C, 40 sec at 60 °C, and 30 sec at 72 °C.), and a final hold at 4° C. After the reaction, 0.5 ul of platinum Taq Antibodies were added in the reaction mixture to inhibit the KlenTaq DNA polymerase.

[0398] PCR reaction. Two each of 10-plex PCR reactions were carried out using the linear amplification products divided equally into two 10 µl parts. The multiplex reaction consisted of two 10-plex reactions with the addition of marker-specific forward primers, universal primer (which base pairing with the long tail of linear amplification primer.) and other PCR reagents.

[0399] The PCR reaction was performed in a 20 µl of mixture prepared with 2 µl of Gotaq Flexi buffer 5x without Magnesium (Promega, Madison, Wis.), 1.4 µl of MgCl₂ at 25 mM (Promega, Madison, Wis.), 0.4 µl of dNTPs (with dATP, dCTP, dGTP and dUTP, 10 mM each) (Promega, Madison, Wis.), 0.2 µl of tween20(5%), 1 µl total of each corresponding 10 plex marker-specific forward primers at 0.5 µM each for condition A, and 0.25 µM for condition B, 0.4 µl of Antarctic Thermolabile UDG (1u/µl)(New England Biolab, Ipswich, MA), 0.29 µl of

RNAseH2 (IDT) at 20 mU/μl , 1.6 μl of KlenTaq polymerase (DNA Polymerase Technology, St. Louis, Mo.) that was mixed with Platinum Taq Antibody (Invitrogen/Thermo Fisher, Waltham, Mass.) (The mixture is prepared by adding 1 μl of KlenTaq polymerase at 50 U/μl with 10 μl of Platinum Taq Antibody at 5 U/μl), and 9 μl of corresponding linear amplification products, and 1 μl of universal primer 2000 (20 uM) . PCR reactions were carried out in a ProFlex PCR system thermocycler (Applied Biosystems/ThermoFisher, Waltham, Mass.) using the following program : 10 min at 37 °C, 40 cycles of (20 sec at 94 °C, 40 sec at 60 °C. and 30 sec at 72 °C), 10 min at 99.5 °C., and a final hold at 4 °C.

[0400] LDR step. The LDR reaction was performed in a 10 μl of mixture that was prepared by combining: 4.82 μl of nuclease free water (IDT), 1 μl of 10X AK16D ligase reaction buffer [1x buffer contains 20 mM Tris-HCl pH 8.5 (Bio-Rad, Hercules, Calif.), 5 mM MgCl₂ (Sigma-Aldrich, St. Louis, Mo.), 50 mM KCl (Sigma-Aldrich, St. Louis, Mo.), 10 mM DTT (Sigma-Aldrich, St. Louis, Mo.) and 20 μg/ml of BSA (Sigma Aldrich, St. Louis, Mo.)], 0.25 μl of 40 mM DTT (Sigma-Aldrich, St. Louis, Mo.), 0.2 μl of 50 mM NAD⁺ (Sigma-Aldrich, St. Louis, Mo.), 0.25 μl of 20 mU/μl RNAseH2 (IDT), 0.2 μl of corresponding 10 plex specific LDR upstream and downstream probes at 500 nM each, 0.28 μl of purified AK16D ligase (at 0.88 μM), and 3 μl of the corresponding PCR amplification products from second steps. LDR reactions were run in a ProFlex PCR system thermocycler (Applied Biosystems/Thermo-Fisher; Waltham, Mass.) using the following program: 20 cycles of (10 sec at 94 °C, and 4 min at 60 °C) followed by a final hold at 4° C.

[0401] qPCR step. The qPCR is a uni-plex reaction, carried out for each marker. The qPCR step was performed in 10 μl of reaction volume by combining: 3 μl of nuclease free water (IDT), 5 μl of 2x TaqMan® Fast Universal PCR Master Mix (Fast Amplitaq, UDG and dUTP) from Applied Biosystems (Applied Biosystems/ThermoFisher; Waltham, Mass.), 0.5 μl of TaqMan™ Assay forward primer and reverse primer (concentration is 5 μM each primer), 0.5 μl of 5 μM Taqman™ probe, and 1 μl of LDR reaction products. qPCR reactions were run in a ViiA7 real-time thermo-cycler from Applied Biosystems (Applied Biosystems/Thermo-Fisher; Waltham, Mass.), using MicroAmp® Fast-96-Well Reaction 0.1 ml plates sealed with MicroAmp™ Optical adhesive film (Applied Biosystems/ThermoFisher; Waltham, Mass.), with the following setting: fast block, Standard curve as experiment type, ROX as passive reference, Ct as quantification method (automatic threshold, but adjusted to 0.05 when needed), TAMRA as reporter, and NFQ-MGB as quencher. The program employed was: 2 min at 50 °C, and 45 cycles of (1 sec at 95 °C, and 20 sec at 60 °C.). The results are shown in Figures 72A-B and 73A-B and Table 47 below.

Table 47. Ct values for each gene (genes 1-10) in Example 2.

CRC-Marker		1	2	3	4	5	6	7	8	9	10
Gene		GATA5-S2	ZNF54-2-S1	RCN-3-S1	MYO1-5B-S1	ANKRD13B-S1	FAM115A-S1	RGS10-S1	HCG4-S1	STK32B-S1	CNRIP1-S1
starting number for primer		5421	5431	5441	5451	5461	5471	5491	5501	5511	5521
200GE of HT29 + 7,500 GE of Roche	No Univ. Primer; 25 nM primers	11.1	12.6	15.9	17.3	24.0	14.1	10.4	7.8	27.4	21.8
200GE of HT29 + 7,500 GE of Roche	With Univ. Primer; 25 nM primers	9.4	8.9	11.7	21.8	12.4	13.1	20.7	6.0	23.8	15.0
1st Linaer Amp_N TC	25 nM primers in step 1 and 2	30.3	38.4	35.1	36.9	37.1	No Ct	40.4	34.8	28.8	No Ct
2nd PCR_NT C	25 nM primers in step 1 and 2	30.1	36.4	34.1	38.0	35.7	No Ct	No Ct	36.7	29.3	39.1
200GE of HT29 + 7,500 GE of Roche	No Univ. Primer; 12 nM primers	20.1	28.7	19.0	21.0	18.8	15.6	15.7	15.9	22.5	24.5
200GE of HT29 + 7,500 GE of Roche	With Univ. Primer; 12 nM primers	11.0	24.1	11.0	16.6	7.5	9.6	7.2	9.4	28.3	13.3
1st Linear Amp NTC	12 nM primers in step 1 and 2	28.5	42.7	35.0	37.7	36.5	41.1	No Ct	35.0	29.2	No Ct
2nd PCR_NTC	12 nM primers in step 1 and 2	30.2	36.5	34.5	38.6	34.6	No Ct	No Ct	34.3	29.5	39.3

Table 47 continued. Ct values for each gene (genes 11-20) in Example 2.

