

TISSUE-SPECIFIC METHYLATION MARKER

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. provisional application no. 62/643,649, filed March 15, 2018, and U.S. provisional application no. 62/769,928, filed November 20, 2018, each of which is incorporated herein by reference in its entirety.

BACKGROUND

[0002] Quantitative measurement of DNA from different tissues to circulating DNA can potentially provide important information regarding the presence of many different pathological conditions. However, existing methods involving genome-wide bisulfite sequencing are relatively expensive and may present challenges for analysis. More cost-effective approaches for measuring DNA derived from different tissues would be useful.

[0003] Detection of circulating cell-free DNA derived from cancer cells, often known as liquid biopsy, is increasingly used for the management of cancer patients. For example, the detection of epidermal growth factor receptor (EGFR) mutations in plasma can correlate well with the mutational status in the tumor tissues and can predict the responsiveness to EGFR tyrosine kinase inhibitors. In addition to point mutations, other cancer-associated genetic and genomic alterations, including copy number changes and altered fragmentation patterns could also be detected in the cell-free plasma of cancer patients. Patients identified by screening of plasma DNA can potentially have significantly earlier stage distribution and superior progression-free survival compared with patients who do not undergo screening.

SUMMARY

[0004] One aspect of the present disclosure provides a method of determining whether an organism having cancer of a first tissue has a cancer located at a second tissue, the method comprising: (a) obtaining cell-free DNA molecules from a first biological sample of the organism having cancer of a first tissue; (b) performing an assay on the cell-free DNA molecules to determine a first methylation status of a target sequence in the cell-free DNA molecules, wherein the first methylation status of the target sequence indicates that a cell-free DNA molecule comprising the target sequence is from a second tissue of the organism, wherein the first tissue and the second tissue are different; (c) determining an absolute amount of cell-free DNA molecules from the first biological sample comprising the target sequence with the first methylation status; and (d) determining whether the organism has cancer at the second tissue based on the absolute amount.

[0005] In some cases, the methylation status comprises a methylation level. In some cases, the assay comprises isolating the cell-free DNA molecule comprising the target sequence from the first biological sample. In some cases, the assay comprises isolating the cell-free DNA molecule comprising the target sequence in an oil emulsion. In some cases, the assay comprises hybridizing the cell-free DNA molecule comprising the target sequence to a probe. In some cases, the probe hybridizes to the target sequence. In some cases, an affinity of hybridization of the probe to the target sequence depends on the first methylation status of the target sequence in the first biological sample. In some cases, the probe hybridizes to the target sequence when a methylation site of the target sequence is methylated in the first biological sample. In some cases, the probe hybridizes to the target sequence when a methylation site of the target sequence is unmethylated in the first biological sample. In some cases, the assay comprises detecting the hybridization of the probe to the target sequence.

[0006] In some cases, the assay comprises amplifying the cell-free DNA molecule. In some cases, the amplifying comprises use of a pair of primers. In some cases, an affinity of hybridization of at least one primer of the pair of primers to the target sequence depends on the first methylation status of the target sequence. In some cases, the at least one primer of the pair of primers hybridizes to the target sequence when a methylation site of the target sequence is methylated in the first biological sample. In some cases, the at least one primer of the pair of primers hybridizes to the target sequence when a methylation site of the target sequence is unmethylated in the first biological sample. In some cases, the assay comprises bisulfite conversion of unmethylated cytosine residues in the cell-free DNA molecule to uracil. In some cases, the assay comprises performing methylation-aware sequencing of cell-free DNA molecules from the first biological sample.

[0007] In some cases, the target sequence comprises at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, or at least 10 methylation sites. In some cases, the target sequence comprises at least 5 methylation sites. In some cases, the first methylation status comprises methylation density for individual sites within the target sequence, a distribution of methylated/unmethylated sites over a contiguous region within the target sequence, a pattern or level of methylation for each individual methylation site within the target sequence, or non-CpG methylation. In some cases, the target sequence comprises a higher methylation density in the first tissue than in the second tissue. In some cases, the first methylation status comprises methylation density for individual sites with the target sequence, and the target sequence has a methylation density in the first tissue that is at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%. In some cases, the target sequence comprises a methylation density in the first tissue that is more than 50%. In some cases, the target sequence comprises a

methylation density in the second tissue that is at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0%. In some cases, the target sequence comprises a methylation density in the second tissue that is less than 20%. In some cases, the target sequence comprises a lower methylation density in the first tissue than in the second tissue. In some cases, the target sequence comprises a methylation density in the first tissue that is at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0%. In some cases, the target sequence comprises a methylation density in the first tissue that is less than 50%. In some cases, the target sequence comprises a methylation density in the second tissue that is at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100%. In some cases, the target sequence comprises a methylation density in the second tissue that is more than 80%.

[0008] In some cases, the first tissue comprises liver tissue, and the target sequence comprises a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 1. In some cases, the first tissue comprises liver tissue, and wherein the assay comprises amplification using a primer comprising SEQ ID NO: 2, a primer comprising SEQ ID NO: 3, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 4 for detection of the target sequence. In some cases, the amplification further comprises use of a primer comprising SEQ ID NO: 5, a primer comprising SEQ ID NO: 6, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 7 for detection of the target sequence.

[0009] In some cases, the first tissue comprises colon tissue, and the target sequence comprises a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 8. In some cases, the first tissue comprises colon tissue, and wherein the amplification comprises use of a primer comprising SEQ ID NO: 9, a primer comprising SEQ ID NO: 10, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 11 for detection of the target sequence. In some cases, the methylation-specific amplification further comprises use of a primer comprising SEQ ID NO: 12, a primer comprising SEQ ID NO: 13, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 14 for detection of the target sequence.

[0010] In some cases, the cancer is selected from the group consisting of: bladder cancer, bone cancer, brain tumor, breast cancer, carcinoma of cervix, colorectal cancer, esophageal cancer, gastrointestinal cancer, hematopoietic malignancy, head and neck squamous cell carcinoma, leukemia, liver cancer, lung cancer, lymphoma, myeloma, nasal cancer, nasopharyngeal cancer, oral cancer, oropharyngeal cancer, ovarian cancer, prostate cancer, sarcoma, stomach cancer, melanoma, and thyroid cancer. In some cases, the cancer comprises hepatocellular carcinoma or colorectal carcinoma.