CRC-Marker		11	12	13	14	15	16	17	18	19	20
Gene		VIM-S1	CLIP4-S1	GSG1L-S1	PP1R16B-S1	KCNA3	GDF6-S1	ADHFE1-S1	THBD-S1	SEPT9-S4	SEMA3B-S1
starting number for primer		5001	5021	5051	5061	5071	5081	5101	5331	5351	5401
200GE of HT29 + 7,500 GE of Roche	No Univ. Primer; 25 nM primers	20.2	7.5	7.2	19.5	No Ct	21.6	11.7	10.0	10.0	17.8
200GE of HT29 + 7,500 GE of Roche	With Univ. Primer; 25 nM primers	No Ct	7.4	7.1	32.1	No Ct	17.9	18.1	6.5	9.8	9.0
1st Linaer Amp_N TC	25 nM primers in step 1 and 2	No Ct	No Ct	No Ct	31.6	43.3	24.4	33.8	No Ct	10.2	No Ct
2nd PCR_NT C	25 nM primers in step 1 and 2	No Ct	No Ct	No Ct	32.1	No Ct	29.9	33.8	No Ct	29.3	No Ct
200GE of HT29 + 7,500 GE of Roche	No Univ. Primer; 12 nM primers	17.9	11.7	16.8	18.9	30.9	19.0	16.6	16.2	10.7	11.8
200GE of HT29 + 7,500 GE of Roche	With Univ. Primer; 12 nM primers	8.6	7.3	6.1	8.8	42.3	12.5	19.5	8.6	7.1	5.8
1st Linear Amp NTC	12 nM primers in step 1 and 2	No Ct	No Ct	No Ct	32.5	34.2	30.2	33.8	No Ct	30.3	No Ct
2nd PCR_NTC	12 nM primers in step 1 and 2	No Ct	No Ct	No Ct	31.9	No Ct	30.1	33.7	No Ct	29.3	No Ct

[0402] Although preferred embodiments have been depicted and described in detail herein, it will be apparent to those skilled in the relevant art that various modifications, additions, substitutions, and the like can be made without departing from the spirit of the present

application and these are therefore considered to be within the scope of the present application as defined in the claims which follow.

WHAT IS CLAIMED:

1. A method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues, said method comprising:

providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues;

subjecting the nucleic acid molecules in the sample to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil residues to produce a treated sample;

providing one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules;

providing one or more primary oligonucleotide primer sets, each primary oligonucleotide primer set comprising (a) a first primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer, wherein the first or second primary oligonucleotide primer further comprises a 5' primer-specific portion;

blending the treated sample, the one or more first primary oligonucleotide primers of the primer sets, a deoxynucleotide mix, and a DNA polymerase to form one or more polymerase extension reaction mixtures;

subjecting the one or more polymerase extension reaction mixtures to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence;

blending the one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the

primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase to form one or more first polymerase chain reaction mixtures;

subjecting the one or more first polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof;

providing one or more oligonucleotide probe sets, each probe set comprising (a) a first oligonucleotide probe having a 5' primer-specific portion and a 3' DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and (b) a second oligonucleotide probe having a 5' DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion and a 3' primer-specific portion, and wherein the first and second oligonucleotide probes of a probe set are configured to hybridize, in a base specific manner, on a complementary nucleotide sequence of a first polymerase chain reaction product;

blending the first polymerase chain reaction products with a ligase and the one or more oligonucleotide probe sets to form one or more ligation reaction mixtures;

subjecting the one or more ligation reaction mixtures to one or more ligation reaction cycles whereby the first and second oligonucleotide probes of the one or more oligonucleotide probe sets are ligated together, when hybridized to their complementary sequences, to form ligated product sequences in the ligation reaction mixture wherein each ligated product sequence comprises the 5' primer-specific portion, the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and the 3' primer-specific portion;

providing one or more secondary oligonucleotide primer sets, each secondary oligonucleotide primer set comprising (a) a first secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the ligated product sequence and (b) a second secondary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the ligated product sequence;

blending the ligated product sequences, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil

(dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase to form one or more second polymerase chain reaction mixtures;

subjecting the one or more second polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products; and

detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

2. A method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues, said method comprising:

providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues;

subjecting the nucleic acid molecules in the sample to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample;

providing one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules;

providing one or more first primary oligonucleotide primer(s) that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more methylated or hydroxymethylated residue;

blending the treated sample, the one or more first primary oligonucleotide primers, a deoxynucleotide mix, and a DNA polymerase to form one or more polymerase extension reaction mixtures;

subjecting the one or more polymerase extension reaction mixtures to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence;

providing one or more secondary oligonucleotide primer sets, each secondary oligonucleotide primer set comprising (a) a first secondary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that is complementary to a portion of the polymerase extension product formed from the first primary oligonucleotide primer and (b) a second secondary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first secondary oligonucleotide primer;

blending the one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix, and a DNA polymerase to form one or more first polymerase chain reaction mixtures;

subjecting the one or more first polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures, and conditions suitable for carrying out two or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising a 5' primer-specific portion of the first secondary oligonucleotide primer, a DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence-specific or complement sequence-specific portion, and a complement of the 5' primer-specific portion of the second secondary oligonucleotide primer;

providing one or more tertiary oligonucleotide primer sets, each tertiary oligonucleotide primer set comprising (a) a first tertiary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first polymerase chain reaction products and (b) a second tertiary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the first polymerase chain reactions product sequence;

blending the first polymerase chain reaction products, the one or more tertiary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)

containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase to form one or more second polymerase chain reaction mixtures;

subjecting the one or more second polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products; and

detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

3. A method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues, said method comprising:

providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues;

subjecting the nucleic acid molecules in the sample to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample;

providing one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules present in the sample;

providing one or more primary oligonucleotide primer sets, each primary oligonucleotide primer set comprising (a) a first primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary

oligonucleotide primer, wherein the first or second primary oligonucleotide primer further comprises a 5' primer-specific portion;

blending the treated sample, the one or more first primary oligonucleotide primers of the primer sets, a deoxynucleotide mix, and a DNA polymerase to form one or more polymerase extension reaction mixtures;

subjecting the one or more polymerase extension reaction mixtures to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence;

blending the one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules in the reaction mixture, a deoxynucleotide mix, and a DNA polymerase to form one or more first polymerase chain reaction mixtures;

subjecting the one or more first polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the first polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reaction products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof;

providing one or more secondary oligonucleotide primer sets, each secondary oligonucleotide primer set comprising (a) a first secondary oligonucleotide primer having a 3' portion that is complementary to a portion of a first polymerase chain reaction product formed from the first primary oligonucleotide primer and (b) a second secondary oligonucleotide primer having a 3' portion that comprises a nucleotide sequence that is complementary to a portion of a first polymerase chain reaction product formed from the first secondary oligonucleotide primer;

blending the first polymerase chain reaction products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase to form one or more second polymerase chain reaction mixtures;

subjecting the one or more second polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out two or more polymerase

chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products; and

detecting and distinguishing the second polymerase chain reactions products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

4. A method for identifying, in a sample, one or more parent nucleic acid molecules containing a target nucleotide sequence differing from nucleotide sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues, said method comprising:

providing a sample containing one or more parent nucleic acid molecules potentially containing the target nucleotide sequence differing from the nucleotide sequences in other parent nucleic acid molecules by one or more methylated or hydroxymethylated residues;

subjecting the nucleic acid molecules in the sample to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues to produce a treated sample;

providing one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules present in the sample;

providing one or more primary oligonucleotide primer sets, each primary oligonucleotide primer set comprising (a) a first primary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a sequence in the parent nucleic acid molecule adjacent to the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence containing the one or more converted methylated or hydroxymethylated residue and (b) a second primary oligonucleotide primer having a 5' primer-specific portion and a 3' portion that comprises a nucleotide sequence that is complementary to a portion of an extension product formed from the first primary oligonucleotide primer;

blending the treated sample, the one or more first primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, a deoxynucleotide mix, and a DNA polymerase to form one or more polymerase extension reaction mixtures;

subjecting the one or more polymerase extension reaction mixtures to conditions suitable for carrying out one or more polymerase extension reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming primary extension products comprising the complement of the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence;

blending the one or more polymerase extension reaction mixtures comprising the primary extension products, the one or more second primary oligonucleotide primers of the one or more primary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules in the reaction mixture, a deoxynucleotide mix, and a DNA polymerase to form one or more first polymerase chain reaction mixtures;

subjecting the one or more first polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the polymerase chain reaction mixtures and for carrying out one or more first polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment, thereby forming first polymerase chain reactions products comprising the DNA repair enzyme and DNA deaminase enzyme-treated target nucleotide sequence or a complement thereof;

providing one or more secondary oligonucleotide primer sets, each secondary oligonucleotide primer set comprising (a) a first secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first polymerase chain reaction products or their complements and (b) a second secondary oligonucleotide primer comprising a nucleotide sequence that is complementary to the 3' primer-specific portion of the first polymerase chain reaction products or their complements;

blending the first polymerase chain reaction products, the one or more secondary oligonucleotide primer sets, the one or more enzymes capable of digesting deoxyuracil (dU)-containing nucleic acid molecules, a deoxynucleotide mix including dUTP, and a DNA polymerase to form one or more second polymerase chain reaction mixtures;

subjecting the one or more second polymerase chain reaction mixtures to conditions suitable for digesting deoxyuracil (dU)-containing nucleic acid molecules present in the second polymerase chain reaction mixtures and for carrying out one or more polymerase chain reaction cycles comprising a denaturation treatment, a hybridization treatment, and an extension treatment thereby forming second polymerase chain reaction products; and

detecting and distinguishing the second polymerase chain reaction products in the one or more second polymerase chain reaction mixtures to identify the presence of one or more parent nucleic acid molecules containing target nucleotide sequences differing from nucleotide

sequences in other parent nucleic acid molecules in the sample by one or more methylated or hydroxymethylated residues.

5. The method of any one of claims 1 through 4 further comprising:
contacting the sample with DNA repair enzymes to repair damaged DNA, abasic sites, oxidized bases, or nicks in the DNA.

6. The method of any one of claims 1 through 4 further comprising:
contacting the sample with at least a first methylation sensitive enzyme to form one or more restriction enzyme reaction mixtures prior to, or concurrent with, said blending to form one or more polymerase extension reaction mixtures, wherein said first methylation sensitive enzyme cleaves nucleic acid molecules in the sample that contain one or more unmethylated residues within at least one methylation sensitive enzyme recognition sequence, and whereby said detecting involves detection of one or more parent nucleic acid molecules containing the target nucleotide sequence, wherein said parent nucleic acid molecules originally contained one or more methylated or hydroxymethylated residues.

7. The method of any one of claims 1 through 4 further comprising:
contacting the sample with an immobilized methylated or hydroxymethylated nucleic acid binding protein or antibody to selectively bind and enrich for methylated or hydroxymethylated nucleic acid in the sample.

8. The method of any one of claims 1 through 4, wherein one or more primary or secondary oligonucleotide primers comprises a portion that has no or one nucleotide sequence mismatch when hybridized in a base-specific manner to the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof, but have one or more additional nucleotide sequence mismatches that interferes with polymerase extension when said primary or secondary oligonucleotide primers hybridize in a base-specific manner to a corresponding nucleotide sequence portion in DNA repair enzyme and DNA deaminase enzyme-treated unmethylated nucleic acid sequence or complement sequence thereof.

9. The method of any one of claims 1 through 4, wherein one or both primary oligonucleotide primers of the primary oligonucleotide primer set and/or one or both secondary oligonucleotide primers of the secondary oligonucleotide primer set have a 3' portion

comprising a cleavable nucleotide or nucleotide analogue and a blocking group, such that the 3' end of said primer or primers is unsuitable for polymerase extension, said method further comprising:

cleaving the cleavable nucleotide or nucleotide analog of one or both oligonucleotide primers during said hybridization treatment, thereby liberating free 3'OH ends on one or both oligonucleotide primers prior to said extension treatment.

10. The method of claim 9, wherein one or more primary or secondary oligonucleotide primers comprises a sequence that differs from the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof, said difference is located two or three nucleotide bases from the liberated free 3'OH end.

11. The method of claim 9, wherein the cleavable nucleotide comprises one or more RNA bases.

12. The method of any one of claims 1 through 4, further comprising:
providing one or more blocking oligonucleotide primers comprising one or more mismatched bases at the 3' end or one or more nucleotide analogs and a blocking group at the 3' end, such that the 3' end of said blocking oligonucleotide primer is unsuitable for polymerase extension when hybridized in a base-specific manner to wild-type nucleic acid sequence or complement sequence thereof, wherein said blocking oligonucleotide primer comprises a portion having a nucleotide sequence that is the same as a nucleotide sequence portion in the wild-type nucleic acid sequence or complement sequence thereof to which the blocking oligonucleotide primer hybridizes but has one or more nucleotide sequence mismatches to a corresponding nucleotide sequence portion in the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof and

blending the one or more blocking oligonucleotide primers with the sample or products subsequently produced from the sample prior to a polymerase extension reaction, polymerase chain reaction, or ligation reaction, whereby during the hybridization step said one or more blocking oligonucleotide primers preferentially hybridize in a base-specific manner to a wild-type nucleic acid sequence or complement sequence thereof, thereby interfering with polymerase extension or ligation during reaction of a primer or probes hybridized in a base-

specific manner to the DNA repair enzyme and DNA deaminase enzyme-treated unmethylated sequence or complement sequence thereof.

13. The method of claim 3, wherein the first secondary oligonucleotide primer has a 5' primer-specific portion and the second secondary oligonucleotide primer has a 5' primer-specific portion, said one or more secondary oligonucleotide primer sets further comprising a third secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the first secondary oligonucleotide primer and (d) a fourth secondary oligonucleotide primer comprising the same nucleotide sequence as the 5' primer-specific portion of the second secondary oligonucleotide primer.

14. The method of any one of claims 1 through 3 further comprising:
providing one or more third primary oligonucleotide primers comprising the same nucleotide sequence as the 5' primer-specific portion of the first or second primary oligonucleotide primer; and

blending the one or more third primary oligonucleotide primers in the one or more first polymerase chain reaction mixtures.

15. The method of any one of claims 1 through 4, wherein the DNA repair enzyme is the ten-eleven translocation (TET2) dioxygenase and the DNA deaminase enzyme is an apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC cytidine deaminase).

16. The method of claim 1, wherein the second oligonucleotide probe of the oligonucleotide probe set further comprises a unitaq detection portion, thereby forming ligated product sequences comprising the 5' primer-specific portion, the target-specific portions, the unitaq detection portion, and the 3' primer-specific portion, said method further comprising:

providing one or more unitaq detection probes, wherein each unitaq detection probe hybridizes to a complementary unitaq detection portion and said detection probe comprises a quencher molecule and a detectable label separated from the quencher molecule;

adding the one or more unitaq detection probes to the second polymerase chain reaction mixture; and

hybridizing the one or more unitaq detection probes to complementary unitaq detection portions on the ligated product sequence or complement thereof during said subjecting the second polymerase chain reaction mixture to conditions suitable for one or more polymerase

chain reaction cycles, wherein the quencher molecule and the detectable label are cleaved from the one or more unitaq detection probes during the extension treatment and said detecting involves the detection of the cleaved detectable label.

17. The method of any one of claims 2 through 4, wherein one primary oligonucleotide primer or one secondary oligonucleotide primer further comprises a unitaq detection portion, thereby forming extension product sequences comprising the 5' primer-specific portion, the target-specific portions, the unitaq detection portion, and the complement of the other 5' primer-specific portion, and complements thereof, said method further comprising:

providing one or more unitaq detection probes, wherein each unitaq detection probe hybridizes to a complementary unitaq detection portion and said detection probe comprises a quencher molecule and a detectable label separated from the quencher molecule;

adding the one or more unitaq detection probes to the one or more polymerase chain reaction mixtures; and

hybridizing the one or more unitaq detection probes to complementary unitaq detection portions on the ligated product sequence or complement thereof during polymerase chain reaction cycles after the first polymerase chain reaction, wherein the quencher molecule and the detectable label are cleaved from the one or more unitaq detection probes during the extension treatment and said detecting involves the detection of the cleaved detectable label.