[0011] In some cases, the method further comprises determining a classification of the cancer in the second tissue. In some cases, the determining the classification of cancer in the second

tissue comprises assessing cell-free nucleic acid molecules from a second biological sample from the organism. In some cases, the assessing comprises determining methylation profile, copy number variations, single polymorphism (SNP) profile, or fragmentation pattern of the cell-free nucleic acid molecules from the second biological sample. In some cases, the assessing comprises determining an amount of cell-free nucleic acid molecules from the second biological sample that are from a pathogen. In some cases, the second biological sample is same as the first biological sample. In some cases, wherein the second biological sample is different than the first biological sample.

[0012] One aspect of the present disclosure provides a system configured to perform the method as provided herein.

[0013] One aspect of the present disclosure provides a non-transitory computer-readable medium comprising a series of instructions for controlling a computer system to perform the method as disclosed herein.

[0014] One aspect of the present disclosure provides a method of analyzing a biological sample of an organism. The method can comprise: amplifying a first tissue-specific marker in a cell-free DNA molecule from the biological sample based on methylation status of the first tissue-specific marker, wherein the first tissue-specific marker comprises a predetermined sequence with one or more differentiated methylation sites having a first methylation status in a first tissue of the organism and a second methylation status in other tissues of the organism, and wherein the first and second methylation statuses are different; identifying a tissue of origin for the cell-free DNA molecule by detecting the amplification of the first tissue-specific marker; and determining an absolute amount of cell-free DNA molecules from the biological sample that are from the first tissue of the organism.

[0015] In some cases, the method further comprises bisulfite conversion of unmethylated cytosine residues to uracil before the amplification. In some cases, the method further comprises isolating the cell-free DNA molecule from other DNA molecules from the biological sample before the amplification. In some cases, the amplification comprises use of a methylation-specific primer complementary to the first tissue-specific marker and annealing to at least a portion of the one or more differentiated methylation sites. In some cases, the identifying the tissue of origin comprises determining that the cell-free DNA molecule is from the first tissue if the first tissue-specific marker in the cell-free DNA molecule is amplified by primers that are configured to amplify first tissue-specific marker methylated at the first methylation status.

[0016] In some cases, the one or more differentiated methylation sites have a higher methylation density in the first tissue than in the second tissue. In some cases, the one or more differentiated methylation sites have a methylation density in the first tissue that is at least 50%,

at least 60%, at least 70%, at least 80%, at least 90%, or at least 95%. In some cases, the one or more differentiated methylation sites have a methylation density in the first tissue that is more than 50%. In some cases, the one or more differentiated methylation sites have a methylation density in the second tissue that is at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0%. In some cases, the one or more differentiated methylation sites have a methylation density in the second tissue that is less than 20%.

[0017] In some cases, the one or more differentiated methylation sites have a lower methylation density in the first tissue than in the second tissue. In some cases, the one or more differentiated methylation sites have a methylation density in the first tissue that is at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0%. In some cases, the one or more differentiated methylation sites have a methylation density in the first tissue that is less than 50%. In some cases, the one or more differentiated methylation sites have a methylation density in the second tissue that is at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100%. In some cases, the one or more differentiated methylation sites have a methylation density in the second tissue that is more than 80%.

[0018] In some cases, the first tissue comprises liver tissue, and the first tissue-specific marker comprises a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 1. In some cases, the first tissue comprises liver tissue, and wherein the amplification comprises use of a primer comprising SEQ ID NO: 2, a primer comprising SEQ ID NO: 3, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 4 for detection of the first tissue-specific marker. In some cases, the amplification further comprises use of a primer comprising SEQ ID NO: 5, a primer comprising SEQ ID NO: 6, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 7 for detection of the first tissue-specific marker. In some cases, the first tissue comprises colon tissue, and the first tissue-specific marker comprises a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 8. In some cases, the first tissue comprises colon tissue, and wherein the amplification comprises use of a primer comprising SEQ ID NO: 9, a primer comprising SEQ ID NO: 10, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 11 for detection of the first tissue-specific marker. In some cases, the methylation-specific amplification further comprises use of a primer comprising SEQ ID NO: 12, a primer comprising SEQ ID NO: 13, or both, or use of a detectably-labeled probe comprising SEQ ID NO: 14 for detection of the first tissue-specific marker.

[0019] In some cases, the method further comprises determining an amount of cell-free DNA molecules from the biological sample that are derived from a second tissue based on methylation

pattern of a second tissue-specific marker, wherein the first and second tissues are different. In some cases, the second tissue belongs to the organism.

[0020] In some cases, the method further comprises diagnosing, monitoring, or prognosticating a cancer in the first tissue based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue of the organism. In some cases, the cancer is selected from the group consisting of: bladder cancer, bone cancer, brain tumor, breast cancer, carcinoma of cervix, colorectal cancer, esophageal cancer, gastrointestinal cancer, hematopoietic malignancy, head and neck squamous cell carcinoma, leukemia, liver cancer, lung cancer, lymphoma, myeloma, nasal cancer, nasopharyngeal cancer, oral cancer, oropharyngeal cancer, ovarian cancer, prostate cancer, sarcoma, stomach cancer, melanoma, and thyroid cancer. In some cases, the cancer comprises hepatocellular carcinoma or colorectal carcinoma. In some cases, the diagnosing or monitoring comprises determining the size of a tumor in the first tissue based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue of the organism. In some cases, the diagnosing or monitoring comprises determining whether the cancer has metastasized to the first tissue based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue of the organism.

[0021] In some cases, the first tissue comprises a transplanted organ. In some cases, the method provided herein further comprises assessing organ transplant based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue.

[0022] Another aspect of the present disclosure provides a composition for determining amount of cell-free DNA molecules from a biological sample that are from liver of an organism. The composition can comprise a pair of primers for amplification of a liver-specific marker based on methylation status of the liver-specific marker, wherein the liver-specific marker comprises a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 1. In some cases, the pair of primers comprises a primer comprising SEQ ID NO: 2, and a primer comprising SEQ ID NO: 3. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 4 for detection of the liver-specific marker. In some cases, the composition further comprises a primer comprising SEQ ID NO: 5 and a primer comprising SEQ ID NO: 6. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 7 for detection of the liver-specific marker.