18. The method of claim 1, wherein one or both oligonucleotide probes of the oligonucleotide probe set comprises a portion that has no or one nucleotide sequence mismatch when hybridized in a base-specific manner to the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof, but have one or more additional nucleotide sequence mismatches that interferes with ligation when said oligonucleotide probe hybridizes in a base-specific manner to a corresponding nucleotide sequence portion in the DNA repair enzyme and DNA deaminase enzyme-treated unmethylated nucleic acid sequence or complement sequence thereof.

19. The method of claim 1, wherein the 3' portion of the first oligonucleotide probe of the oligonucleotide probe set comprises a cleavable nucleotide or nucleotide analogue and a blocking group, such that the 3' end is unsuitable for polymerase extension or ligation, said method further comprising;

cleaving the cleavable nucleotide or nucleotide analog of the first oligonucleotide probe when said probe is hybridized to its complementary target nucleotide sequence of the primary extension product, thereby liberating a 3'OH on the first oligonucleotide probe prior to said ligating.

20. The method of claim 19, wherein one or more first oligonucleotide probe of the oligonucleotide probe set comprises a sequence that differs from the target nucleic acid sequence or DNA repair enzyme and DNA deaminase enzyme-treated methylated or hydroxymethylated nucleic acid sequence or complement sequence thereof, said difference is located two or three nucleotide bases from the liberated free 3'OH end.

21. The method of claim 1, wherein the second oligonucleotide probe has, at its 5' end, an overlapping identical nucleotide with the 3' end of the first oligonucleotide probe, and, upon hybridization of the first and second oligonucleotide probes of a probe set at adjacent positions on a complementary target nucleotide sequence of a primary extension product to form a junction, the overlapping identical nucleotide of the second oligonucleotide probe forms a flap at the junction with the first oligonucleotide probe, said method further comprising:

cleaving the overlapping identical nucleotide of the second oligonucleotide probe with an enzyme having 5' nuclease activity thereby liberating a phosphate at the 5' end of the second oligonucleotide probe prior to said ligating.

22. The method of claim 1, wherein the one or more oligonucleotide probe sets further comprise a third oligonucleotide probe having a target-specific portion, wherein the second and third oligonucleotide probes of a probe set are configured to hybridize adjacent to one another on the target nucleotide sequence with a junction between them to allow ligation between the second and third oligonucleotide probes to form a ligated product sequence comprising the first, second, and third oligonucleotide probes of a probe set.

23. The method of any one of claims 1 through 22, wherein the sample is selected from the group consisting of tissue, cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, cell-free circulating nucleic acids, cell-free circulating tumor nucleic acids, cell-free circulating fetal nucleic acids in pregnant woman, circulating tumor cells, tumor, tumor biopsy, and exosomes.

24. The method of any one of claims 1 through 22, wherein the one or more target nucleotide sequences are low-abundance nucleic acid molecules comprising one or more nucleotide base mutations, insertions, deletions, translocations, splice variants, mRNA, lncRNA, ncRNA, miRNA variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicing, exon insertions, exon deletions, intron insertions, or other rearrangement at the genome level and/or methylated or hydroxymethylated nucleotide bases.

25. The method of claim 24, wherein the low-abundance nucleic acid molecules with one or more nucleotide base mutations, insertions, deletions, translocations, splice variants, mRNA, lncRNA, ncRNA, miRNA variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicings, exon insertions, exon deletions, intron insertions, or other rearrangement at the genome level, and/or methylated or hydroxymethylated nucleotide bases are identified and distinguished from a high-abundance of nucleic acid molecules in the sample having a similar nucleotide sequence as the low abundance nucleic acid molecules but without the one or more nucleotide base mutations, insertions, deletions, translocations, splice variants, mRNA, lncRNA, ncRNA, miRNA variants, alternative transcripts, alternative start sites, alternative coding sequences, alternative non-coding sequences, alternative splicing, exon insertions, exon deletions, intron insertions, or other rearrangement at the genome level, and/or methylated or hydroxymethylated nucleotide bases.

26. The method of claim 25, wherein the copy number of one or more low-abundance target nucleotide sequences are quantified relative to the copy number of the high-abundance nucleic acid molecules in the sample.

27. The method of any one of claims 1 through 22, wherein the one or more target nucleotide sequences are quantified or enumerated.

28. The method of claim 27, wherein the one or more target nucleotide sequences are quantified or enumerated relative to other nucleotide sequences in the sample or other samples undergoing the identical subsequent steps.

29. The method of claim 28, wherein the relative copy number of one or more target nucleotide sequences are quantified or enumerated.

30. The method of any one of claims 1 through 22, further comprising:
diagnosing or prognosing a disease state based on said identifying.

31. The method of any one of claims 1 through 22, further comprising:
distinguishing a genotype or disease predisposition based on said identifying.

32. A method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 6-12 markers, 12-24 markers, 24-36 markers, 36-48 markers, 48-72 markers, 72-96 markers, or > 96 markers, wherein each marker in a given set is selected by having any one or more of the following criteria:

present, or above a cutoff level, in > 50% of biological samples of the disease cells or tissue from individuals diagnosed with the disease state;

absent, or below a cutoff level, in > 95% of biological samples of the normal cells or tissue from individuals without the disease state;

present, or above a cutoff level, in > 50% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with the disease state;

absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without the disease state;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with the disease state;

and, wherein at least 50% of the markers in a set each comprise one or more methylated or hydroxymethylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated or hydroxymethylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with the disease state, said method comprising:

obtaining the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the

biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with the disease state if a minimum of 2 or 3 markers are present or above a cutoff level in a marker set comprising from 6-12 markers; or a minimum of 3, 4, or 5 markers are present or above a cutoff level in a marker set comprising from 12-24 markers; or a minimum of 3, 4, 5, or 6 markers are present or above a cutoff level in a marker set comprising from 24-36 markers; or a minimum of 4, 5, 6, 7, or 8 markers are present or above a cutoff level in a marker set comprising from 36-48 markers; or a minimum of 6, 7, 8, 9, 10, 11, or 12 markers are present or above a cutoff level in a marker set comprising from 48-72 markers, or a minimum of 7, 8, 9, 10, 11, 12 or 13 markers are present or above a cutoff level in a marker set comprising from 72-96 markers, or a minimum of 8, 9, 10, 11, 12, 13 or "n"/12 markers are present or above a cutoff level in a marker set comprising 96 to "n" markers, when "n" > 168 markers.

33. A method of diagnosing or prognosing a disease state of a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-

specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 48-72 total cancer markers, 72-96 total cancer markers or ≥ 96 total cancer markers, wherein on average greater than one quarter such markers in a given set cover each of the aforementioned major cancers being tested, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in $> 50\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in $> 50\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer, said method comprising:

obtaining a biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-

hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 48-72 total cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 72-96 total cancer markers; or a minimum of 6 or “n”/18 markers are present or are above a cutoff level in a marker set comprising 96 to “n” total cancer markers, when “n” > 96 total cancer markers.

34. The method of claim 33, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in > 66% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in > 66% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in > 95% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer.