[0023] Another aspect of the present disclosure provides a composition for determining amount of cell-free DNA molecules from a biological sample that are from colon of an organism, comprising a pair of primers for amplification of a colon-specific marker based on methylation status of the colon-specific marker, wherein the colon-specific marker comprises a

polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 8. In some cases, the pair of primers comprises a primer comprising SEQ ID NO: 9, and a primer comprising SEQ ID NO: 10. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 11 for detection of the colon-specific marker. In some cases, the composition further comprises a primer comprising SEQ ID NO: 12 and a primer comprising SEQ ID NO: 13. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 14 for detection of the colon-specific marker.

INCORPORATION BY REFERENCE

[0024] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0025] The novel features of the disclosure are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present disclosure will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the disclosure are utilized, and the accompanying drawings of which:

[0026] **Fig. 1** is a schematic drawing of plasma cell-free DNA origins in patients with colorectal cancer with liver metastasis.

[0027] **Fig. 2** shows correlation between fractional concentrations of liver-derived DNA in the plasma of liver transplant recipients based on liver-specific methylation marker analysis (droplet digital PCR) and donor-specific alleles analysis by sequencing.

[0028] **Figs. 3A** and **3B** show absolute concentrations (**Fig. 3A**) and fractional concentrations (**Fig. 3B**) of liver-derived DNA using droplet digital PCR in the plasma of healthy subjects, chronic HBV carriers, cirrhotic patients and HCC patients.

[0029] **Figs. 4A** and **4B** show correlation between the maximal dimension of the tumor and the absolute concentration (**Fig. 4A**) and the fractional concentration (**Fig. 4B**) of liver-derived DNA using droplet digital PCR in the plasma of HCC patients.

[0030] **Figs. 5A-5D** show the plasma concentrations of colon- and liver-derived DNA by droplet digital PCR in healthy subjects and CRC patients with and without liver metastases. (**Fig. 5A**) absolute concentration of colon-derived DNA, (**Fig. 5B**) fractional concentration of colon-derived DNA, (**Fig. 5C**) absolute concentration of liver-derived DNA, and (**Fig. 5D**) fractional concentration of liver-derived DNA.

[0031] **Fig. 6** shows ROC curves for using the absolute and fractional concentrations of liver- and colon-derived DNA to differentiate CRC patients with and without liver metastasis. “AUC” denotes area under curve.

[0032] **Fig. 7** shows the methylation densities of the CpG sites within the protein tyrosine kinase 2 beta (PTK2B) gene region.

[0033] **Fig. 8** shows the methylation densities of the CpG sites within the Sestrin 3 (*SESN3*) gene region.

[0034] **Fig. 9** shows a ROC curve for using the absolute and fractional concentrations of liver-derived DNA to differentiate HCC patients from non-HCC subjects.

[0035] **Fig. 10** shows a computer control system that can be programmed or otherwise configured to implement methods provided herein.

[0036] **Fig. 11** shows a diagram of the methods and systems as disclosed herein.

DETAILED DESCRIPTION

[0037] Provided herein are methods, compositions, and systems for quantifying cell-free nucleic acid molecules, e.g., cell-free DNA molecules, e.g., plasma DNA, that are from specific tissues using tissue-specific markers. Also provided herein are clinical applications of these markers, for instance, but not limited to, in diagnosis, monitoring, and prognostication of cancer, in detection of metastatic cancer, and in some case, in assessment of organ transplantation.

[0038] In addition to cancer-specific changes, there can be a general increase in DNA released from an organ affected by a cancer into the circulation. For instance, the level of liver-derived DNA can be increased in liver cancer patients. Without wishing to be bound by a certain theory, the increased level of DNA released by the organ affected by the cancer can be due to the direct release of DNA from the tumor cells or the increased turnover of non-tumor cells invaded by the cancer. Such an increase in release of tissue specific DNA can also be observed in organ transplant recipients who experienced acute rejection as a result of increased cell turnover. In some cases, the proportional contribution of the cells within the transplanted organ into plasma DNA determined by methylation deconvolution correlates well with that determined based on the analysis of donor-specific alleles.

[0039] In some cases, methylation deconvolution using genome-wide bisulfite sequencing can be challenging because of the relatively high cost and long turn-around time. Moreover, in some cases, methylation deconvolution can determine the relative or proportional contributions of different organs rather than the absolute concentration of DNA derived from each organ. In scenarios where DNA from more than one organ would be released into the circulation, the measurement of absolute concentration of DNAs derived from a specific organ can be more

informative. For example, in a patient with colorectal cancer (CRC) metastasizing to the liver, an increased amount of DNA would be released from the liver into the circulation. However, the fraction of liver DNA in plasma might show a paradoxical reduction because of the even larger degree of increase in DNA released by the tumor cells which have originated from the colon. Therefore, the development of a method that can accurately determine the absolute amount of DNA with a tissue-specific methylation pattern can be useful.

[0040] I. Tissue-specific Marker

[0041] Provided herein are tissue-specific markers that can identify a tissue of origin for a cell-free DNA molecule. In some cases, the tissue-specific marker is a polynucleotide sequence of the genome of an organism. In some cases, the tissue-specific marker comprises a differentiated methylated region (DMR) which is identified based on the methylation status of one or more differentiated methylation sites contained within the marker polynucleotide sequence. In some cases, the one or more differentiated methylation sites comprise one or more CpG sites. In some cases, the one or more differentiated methylation sites comprise one or more non-CpG sites. In some cases, a tissue-specific marker as discussed herein can be referred to as a target sequence.

[0042] In some cases, the differentiated methylation sites of the tissue-specific marker have a first methylation status in a first tissue of the organism, whereas a second methylation status in a different second tissue of the organism. The first and second methylation statuses can be different so that the first and second tissues can be differentiated based on the methylation status of the tissue-specific marker.

[0043] In some cases, the differentiated methylation sites of the tissue-specific marker have a first methylation status in a first tissue of the organism, whereas a second methylation status in all other tissues of the organism. The first and second methylation statuses can be different so that the first tissue can be differentiated from all other tissues of the organism based on the methylation status of the tissue-specific marker.

[0044] In some cases, the tissue-specific marker comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 differentiated methylation sites. In some cases, the tissue-specific marker comprises at least 5 differentiated methylation sites. A methylated nucleotide or a methylated nucleotide base can refer to the presence of a methyl moiety on a nucleotide base, where the methyl moiety is not present in a recognized typical nucleotide base. For example, cytosine does not contain a methyl moiety on its pyrimidine ring, but 5-methylcytosine contains a methyl moiety at position 5 of its pyrimidine ring. In this case, cytosine is not a methylated nucleotide and 5-methylcytosine is a methylated nucleotide. In another example, thymine contains a methyl moiety at position 5 of its pyrimidine ring, however,

for purposes herein, thymine is not considered a methylated nucleotide when present in DNA since thymine is a typical nucleotide base of DNA. Typical nucleoside bases for DNA are thymine, adenine, cytosine and guanine. Typical bases for RNA are uracil, adenine, cytosine and guanine. Correspondingly a "methylation site" can be the location in the target gene nucleic acid region where methylation has, or has the possibility of occurring. For example a location containing CpG is a methylation site wherein the cytosine may or may not be methylated. A "site" can correspond to a single site, which can be a single base position or a group of correlated base positions, e.g., a CpG site. A methylation site can refer to a CpG site, or a non-CpG site of a DNA molecule that has the potential to be methylated. A CpG site can be a region of a DNA molecule where a cytosine nucleotide is followed by a guanine nucleotide in the linear sequence of bases along its 5' to 3' direction and that is susceptible to methylation either by natural occurring events in vivo or by an event instituted to chemically methylate the nucleotide in vitro. A non-CpG site can be a region that does not have a CpG dinucleotide sequence but is also is susceptible to methylation either by natural occurring events in vivo or by an event instituted to chemically methylate the nucleotide in vitro. A locus or region can correspond to a region that includes multiple sites.

[0045] The methylation status of the tissue-specific marker can comprise methylation density for individual sites within the marker region, a distribution of methylated/unmethylated sites over a contiguous region within the marker, a pattern or level of methylation for each individual methylation site within the marker that contains more than one sites, and non-CpG methylation. In some cases, the methylation status of the tissue-specific marker comprises methylation level (or methylation density) for individual differentiated methylation sites. The methylation density can refer to, for a given methylation site, a fraction of nucleic acid molecules methylated at the given methylation site over the total number of nucleic acid molecules of interest that contain such methylation site. For instance, the methylation density of a first methylation site in liver tissue can refer to a fraction of liver DNA molecules methylated at the first site over the total liver DNA molecules. In some cases, the methylation status comprises coherence of methylation/unmethylation status among individual differentiated methylation sites.

[0046] In some cases, the tissue-specific marker comprises methylation sites that are hypermethylated in a first tissue, but are hypomethylated in a second tissue. For instance, the tissue-specific marker can comprise one or more methylation sites that are hypermethylated in liver tissue, by which it can mean the one or more methylation sites have an at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% methylation density in liver tissue; in contrast, the one or more methylation sites can have an at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0% methylation density in other tissues, such as,

but not limited to, blood cells, lung, esophagus, stomach, small intestines, colon, pancreas, urinary bladder, heart, and brain. The tissue-specific marker can comprise at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 methylation sites that are hypermethylated in a first tissue, but hypomethylated in a second tissue. In some cases, the tissue-specific marker comprises at least 5 methylation sites that are hypermethylated in a first tissue, but hypomethylated in a second tissue.

[0047] The tissue-specific marker can comprise at most 300 base-pairs (bp), at most 250 bp, at most 225 bp, at most 200 bp, at most 190 bp, at most 185 bp, at most 180 bp, at most 175 bp, at most 170 bp, at most 169 bp, at most 168 bp, at most 167 bp, at most 166 bp, at most 165 bp, at most 164 bp, at most 163 bp, at most 162 bp, at most 161 bp, at most 160 bp, at most 150 bp, at most 140 bp, at most 120 bp, or at most 100 bp. In some cases, the tissue-specific marker comprises at most 166 bp.

[0048] In some cases, the liver-specific marker as provided herein comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 methylation sites that are hypermethylated in liver tissue, but hypomethylated in other tissues. In some cases, the liver-specific marker as provided herein comprises at least 5 methylation sites that are hypermethylated in liver tissue, but hypomethylated in other tissues. Each of the methylation sites that are hypermethylated in liver tissue can have an at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% methylation density in liver tissue, and an at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0% methylation density in other tissues, such as, but not limited to, blood, brain, thymus, pancreas, kidney. In some cases, each of the methylation sites that are hypermethylated in liver tissue can have a more than 50% methylation density in liver tissue, and a less than 20% methylation density in other tissues, such as, but not limited to, blood cells, lung, esophagus, stomach, small intestines, colon, pancreas, urinary bladder, heart, and brain.

[0049] In some cases, the tissue-specific marker comprises methylation sites that are hypomethylated in a first tissue, but are hypermethylated in a second tissue. In some cases, the tissue-specific marker comprises methylation sites that are hypomethylated in a first tissue, but are hypermethylated in other tissues. For instance, the tissue-specific marker can comprise one or more methylation sites that are hypomethylated in liver tissue, by which it can mean the one or more methylation sites have an at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0% methylation density in liver tissue; in contrast, the one or more methylation sites can have an at least 50%, at least 60%, at least 70%, at least 80%, at least 90%,

or at least 95% methylation density in other tissues, such as, but not limited to, blood cells, lung, esophagus, stomach, small intestines, colon, pancreas, urinary bladder, heart, and brain. The tissue-specific marker can comprise at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 methylation sites that are hypomethylated in a first tissue, but hypermethylated in a second tissue. In some cases, the tissue-specific marker comprises at least 5 methylation sites that are hypomethylated in a first tissue, but hypermethylated in a second tissue.

[0050] In some cases, the liver-specific marker as provided herein comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 methylation sites that are hypomethylated in liver tissue, but hypermethylated in other tissues. In some cases, the liver-specific marker as provided herein comprises at least 5 methylation sites that are hypomethylated in liver tissue, but hypermethylated in other tissues. Each of the methylation sites that are hypomethylated in liver tissue can have an at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, or at most 5% methylation density in liver tissue, and an at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% methylation density in other tissues, such as, but not limited to, blood, brain, thymus, pancreas, kidney. In some cases, each of the methylation sites that are hypomethylated in liver tissue can have a less than 50% methylation density in liver tissue, and a more than 80% methylation density in other tissues, such as, but not limited to, blood cells, lung, esophagus, stomach, small intestines, colon, pancreas, urinary bladder, heart, and brain.