35. A method of diagnosing or prognosing a disease state of and identifying the most likely specific tissue(s) of origin of a solid tissue cancer in the following groups: Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast

lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 36-48 group-specific cancer markers, 48-64 group-specific cancer markers or ≥ 64 group-specific cancer markers, wherein on average greater than one third such markers in a given set cover each of the aforementioned cancers being tested within that group, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in $> 50\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in $> 50\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer, said method comprising:

obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood,

plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dextyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 36-48 group-specific cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 48-64 group-specific cancer markers; or a minimum of 6 or "n"/12 markers are present or are above a cutoff level in a marker set comprising 64 to "n" group-specific cancer markers, when "n" > 64 group-specific cancer markers.

36. The method of claim 35, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in > 66% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in > 66% of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer.

37. A method of diagnosing or prognosing a disease state of a gastrointestinal cancer including colorectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 6-12 markers, 12-18 markers, 18-24 markers, 24-36 markers, 36-48 markers or ≥ 48 markers, wherein each marker is selected by having any one or more of the following criteria for gastrointestinal cancer:

present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with gastrointestinal cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without gastrointestinal cancer;

present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with gastrointestinal cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without gastrointestinal cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with gastrointestinal cancer;

and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with gastrointestinal cancer, said method comprising:

obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or

cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with gastrointestinal cancer if a minimum of 2, 3 or 4 markers are present or are above a cutoff level in a marker set comprising from 6-12 markers; or a minimum of 2, 3, 4, or 5 markers are present or are above a cutoff level in a marker set comprising from 12-18 markers; or a minimum of 3, 4, 5, or 6 markers are present or are above a cutoff level in a marker set comprising from 18-24 markers; or a minimum of 3, 4, 5, 6, 7, or 8 markers are present or are above a cutoff level in a marker set comprising from 24-36 markers; or a minimum of 4, 5, 6, 7, 8, 9, or 10 markers are present or are above a cutoff level in a marker set comprising from 36-48 markers; or a minimum of 5, 6, 7, 8, 9, 10, 11, 12, or "n"/12 markers are present or are above a cutoff level in a marker set comprising 48 to "n" markers, when "n" > 48 markers.

38. A method of diagnosing or prognosing a disease state of a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-

specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 36-48 total cancer markers, 48-64 total cancer markers, or ≥ 64 total cancer markers, wherein on average greater than half of such markers in a given set cover each of the aforementioned major cancers being tested, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

- present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

- absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer;

- present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

- absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

- present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

- and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer, said method comprising:

- obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

- fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

- subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-

hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set comprising from 36-48 total cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 48-64 total cancer markers; or a minimum of 6 or “n”/12 markers are present or are above a cutoff level in a marker set comprising 64 to “n” total cancer markers, when “n” > 96 total cancer markers.

39. A method of diagnosing or prognosing a disease state of and identifying the most likely specific tissue(s) of origin of a solid tissue cancer in the following groups: Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in > 75% of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in > 95% of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

and, wherein at least 50% of the markers in a set each comprise one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer, said method comprising:

obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into dexoyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals are diagnosed or prognosed with a solid-tissue cancer if a minimum of 4 markers are present or are above a cutoff level in a marker set

comprising from 24-36 group-specific cancer markers; or a minimum of 5 markers are present or are above a cutoff level in a marker set comprising from 36-48 group-specific cancer markers; or a minimum of 6 or “n”/8 markers are present or are above a cutoff level in a marker set comprising 48 to “n” group-specific cancer markers, when “n” > 48 group-specific cancer markers.

40. A method of diagnosing or prognosing a disease state to guide and monitor treatment of a solid tissue cancer in one or more of the following groups; Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

and, wherein at least 50% of the markers in a set each comprise of one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer, said method comprising:

obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals with a given tissue-specific cancer will on average have from approximately one-quarter to about one-half or more of the markers scored as present, or are above a cutoff level in the tested marker set, wherein to guide and monitor subsequent treatment, a portion or all of the identified markers scored as present or the identified markers as above a cutoff level in the tested marker set are deemed the "patient-specific marker set", and retested on a subsequent biological sample from the individual during the treatment protocol, to monitor for loss of marker signal, wherein if a minimum of 3 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 12-24 markers; or if a minimum of 4 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 24-36 markers; or a minimum of 5 markers remain present or remain above a cutoff level in a patient-specific marker set comprising from 36-48 markers; or a

minimum of 6 or “n”/8 markers remain present or remain above a cutoff level in a patient-specific marker set comprising 48 to “n” markers, when “n” > 48 markers after the treatment protocol has been administered, then the continuing presence of said markers may guide a decision to change the cancer treatment therapy.

41. A method of diagnosing or prognosing a disease state for recurrence of a solid tissue cancer in one or more of the following groups; Group 1 (colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma); Group 2 (breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma); Group 3 (lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma); Group 4 (prostate adenocarcinoma, invasive urothelial bladder cancer); and/or Group 5 (liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma) based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, wherein the plurality of markers is in a set comprising from 24-36 group-specific cancer markers, 36-48 group-specific cancer markers, or ≥ 48 group-specific cancer markers, wherein on average greater than one half of such markers in a given set cover each of the aforementioned cancers being tested within that group, wherein each marker in a given set for a given solid tissue cancer is selected by having any one or more of the following criteria for that solid tissue cancer:

present, or above a cutoff level, in $> 75\%$ of biological samples of a given cancer tissue from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples of the normal tissue from individuals without that given solid tissue cancer;

present, or above a cutoff level, in $> 75\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

absent, or below a cutoff level, in $> 95\%$ of biological samples comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals without that given solid tissue cancer;

present with a z-value of > 1.65 in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from individuals diagnosed with a given solid tissue cancer;

and, wherein at least 50% of the markers in a set each comprise of one or more methylated residues, and/or wherein at least 50% of the markers in a set that are present, or

above a cutoff level, or present with a z-value of > 1.65 comprise of one or more methylated residues, in the biological sample comprising cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, or fractions thereof, from at least 50% of individuals diagnosed with a given solid tissue cancer, said method comprising:

obtaining the biological sample, the biological sample including cell-free DNA, RNA, and/or protein originating from the cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

fractionating the sample into one or more fractions, wherein at least one fraction comprises exosomes, tumor-associated vesicles, other protected states, or cell-free DNA, RNA, and/or protein;

subjecting the nucleic acid molecules in one or more fractions to a treatment with one or more DNA repair enzymes under conditions suitable to convert 5-methylated and 5-hydroxymethylated cytosine residues to 5-carboxycytosine residues, followed by treatment with one or more DNA deamination enzymes under conditions suitable to convert unmethylated cytosine but not 5-carboxycytosine residues into deoxyuracil (dU) residues;

carrying out at least two enrichment steps for 50% or more disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers during either said fractionating step and/or by carrying out a nucleic acid amplification step; and

performing one or more assays to detect and distinguish the plurality of cancer - specific and/or cell/tissue-specific DNA, RNA, and/or protein markers, thereby identifying their presence or levels in the sample, wherein individuals with a given tissue-specific cancer will on average have from approximately one-quarter to about one-half or more of the markers scored as present, or are above a cutoff level in the tested marker set, wherein to monitor for recurrence, a portion or all of the markers scored as being present, or the markers scored as above a cutoff level in the tested marker set are deemed the "patient-specific marker set", and retested on subsequent biological samples from the individual after a successful treatment, to monitor for gain of marker signal, wherein if a minimum of 3 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 12-24 markers; or if a minimum of 4 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 24-36 markers; or a minimum of 5 markers reappear or rise above a cutoff level in a patient-specific marker set comprising from 36-48 markers; or a minimum of 6 or " n "/8 markers reappear or rise above a cutoff level in a patient-specific marker set comprising 48 to " n " markers, when " n " > 48 markers after the treatment protocol has been administered, then the reappearance or rise or

rise above a cutoff level in a patient-specific marker set may guide a decision to resume the cancer treatment therapy or change to a new cancer treatment therapy.