[0051] In some cases, the colon-specific marker as provided herein comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 methylation sites that are hypermethylated in colon tissue, but hypomethylated in other tissues. In some cases, the colon-specific marker as provided herein comprises at least 5 methylation sites that are hypermethylated in colon tissue, but hypomethylated in other tissues. Each of the methylation sites that are hypermethylated in colon tissue can have an at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% methylation density in colon tissue, and an at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, at most 5%, or 0% methylation density in other tissues, such as, but not limited to, blood, brain, thymus, pancreas, kidney. In some cases, each of the methylation sites that are hypermethylated in colon tissue can have a more than 50% methylation density in colon tissue, and a less than 20% methylation density in other tissues, such as, but not limited to, blood cells, lung, esophagus, stomach, small intestines, liver, pancreas, urinary bladder, heart, and brain.

[0052] In some cases, the colon-specific marker as provided herein comprises at least 1, at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least 12, at least 15, at least 20, at least 25, at least 30, at least 50 methylation sites that are hypomethylated in colon tissue, but hypermethylated in other tissues. In some cases, the colon-specific marker as provided herein comprises at least 5 methylation sites that are hypomethylated in colon tissue, but hypermethylated in other tissues. Each of the methylation sites that are hypomethylated in colon tissue can have an at most 50%, at most 40%, at most 30%, at most 20%, at most 10%, or at most 5% methylation density in colon tissue, and an at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or 100% methylation density in other tissues, such as, but not limited to, blood, brain, thymus, pancreas, kidney. In some cases, each of the methylation sites that are hypomethylated in colon tissue can have a less than 50% methylation density in colon tissue, and a more than 80% methylation density in other tissues, such as, but not limited to, blood cells, lung, esophagus, stomach, small intestines, liver, pancreas, urinary bladder, heart, and brain.

[0053] Also provided herein is a liver-specific marker for identifying liver-derived DNA molecules. A liver-specific marker can be located in the exonic region of the protein tyrosine kinase 2 beta (PTK2B) gene on chromosome 8. The eight CpG sites within the liver-specific DMR can be hypermethylated in the liver but hypomethylated in other tissues and blood cells. A liver-specific marker as provided herein can comprise a polynucleotide sequence having at least about 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 1. A liver-specific marker as provided herein can comprise SEQ ID NO: 1.

[0054] Also provided herein is a colon-specific marker for identifying colon-derived DNA molecules. A colon-specific marker can be located in the exonic region of the Sestrin 3 (SESN3) gene on chromosome 11. All the six CpG sites located within the colon-specific DMR can be hypermethylated in the colon but hypomethylated in other tissues and blood cells. A colon-specific marker can comprise a polynucleotide sequence having at least about 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 8.

[0055] As used herein, the term “identity” or “percent identity” between two or more nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence). The nucleotides at corresponding positions can then be compared, and the percent identity between the two sequences can be a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). For example, a position in the first sequence may be occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between

the two sequences may be a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences. In some cases, the length of a sequence aligned for comparison purposes is at least about: 30%, 40%, 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 95%, of the length of the reference sequence. A BLAST® search can determine homology between two sequences. The homology can be between the entire lengths of two sequences or between fractions of the entire lengths of two sequences. The two sequences can be genes, nucleotides sequences, protein sequences, peptide sequences, amino acid sequences, or fragments thereof. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A non-limiting example of such a mathematical algorithm can be those described in Karlin, S. and Altschul, S., Proc. Natl. Acad. Sci. USA, 90- 5873-5877 (1993). Such an algorithm can be incorporated into the NBLAST and XBLAST programs (version 2.0), as described in Altschul, S. et al., Nucleic Acids Res., 25:3389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, any relevant parameters of the respective programs (*e.g.*, NBLAST) can be used. For example, parameters for sequence comparison can be set at score= 100, word length= 12, or can be varied (*e.g.*, W=5 or W=20). Other examples include the algorithm of Myers and Miller, CABIOS (1989), ADVANCE, ADAM, BLAT, and FASTA.

[0056] Also provided herein are methods of identifying tissue-specific markers. The method can comprise comparing methylation status across the genome among different tissue samples. Publicly available databases, such as, databases from RoadMap Epigenomics Project (Roadmap Epigenomics Consortium et al. Nature 2015;518:317-30) and BLUEPRINT project (Martens et al. Haematologica 2013;98:1487-9), can be utilized for bioinformatics analysis in order to screen for potential tissue-specific markers. In some cases, experimental validation is desirable. For instance, methylation-aware sequencing, such as bisulfite sequencing, can be performed to validate the methylation status among different tissues. In some cases, methylation-specific amplification can also be used for a relatively more target-orientated validation.

[0057] II. Methods of Analyzing Cell-free DNA Molecules from a Tissue

[0058] Provided herein are methods of analyzing a biological sample of an organism. A method as provided herein can determine an absolute amount of cell-free nucleic acid molecules, *e.g.*, cell-free DNA molecules, from a tissue of an organism. A method can comprise identifying a cell-free DNA molecule from a biological sample as a cell-free DNA molecule from a first tissue of the organism when the cell-free DNA molecule comprises a first tissue-specific marker having a first methylation status. In some cases, the first tissue-specific marker comprises a predetermined sequence with one or more differentiated methylation sites. In some cases, the

first tissue-specific marker can have a first methylation status in a first tissue and a second methylation status in other tissues of the organism, and the first methylation status and the second methylation status can be different.

[0059] As provided herein, a method can comprise assessing methylation status of a cell-free DNA molecule. In some cases, a method can further comprise bisulfite conversion of unmethylated cytosine residues to uracil before the amplification. In some cases, a method can comprise conversion of either methylated or unmethylated cytosine residues by any other method, so that the converted residues can be differentiated by the subsequent detection method, e.g. primer-based amplification. In some cases, the cell-free DNA molecule can be digested with a methylation-sensitive enzyme that digests a DNA molecule at one or more specific methylation sites when the methylation sites are methylated or unmethylated. The methylation-sensitive enzyme can thus be used to differentiate methylated and unmethylated DNA molecules. Non-limiting examples of methylation-sensitive enzymes that can be used in the methods provided herein can include Aat II, Acc II, Aor13H I, Aor51H I, BspT104 I, BssH II, Cfr10 I, Cla I, Cpo I, Eco52 I, Hae II, Hap II, Hha I, Mlu I, Nae I, Not I, Nru I, Nsb I, PmaC I, Psp1406 I, Pvu I, Sac II, Sal I, Sma I, SnaB I, and any combinations thereof.

[0060] A method as provided herein can comprise amplifying a first tissue-specific marker in a cell-free DNA molecule from the biological sample based on the methylation status of the first tissue-specific marker. The method can further comprise identifying a tissue of origin for the cell-free DNA molecule by detecting the amplification of the first tissue-specific marker. A method can further comprise determining an absolute amount of cell-free DNA molecules from the biological sample that are from the first tissue of the organism. Amplification reaction can refer to a process for copying nucleic acid one or more times. In some cases, the method of amplification includes but is not limited to polymerase chain reaction (PCR), self-sustained sequence reaction, ligase chain reaction, rapid amplification of cDNA ends, polymerase chain reaction and ligase chain reaction, Q-beta phage amplification, strand displacement amplification, or splice overlap extension polymerase chain reaction. In some cases, a single molecule of nucleic acid is amplified, for example, by digital PCR.

[0061] In some cases, the amplification comprises use of a methylation-specific primer complementary to the first tissue-specific marker and annealing to at least a portion of the one or more methylation sites. A methylation-specific primer can refer to a primer that can differentiate methylated and unmethylated target sequence. For a given target sequence that contains methylation sites, a methylation-specific primer can be designed to cover at least a portion of the methylation sites. In some cases, when bisulfite conversion is performed before amplification, the nucleotide residue on the methylation-specific primer for a given methylation site can be

designed to be complementary to unconverted cytosine residue at the site for detecting methylated target sequence, whereas for detecting unmethylated target sequence, the nucleotide residue on the methylation-specific primer can be designed to be complementary to converted residue at the site (e.g. uracil residue).

[0062] In some cases, a method can further comprise isolating the cell-free DNA molecule from other DNA molecules from the biological sample before the amplification. By physically isolating the cell-free DNA molecule, a “digital” assay of the cell-free DNA molecules from the biological sample can be performed, such as, but not limited to, digital PCR, e.g. droplet digital PCR. The isolation method can be any method known to a skilled artisan, such as, but not limited to, isolation by micro well plates, capillaries, oil emulsion, and arrays of miniaturized chambers. The digital PCR reaction can be performed using any known techniques in the art, such as, microfluidics-based, or emulsion-based, e.g. BEAMing (Dressman et al. *Proc Natl Acad Sci USA* 2003; 100: 8817-8822).

[0063] In some cases, the first tissue comprises liver tissue, and the method comprises use of a liver-specific marker as discussed above. In some cases, the liver-specific marker, e.g. SEQ ID NO: 1, comprises sites hypermethylated in liver but hypomethylated in other tissues. In these cases, the method can comprise primers for detecting methylated liver tissue-specific marker (“primers for methylated assay”). For instance, the primers for methylated assay can comprise a primer comprising SEQ ID NO: 2, a primer comprising SEQ ID NO: 3, or both. In some cases, the primers provided herein are used for amplification reactions after bisulfite conversion. Alternatively or cumulatively, the method can comprise use of a detectably-labeled probe comprising SEQ ID NO: 4, for detection of the methylated liver tissue-specific marker. Optionally, the method can further comprise use of primers for detecting unmethylated liver tissue-specific marker (“primers for unmethylated assay”). The primers for unmethylated assay can comprise a primer comprising SEQ ID NO: 5, a primer comprising SEQ ID NO: 6, or both. Alternatively or cumulatively, the method can comprise use of a detectably-labeled probe comprising SEQ ID NO: 7, for detection of the unmethylated liver tissue-specific marker.

[0064] A primer, a probe, or an oligonucleotide, can be used interchangeably herein and can refer to a polymer of more than one, e.g., 2, 4, 6, 8, 10, 14, 18, 20, or 40 nucleotides or chemically modified nucleotides that are linked together via phosphodiester bond. A primer can comprise about 20 to about 30 nucleotides. A primer can comprise at least 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, or 40 nucleotides. Chemical modifications to the constitute nucleotides of a primer can be introduced to modify certain properties of the primer, for instance, increasing its stability, increasing hybridization specificity, and labeling with a detectable signal. Chemical modifications that can be used in the compositions provided herein

can include covalent modifications, attachment chemistry, and modification of nucleotide bases. Chemical modifications can include phosphorylation, addition of biotin, cholesteryl-TEG, amino-modifiers (e.g., C6, C12, or dT), azide, alkyne, thiol modifiers, fluorophore, dark quencher, and spacer. A primer can comprise one or more phosphorothioate bonds, or one or more modified bases, such as 2-aminopurine, 2,6-diaminopurine (2-amino-dA), 5-bromo dU, deoxyUridine, inverted dT, inverted dideoxy-T, dideoxy-C, 5-methyl dC, deoxyinoside, super T®, super G®, locked nucleic acids, 5-nitroindole, 2'-O-methyl RNA bases, hydroxymethyl dC, iso-dC, iso-dG, fluoro bases (e.g., fluoro C, U, A, G, T), and 2'-O-methoxy-ethyl bases (e.g., 2-methoxyethoxy A, G, U, C, T). The detectably-labeled probe can comprise any detectable label known to one skilled in the art, for instance, any suitable fluorophore. The PCR reaction (e.g., digital PCR or real-time PCR) can be monitored via any suitable optical methods, magnetic methods, electronic methods, or any other techniques available to one of skill in the art.

[0065] In some cases, the first tissue comprises colon tissue, and the method comprises use of a colon-specific marker as discussed above. In some cases, the colon-specific marker, e.g. SEQ ID NO: 8, comprises sites hypermethylated in colon but hypomethylated in other tissues. In these cases, the primers for methylated assay can comprise a primer comprising SEQ ID NO: 9, a primer comprising SEQ ID NO: 10, or both. In some cases, the primers provided herein are used for amplification reactions after bisulfite conversion. Alternatively or cumulatively, the method can comprise use of a detectably-labeled probe comprising SEQ ID NO: 11, for detection of the methylated colon-specific marker. Optionally, the method can further comprise use of primers for detecting unmethylated colon-specific marker ("primers for unmethylated assay"). The primers for unmethylated assay can comprise a primer comprising SEQ ID NO: 12, a primer comprising SEQ ID NO: 13, or both. Alternatively or cumulatively, the method can comprise use of a detectably-labeled probe comprising SEQ ID NO: 14, for detection of the unmethylated colon-specific marker.