42. The method of any one of claims 32 through 41, wherein the at least two enrichment steps comprise of one or more of the following steps:

capturing or separating exosomes or extracellular vesicles or markers in other protected states; capturing or separating a platelet fraction; capturing or separating circulating tumor cells; capturing or separating RNA-containing complexes; capturing or separating cfDNA-nucleosome or differentially modified cfDNA-histone complexes; capturing or separating protein targets or protein target complexes; capturing or separating auto-antibodies; capturing or separating cytokines; capturing or separating methylated or hydroxymethylated cfDNA; capturing or separating marker specific DNA, cDNA, miRNA, lncRNA, ncRNA, or mRNA, or amplified complements, by hybridization to complementary capture probes in solution, on magnetic beads, or on a microarray; amplifying miRNA markers, non-coding RNA markers (lncRNA & ncRNA markers), mRNA markers, exon markers, splice-variant markers, translocation markers, or copy number variation markers in a linear or exponential manner via a polymerase extension reaction, polymerase chain reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific polymerase chain reaction, reverse-transcription reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific ligation reaction, and/or ligation reaction, using DNA polymerase, reverse transcriptase, DNA ligase, RNA ligase, DNA repair enzyme, DNA deaminase enzyme, RNase, RNaseH2, endonuclease, restriction endonuclease, exonuclease, CRISPR, DNA glycosylase or combinations thereof; selectively amplifying one or more target regions containing mutation markers or DNA repair enzyme and DNA deaminase enzyme-treated -converted DNA methylation markers, while suppressing amplification of the target regions containing DNA repair enzyme and DNA deaminase enzyme-treated unmethylated sequence or complement sequence thereof, in a linear or exponential manner via a polymerase extension reaction, polymerase chain reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific polymerase chain reaction, reverse-transcription reaction, DNA repair enzyme and DNA deaminase enzyme-treated -methyl-specific ligation reaction, and/or ligation reaction, using DNA polymerase, reverse transcriptase, DNA ligase, RNA ligase, DNA repair enzyme, DNA deaminase enzyme, RNase, RNaseH2, endonuclease, restriction endonuclease, exonuclease, CRISPR, DNA glycosylase or combinations thereof; preferentially extending, ligating, or amplifying one or more primers or probes whose 3'-OH end has been liberated in an enzyme and sequence-dependent process; using one or more blocking oligonucleotide primers comprising one or more mismatched bases

at the 3' end or comprising one or more nucleotide analogs and a blocking group at the 3' end under conditions that interfere with polymerase extension or ligation during said reaction of target-specific primer or probes hybridized in a base-specific manner to DNA repair enzyme and DNA deaminase enzyme-treated unmethylated sequence or complement sequence thereof.

43. The method of any one of claims 32 through 42, wherein the one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA, or protein markers comprise one or more of the following:

a quantitative real-time PCR method (qPCR); a reverse transcriptase-polymerase chain reaction (RTPCR) method; a DNA repair enzyme and DNA deaminase-treated- qPCR method; a digital PCR method (dPCR); a DNA repair enzyme and DNA deaminase-treated- dPCR method; a ligation detection method, a ligase chain reaction, a restriction endonuclease cleavage method; a DNA or RNA nuclease cleavage method; a micro-array hybridization method; a peptide-array binding method; an antibody-array method; a Mass spectrometry method; a liquid chromatography-tandem mass spectrometry (LC-MS/MS) method; a capillary or gel electrophoresis method; a chemiluminescence method; a fluorescence method; a DNA sequencing method; a DNA repair enzyme and DNA deaminase-treated -DNA sequencing method; an RNA sequencing method; a proximity ligation method; a proximity PCR method; a method comprising immobilizing an antibody-target complex; a method comprising immobilizing an aptamer-target complex; an immunoassay method; a method comprising a Western blot assay; a method comprising an enzyme linked immunosorbent assay (ELISA); a method comprising a high-throughput microarray-based enzyme-linked immunosorbent assay (ELISA); a method comprising a high-throughput flow-cytometry-based enzyme-linked immunosorbent assay (ELISA).

44. The method of any one of claims 32 through 43, wherein the one or more cutoff levels of the one or more assays to detect and distinguish the plurality of disease-specific and/or cell/tissue-specific DNA, RNA, or protein markers comprise one or more of the following calculations, comparisons, or determinations, in the one or more marker assays comparing samples from the disease vs. normal individual:

the marker ΔC_t value is > 2 ; the marker ΔC_t value is > 4 ; the ratio of detected marker-specific signal is > 1.5 ; the ratio of detected marker-specific signal is > 3 ; the ratio of marker concentrations is > 1.5 ; the ratio of marker concentrations is > 3 ; the enumerated marker-specific signals differ by $> 20\%$; the enumerated marker-specific signals differ by $> 50\%$; the marker-specific signal from a given disease sample is $> 85\%$; $> 90\%$; $> 95\%$; $> 96\%$; $> 97\%$; or

> 98% of the same marker-specific signals from a set of normal samples; the marker-specific signal from a given disease sample has a z-score of > 1.03; > 1.28; > 1.65; > 1.75; > 1.88; or > 2.05 compared to the same marker-specific signals from a set of normal samples.

45. A two-step method of diagnosing or prognosing a disease state of cells or tissue based on identifying the presence or level of a plurality of disease-specific and/or cell/tissue-specific DNA, RNA, and/or protein markers in a biological sample of an individual, said two-step method comprising:

obtaining a biological sample, the biological sample including exosomes, tumor-associated vesicles, markers within other protected states, cell-free DNA, RNA, and/or protein originating from the potentially disease state cells or tissue and from one or more other tissues or cells, wherein the biological sample is selected from the group consisting of cells, serum, blood, plasma, amniotic fluid, sputum, urine, bodily fluids, bodily secretions, bodily excretions, and fractions thereof;

applying a first step to the biological samples with an overall sensitivity of > 80% and an overall specificity of > 90% or an overall Z-score of > 1.28 to identify individuals more likely to be diagnosed or prognosed with the disease state; and

applying a second step to biological samples from those individuals identified in the first step with an overall specificity of > 95% or an overall Z-score of > 1.65 to diagnose or prognose individuals with the disease state, wherein said applying the first step and/or said applying the second step is carried out using the method of one of claims 32 through 44.

46. The method of any one of claims 32 through 45, wherein the disease state is a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated cytosine residues of a CpG sequence, or the complement of one or more methylated cytosine residues of a CpG sequence selected from the list in Figure 42 or in Figure 58.

47. The method of any one of claims 32 through 45, wherein the disease state is a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma,

esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated residues of one or more chromosomal sub-regions selected from the list in Figure 43 or in Figure 59.

48. The method of any one of claims 32 through 45, wherein the disease state is a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more miRNA sequences selected from the group consisting of (mir ID , Gene ID): hsa-mir-21 , MIR21; hsa-mir-182 , MIR182; hsa-mir-454 , MIR454; hsa-mir-96 , MIR96; hsa-mir-183 , MIR183; hsa-mir-549 , MIR549; hsa-mir-301a , MIR301A; hsa-mir-548f-1 , MIR548F1; hsa-mir-301b , MIR301B; hsa-mir-103-1 , MIR1031; hsa-mir-18a , MIR18A; hsa-mir-147b , MIR147B; hsa-mir-4326 , MIR4326; and hsa-mir-573 , MIR573 or one or more lncRNA or ncRNA sequences selected from the list in Figure 39.