[0066] The methods provided herein, e.g. use of digital PCR technique, can allow the direct determination of the actual number of the target DNA molecules without the need of calibrators. Other technologies, such as certain sequencing-based methods, such as, but not limited to, bisulfite sequencing and non-bisulfite-based methylation-aware sequencing using the PacBio sequencing platform, can determine the relative or fractional concentration of the DNA from the target tissues in relation to other tissues. The absolute amount can refer to an absolute count of DNA molecules, or in some cases, can also refer to a concentration of DNA molecules, e.g., number, mole, or weight per volume, e.g., copies/mL, mole/L, or mg/L. The analysis of the absolute amount as provided herein can be useful in scenarios when increased amounts of DNA would be released from more than one type of tissues. Methylation deconvolution analysis, based

on sequencing of cell-free nucleic acid molecules, such as disclosed in U.S. Patent Application No. 14/803,692, on the other hand, can provide readout of tissue of origin of cell-free nucleic acids in the form of fractional contribution, e.g., a first tissue contributes A% of cell-free nucleic acids from a biological sample, and a second tissue contributes B% of cell-free nucleic acids from the same biological sample.

[0067] In some cases, the methods, compositions, and systems provided herein can also make use of technologies like, real-time PCR, sequencing and microarray for methylation analysis of cell-free nucleic acids. In some cases, the absolute number of cell-free nucleic acids harboring a tissue-specific marker, such as counting positive reactions in a digital PCR assay, may not be derived directly from methylation analysis by some technologies. However, such absolute number can be calculated indirectly based on concentrations (relative or fractional) of cell-free nucleic acids harboring tissue-specific markers, for instance, by taking the total number or concentration of cell-free nucleic acids in a given volume of biological sample into account. In some cases, the sequencing that can be used in the methods provided herein can include chain termination sequencing, hybridization sequencing, Illumina sequencing (e.g., using reversible terminator dyes), ion torrent semiconductor sequencing, mass spectrophotometry sequencing, massively parallel signature sequencing (MPSS), Maxam-Gilbert sequencing, nanopore sequencing, polony sequencing, pyrosequencing, shotgun sequencing, single molecule real time (SMRT) sequencing, SOLiD sequencing (hybridization using four fluorescently labeled di-base probes), universal sequencing, or any combination thereof. Microarrays having probes targeting methylation sites can also be used for analyzing methylation status of the cell-free DNA molecules in the methods provided herein.

[0068] In some cases, the method provided herein further comprises determining an amount of cell-free DNA molecules from the biological sample that are derived from a second tissue based on methylation pattern of a second tissue-specific marker, wherein the first and second tissues are different. The second tissue can belong to the same organism. The second tissue can also come from a different organism, for instance, a fetus within a pregnant woman.

[0069] III. Methods of Diagnosing, Monitoring, Prognosticating Cancer

[0070] Provided herein are methods of diagnosing, monitoring, prognosticating cancer. A method can comprise determining an absolute amount of cell-free DNA molecules from a first tissue as discussed above, and diagnosing, monitoring, prognosticating a cancer in the first tissue based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue of the organism.

[0071] An absolute amount of cell-free DNA molecules that are from a first tissue can be correlated with a condition of the first tissue. For instance, the amount of liver-derived plasma

DNA molecules can be increased due to increased release of DNA molecules from the liver tissue as a result of tumor growth. In other cases, increased cell turnover, for example, as a result of organ transplant can also lead to increased plasma DNA released from the tissue with the transplant.

[0072] The methods provided herein can comprise determining the size of a tumor in the first tissue based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue of the organism. A predetermined comparison chart, in which the amount of target cell-free DNA molecule and the tumor size are correlated, can be used for the tumor size determination. The detection of the tumor size can aid diagnosis, monitoring, and prognostication of the cancer.

[0073] The methods provided herein can comprise determining whether the cancer has metastasized to the first tissue based on the absolute amount of the cell-free DNA molecules from the biological sample that are from the first tissue of the organism. The absolute amount of target DNA molecules determined by the methods provided herein can provide a desirable differentiation between cancer patients with and without metastasis, as compared to fractional amount of the target DNA molecules.

[0074] Cancer types that the methods, compositions, and systems provided herein can be applicable to can comprise bladder cancer, bone cancer, a brain tumor, breast cancer, carcinoma of cervix, colorectal cancer, esophageal cancer, gastrointestinal cancer, hematopoietic malignancy, head and neck squamous cell carcinoma, leukemia, liver cancer, lung cancer, lymphoma, myeloma, nasal cancer, nasopharyngeal cancer, oral cancer, oropharyngeal cancer, ovarian cancer, prostate cancer, sarcoma, stomach cancer, or thyroid cancer. The metastatic tissue to be assessed by the methods provided herein can comprise bladder, bone, brain, breast, cervix, colon, esophagus, gastrointestinal tract, blood, head, neck, liver, lung, lymph node, nose, nasopharynx, mouth, oropharynx, ovary, prostate, skin, stomach, or thyroid.

[0075] As one skilled in the art would readily appreciate, cancer cells can spread locally by moving into nearby normal tissue, can spread regionally, to nearby lymph nodes, tissues, or organs, and can spread to distant parts of the body. The spread of a cancer from an initial first tissue to a second tissue can be called metastasis, and such a cancer can thus be called metastatic cancer. Exemplary types of cancer metastasis that the methods, compositions, and systems provided herein can be applied to can include metastasis taking place in the sites listed in **Table 1**.

Table 1. Exemplary Cancer Metastasis Sites

Cancer Type	Metastasis Sites
Bladder	Bone, liver, lung
Breast	Bone, brain, liver, lung
Colon	Liver, lung, peritoneum
Kidney	Adrenal gland, bone, brain, liver, lung
Lung	Adrenal gland, bone, brain, liver, other lung
Melanoma	Bone, brain, liver, lung, skin, muscle
Ovary	Liver, lung, peritoneum
Pancreas	Liver, lung, peritoneum
Prostate	Adrenal gland, bone, liver, lung
Rectal	Liver, lung, peritoneum
Stomach	Liver, lung, peritoneum
Thyroid	Bone, liver, lung
Uterus	Bone, liver, lung, peritoneum, vagina

[0076] In some cases, the methods, compositions, and systems as provided herein can be applied to diagnose, monitor, and prognosticate cancer when combined with other techniques that are available to one skilled in the art. In some cases, detection of other molecular markers, e.g., in nucleic acid, e.g., DNA, RNA, such as, copy number aberrations (CNAs), single nucleotide polymorphisms (SNPs), genetic mutations, germline mutations, somatic mutations, nucleic acids from a pathogen, e.g., a virus, e.g., an Epstein-Barr virus, size of cell-free nucleic acids, and fragmentation patterns of cell-free nucleic acids, can also be applied in combination with the methods, compositions, and systems as provided herein. The combination of techniques can help facilitate the detection of level of cancer, including, but not limited to, whether cancer exists, a stage of a cancer, a size of tumor, how many deletions or amplifications of a chromosomal region are involved (e.g. duplicated or tripled), and/or other measure of a severity of a cancer. The level of cancer can be a number of other characters. The level can be zero. The level of cancer can also include premalignant or precancerous conditions associated with deletions or amplifications.