49. The method of any one of claims 32 through 45, wherein the disease state is a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more Exon RNA sequences selected from the list in Figure 40.

50. The method of any one of claims 32 through 45, wherein the disease state is a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein, the one or more markers in a set comprise one or more mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product selected from the list in Figure 41 or from the group consisting of: (Protein name, UniProt ID); Uncharacterized protein C19orf48, Q6RUI8; Protein FAM72B, Q86X60; Protein FAM72D, Q6L9T8; Hydroxyacylglutathione hydrolase-like protein, Q6PII5; Putative methyltransferase NSUN5, Q96P11; RNA pseudouridylate synthase domain-containing protein 1, Q9UJJ7; Collagen triple helix repeat-containing protein 1, Q96CG8; Interleukin-11, P20809; Stromelysin-2, P09238; Matrix metalloproteinase-9, P14780; Podocan-like protein 1, Q6PEZ8; Putative peptide YY-2, Q9NRI6; Osteopontin, P10451; Sulfhydryl oxidase 2, Q6ZRP7; Glypican-2, Q8N158; Macrophage migration inhibitory factor, P14174; Peptidyl-prolyl cis-trans isomerase A, P62937; and Calreticulin, P27797.

51. The method of any one of claims 32 through 45, wherein the disease state is a solid tissue cancer including colorectal adenocarcinoma, stomach adenocarcinoma, esophageal carcinoma, breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, uterine carcinosarcoma, lung adenocarcinoma, lung squamous cell carcinoma, head & neck squamous cell carcinoma, prostate adenocarcinoma, invasive urothelial bladder cancer, liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more mutations, insertions, deletions, copy number changes, or expression changes in a gene selected from the group consisting of TP53 (tumor protein p53), TTN (titin), MUC16 (mucin 16), and KRAS (Kirsten rat sarcoma viral oncogene homolog).

52. The method of any one of claims 32 through 45, wherein the disease state is colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated

cytosine residues of a CpG sequence, or the complement of one or more methylated cytosine residues of a CpG sequence selected from the list in Figure 28 or in Figure 45 or in Figure 60.

53. The method of any one of claims 32 through 45, wherein the disease state is colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, wherein at least 50% of the markers in a set each comprise of one or more methylated residues of one or more chromosomal sub-regions selected from the list in Figure 29 or in Figure 46 or in Figure 61.

54. The method of any one of claims 32 through 45, wherein the disease state is colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, wherein the one or more markers in a set comprise one or more miRNA sequences selected from the list in Figure 23 or the group of (mir ID , Gene ID): hsa-mir-624 , MIR624, or one or more lncRNA or ncRNA sequences selected from the list in Figure 24 or the group of [Gene ID, Coordinate (GRCh38)], ENSEMBL ID: LINC01558, chr6:167784537-167796859, or ENSG00000146521.8..

55. The method of any one of claims 32 through 45, wherein the disease state is colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, wherein the one or more markers in a set comprise one or more Exon RNA sequences selected from the list in Figure 25 or in Figure 44.

56. The method of any one of claims 32 through 45, wherein the disease state is colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, wherein, the one or more markers in a set comprise one or more mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product selected from the list in Figure 26 or Figure 27, or from the group consisting of: (Gene Symbol , Chromosome Band , Gene Title , UniProt ID): SELE , 1q22-q25 , selectin E , P16581; OTUD4 , 4q31.21 , OTU domain containing 4 , Q01804; BPI , 20q11.23 , bactericidal/permeability-increasing protein , P17213; ASB4 , 7q21-q22 , ankyrin repeat and SOCS box containing 4 , Q9Y574; C6orf123 , 6q27 , chromosome 6 open reading frame 123 , Q9Y6Z2; KPNA3 , 13q14.3 , karyopherin alpha 3 (importin alpha 4) , and O00505; NUP98 , 11p15 , nucleoporin 98kDa , P52948 or group of : (Protein name , UniProt ID); Bactericidal permeability-increasing protein (BPI) (CAP 57), or P17213.

57. The method of any one of claims 32 through 45, wherein the disease state is colon adenocarcinoma, rectal adenocarcinoma, stomach adenocarcinoma, or esophageal carcinoma, wherein the one or more markers in a set comprise one or more mutations, insertions, deletions, copy number changes, or expression changes in a gene selected from the group consisting of APC (APC regulator of WNT signaling pathway), ATM (ATM serine/threonine kinase), CSMD1 (CUB and Sushi multiple domains 1), DNAH11 (dynein axonemal heavy chain 11), DST (dystonin), EP400 (E1A binding protein p400), FAT3 (FAT atypical cadherin 3), FAT4 (FAT atypical cadherin 4), FLG (filaggrin), GLI3 (GLI family zinc finger 3), KRAS (Kirsten rat sarcoma viral oncogene homolog), LRP1B (LDL receptor related protein 1B), MUC16 (mucin 16, cell surface associated), OBSCN (obscurin, cytoskeletal calmodulin and titin-interacting RhoGEF), PCLO (piccolo presynaptic cytomatrix protein), PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha), RYR2 (ryanodine receptor 2), SYNE1 (spectrin repeat containing nuclear envelope protein 1), TP53 (tumor protein p53), TTN (titin), and UNC13C (unc-13 homolog C).

58. The method of any one of claims 32 through 45, wherein the disease state is breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma, wherein, at least 50% of the markers in a set each comprise one or more methylated cytosine residues of a CpG sequence, or the complement of one or more methylated cytosine residues of a CpG sequence selected from the list in Figure 47 or in Figure 62.

59. The method of any one of claims 32 through 45, wherein the disease state is breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma, wherein at least 50% of the markers in a set each comprise one or more methylated residues of one or more chromosomal sub-regions selected from the list in Figure 48 or in Figure 63.

60. The method of any one of claims 32 through 45, wherein the disease state is breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma, wherein the one or more markers in a set comprise one or more miRNA sequences (mir ID, Gene ID) selected from hsa-mir-1265, MIR1265.

61. The method of any one of claims 32 through 45, wherein the disease state is breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma, wherein the one or more markers in a set comprise one or more Exon RNA sequences selected from the group consisting of (Exon location, Gene): chr2:179209013-179209087:+ , OSBPL6; chr2:179251788-179251866:+ , OSBPL6; and chr2:179253736-179253880:+ , OSBPL6.

62. The method of any one of claims 32 through 45, wherein the disease state is breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma, wherein the one or more markers in a set comprise one or more mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product selected from the group consisting of (Gene Symbol , Chromosome Band , Gene Title , UniProt ID): RSPO2 , 8q23.1 , R-spondin 2 , Q6UXX9; KLC4 , 6p21.1 , kinesin light chain 4 , Q9NSK0; GLRX , 5q14 , glutaredoxin (thioltransferase) , P35754 and (Protein name, UniProt ID): R-spondin-2 (Roof plate-specific spondin-2) (hRspo2), or Q6UXX9.

63. The method of any one of claims 32 through 45, wherein the disease state is breast lobular and ductal carcinoma, uterine corpus endometrial carcinoma, ovarian serous cystadenocarcinoma, cervical squamous cell carcinoma and adenocarcinoma, or uterine carcinosarcoma, wherein the one or more markers in a set comprise one or more mutations, insertions, deletions, copy number changes, or expression changes in a gene selected from the group consisting of PIK3CA (phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha), and TTN (titin).

64. The method of any one of claims 32 through 45, wherein the disease state is lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated cytosine residues of a CpG sequence, or the complement of one or more methylated cytosine residues of a CpG sequence selected from the list in Figure 49 or in Figure 64.

65. The method of any one of claims 32 through 45, wherein the disease state is lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated

residues of one or more chromosomal sub-regions selected from the list in Figure 50 or in Figure 65.

66. The method of any one of claims 32 through 45, wherein the disease state is lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma, wherein the one or more markers in a set comprise one or more miRNA sequences selected from (mir ID , Gene ID): hsa-mir-28 , MIR28.

67. The method of any one of claims 32 through 45, wherein the disease state is lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma, wherein the one or more markers in a set comprise one or more Exon RNA sequences selected from the group consisting of (Exon location, Gene): chr2: chr1:93307721-93309752:- , FAM69A; chr1:93312740-93312916:- , FAM69A; chr1:93316405-93316512:- , FAM69A; chr1:93341853-93342152:- , FAM69A; chr1:93426933-93427079:- , FAM69A; chr7:40221554-40221627:+ , C7orf10; chr7:40234539-40234659:+ , C7orf10; chr8:22265823-22266009:+ , SLC39A14; chr8:22272293-22272415:+ , SLC39A14; chr14:39509936-39510091:- , SEC23A; and chr14:39511990-39512076:- , SEC23A.

68. The method of any one of claims 32 through 45, wherein the disease state is lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma, wherein the one or more markers in a set comprise one or more mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product selected from the group consisting of (Gene Symbol , Chromosome Band , Gene Title , UniProt ID): STRN3 , 14q13-q21 , striatin, calmodulin binding protein 3 , Q13033; LRRC17 , 7q22.1 , leucine rich repeat containing 17 , Q8N6Y2; FAM69A , 1p22 , family with sequence similarity 69, member A , Q5T7M9; ATF2 , 2q32 , activating transcription factor 2 , P15336; BHMT , 5q14.1 , betaine--homocysteine S-methyltransferase , Q93088; ODZ3/TENM3 , 4q34.3-q35.1 , teneurin transmembrane protein 3 , Q9P273; ZFHX4 , 8q21.11 , zinc finger homeobox 4 , Q86UP3 or (Protein name , UniProt ID): Leucine-rich repeat-containing protein 17 (p37NB) , Q8N6Y2.

69. The method of any one of claims 32 through 45, wherein the disease state is lung adenocarcinoma, lung squamous cell carcinoma, or head & neck squamous cell carcinoma, wherein the one or more markers in a set comprise one or more mutations, insertions, deletions, copy number changes, or expression changes in a gene selected from the group

consisting of CSMD3 (CUB and Sushi multiple domains 3), DNAH5 (dynein axonemal heavy chain 5), FAT1 (FAT atypical cadherin 1), FLG (filaggrin), KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), LRP1B (LDL receptor related protein 1B), MUC16 (mucin 16, cell surface associated), PCLO (piccolo presynaptic cytomatrix protein), PKHD1L1 (PKHD1 like 1), RELN (reelin), RYR2 (ryanodine receptor 2), SI (sucrase-isomaltase), SYNE1 (spectrin repeat containing nuclear envelope protein 1), TP53 (tumor protein p53), TTN (titin), USH2A (usherin), and XIRP2 (xin actin binding repeat containing 2).

70. The method of any one of claims 32 through 45, wherein the disease state is prostate adenocarcinoma or invasive urothelial bladder cancer, wherein at least 50% of the markers in a set each comprise one or more methylated cytosine residues of a CpG sequence, or the complement of one or more methylated cytosine residues of a CpG sequence selected from the list in Figure 51 or in Figure 66.

71. The method of any one of claims 32 through 45, wherein the disease state is prostate adenocarcinoma or invasive urothelial bladder cancer, wherein at least 50% of the markers in a set each comprise one or more methylated residues of one or more chromosomal sub-regions selected from the list in Figure 52 or in Figure 67.

72. The method of any one of claims 32 through 45, wherein the disease state is prostate adenocarcinoma or invasive urothelial bladder cancer, wherein the one or more markers in a set comprise one or more miRNA sequences selected from the list in Figure 74 or one or more lncRNA or ncRNA sequences selected from the group consisting of (mir ID, Gene ID): hsa-mir-491, MIR491; hsa-mir-1468, MIR1468 and one or more lncRNA or ncRNA sequences selected from the group consisting of [Gene ID, Coordinate (GRCh38), ENSEMBL ID]: AC007383.3, chr2:206084605-206086564, ENSG00000227946.1; and LINC00324, chr17:8220642-8224043, ENSG00000178977.3.

73. The method of any one of claims 32 through 45, wherein the disease state is prostate adenocarcinoma or invasive urothelial bladder cancer, wherein the one or more markers in a set comprise one or more Exon RNA sequences (Exon location, Gene) selected from: chr21:45555942-45556055:+, C21orf33.

74. The method of any one of claims 32 through 45, wherein the disease state is prostate adenocarcinoma or invasive urothelial bladder cancer, wherein the one or more

markers in a set comprise one or more mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product selected from the group:(Gene Symbol , Chromosome Band , Gene Title , UniProt ID): PMM1 , 22q13 , phosphomannomutase 1 , Q92871.

75. The method of any one of claims 32 through 45, wherein the disease state is prostate adenocarcinoma or invasive urothelial bladder cancer, wherein the one or more markers in a set comprise one or more mutations, insertions, deletions, copy number changes, or expression changes in a gene selected from the group consisting of BAGE2 (BAGE family member 2), DNM1P47 (dynamin 1 pseudogene 47), FRG1BP (region gene 1 family member B, pseudogene), KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), RP11-156P1.3, TTN (titin), and TUBB8P7 (tubulin beta 8 class VIII pseudogene 7).

76. The method of any one of claims 32 through 45, wherein the disease state is liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated cytosine residues of a CpG sequence, or the complement of one or more methylated cytosine residues of a CpG sequence selected from the list in Figure 56 or in Figure 68.

77. The method of any one of claims 32 through 45, wherein the disease state is liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein at least 50% of the markers in a set each comprise one or more methylated residues of one or more chromosomal sub-regions selected from the list in Figure 57 or in Figure 69.

78. The method of any one of claims 32 through 45, wherein the disease state is liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more miRNA sequences selected from (mir ID , Gene ID): hsa-mir-132 , MIR132 and one or more lncRNA or ncRNA sequences selected from the list in Figure 53.

79. The method of any one of claims 32 through 45, wherein the disease state is liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more Exon RNA sequences selected from the list in Figure 54.

80. The method of any one of claims 32 through 45, wherein the disease state is liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more mRNA sequences, protein expression levels, protein product concentrations, cytokines, or autoantibody to the protein product selected from the list in Figure 55 or from the group consisting of (Protein name , UniProt ID); Gelsolin (AGEL) (Actin-depolymerizing factor) (ADF) (Brevin) , P06396; Pro-neuregulin-2 , O14511; CD59 glycoprotein (1F5 antigen) (20 kDa homologous restriction factor) (HRF-20) (HRF20) (MAC-inhibitory protein) (MAC-IP) (MEM43 antigen) (Membrane attack complex inhibition factor) (MACIF) (Membrane inhibitor of reactive lysis) (MIRL) (Protectin) (CD antigen CD59) , P13987; and Divergent protein kinase domain 2B (Deleted in autism-related protein 1) , Q9H7Y0.

81. The method of any one of claims 32 through 45, wherein the disease state is liver hepatocellular carcinoma, pancreatic ductal adenocarcinoma, or gallbladder adenocarcinoma, wherein the one or more markers in a set comprise one or more mutations, insertions, deletions, copy number changes, or expression changes in a gene selected from the group consisting of KRAS (Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), MUC16 (mucin 16, cell surface associated), MUC4 (mucin 4, cell surface associated), TP53 (tumor protein p53), and TTN (titin).