[0077] In some cases, detection of copy number aberrations (CNAs), such as the methods disclosed in U.S. Patent No. 8,741,811, can be used in combination with the methods provided herein. As discussed above, in some cases, based on absolute amount of cell-free nucleic acid

from a first tissue, the methods provided herein can determine whether a cancer has metastasized into the first tissue. Detection of CNAs, on the other hand, can aid in identifying the origin of the metastatic cancer cells in the first tissue. In some cases, analysis of fragmentation pattern of cell-free nucleic acids, such as the methods disclosed in U.S. Patent Application No. 15/218,497, can be used in combination with the methods, compositions, and systems provided herein. In some cases, the subject methods, compositions, and systems can be used for detecting, monitoring, or prognosticating cancer in a subject in combination with any available approaches. Besides the aforementioned detection methods, one can also perform any appropriate tests like tumor biomarker tests (e.g., alpha-fetoprotein (AFP) for liver cancer, ALK gene for non-small-cell lung cancer, prostate-specific antigen (PSA) for prostate cancer, and thyroglobulin for thyroid cancer), physical examination, radiographic imaging (e.g., computed tomography scan, magnetic resonance imaging, Positron Emission Tomography (PET)), ultra sound examination, endoscopic examinations, biopsy, or cytology tests.

[0078] In some cases, the subject methods, compositions, or systems can be used for monitoring cancer in a subject on a regular, semi-regular, or non-regular schedule. For instance, a subject can take a cancer monitoring examination that makes use of the subject methods, compositions, or systems on a weekly, monthly, quarterly, or annually basis. In some cases, as subject can take such examinations about every 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or more than 12 months. In some cases, the interval between two consecutive examinations can be determined based on the results of the most recent examination, e.g., in some cases, according to a physician's prescription or medical advice.

[0079] IV. Methods of Assessing Organ Transplant

[0080] Provided herein are methods of assessing organ transplant based on determination of an absolute amount of cell-free DNA molecules that are from the transplanted tissue. The transplanted tissues as described herein are considered as a tissue of the subject of concern.

[0081] The method as provided herein can take advantage of the correlation between the amount of cell-free DNA molecules that are from the transplanted tissue and the cell turnover rate in the transplanted tissue. The cell turnover rate can thereby be used as a criterion for assessing the organ transplant.

[0082] V. Composition for Analyzing Cell-free DNA Molecules from a Tissue

[0083] Also provided herein are compositions for analyzing cell-free DNA molecules from a specific tissue, e.g. bone, liver, lung, brain, peritoneum, adrenal gland, skin, muscle, vagina, colon, bladder, breast, kidney, melanoma, ovary, pancreas, prostate, rectal, stomach, thyroid, or uterus.

[0084] A composition for determining amount of cell-free DNA molecules from a biological sample that are from liver of an organism can comprise a pair of primers for amplification of a liver-specific marker based on methylation status of the liver-specific marker. The liver-specific marker can comprise a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 1. In some cases, the pair of primers comprises a primer comprising SEQ ID NO: 2, and a primer comprising SEQ ID NO: 3. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 4 for detection of the liver-specific marker. In some cases, the composition further comprises a primer comprising SEQ ID NO: 5 and a primer comprising SEQ ID NO: 6. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 7 for detection of the liver-specific marker.

[0085] A composition provided herein can comprise a pair of primers for amplification of a liver-specific marker based on methylation status of the colon-specific marker. The colon-specific marker can comprise a polynucleotide sequence having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identity to SEQ ID NO: 8. In some cases, the pair of primers comprises a primer comprising SEQ ID NO: 9, and a primer comprising SEQ ID NO: 10. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 11 for detection of the colon-specific marker. In some cases, the composition further comprises a primer comprising SEQ ID NO: 12 and a primer comprising SEQ ID NO: 13. In some cases, the composition further comprises a detectably-labeled probe comprising SEQ ID NO: 14 for detection of the colon-specific marker.

[0086] VI. Biological Sample

[0087] The biological sample used in methods provided herein can include any tissue or material derived from a living or dead subject. A biological sample can be a cell-free sample. A biological sample can comprise a nucleic acid (e.g., DNA, e.g., genomic DNA or mitochondrial DNA, or RNA) or a fragment thereof. The nucleic acid in the sample can be a cell-free nucleic acid. A sample can be a liquid sample or a solid sample (e.g., a cell or tissue sample). The biological sample can be a bodily fluid, such as blood, plasma, serum, urine, vaginal fluid, fluid from a hydrocele (e.g., of the testis), vaginal flushing fluids, pleural fluid, ascitic fluid, cerebrospinal fluid, saliva, sweat, tears, sputum, bronchoalveolar lavage fluid, discharge fluid from the nipple, aspiration fluid from different parts of the body (e.g., thyroid, breast), etc. Stool samples can also be used. In various embodiments, the majority of DNA in a biological sample that has been enriched for cell-free DNA (e.g., a plasma sample obtained via a centrifugation protocol) can be cell-free (e.g., greater than 50%, 60%, 70%, 80%, 90%, 95%, or 99% of the DNA can be cell-free). The biological sample can be treated to physically disrupt tissue or cell

structure (e.g., centrifugation and/or cell lysis), thus releasing intracellular components into a solution which can further contain enzymes, buffers, salts, detergents, and the like which are used to prepare the sample for analysis.

[0088] Methods, compositions, and systems provided herein can be used to analyze nucleic acid molecules in a biological sample. The nucleic acid molecules can be cellular nucleic acid molecules, cell-free nucleic acid molecules, or both. The cell-free nucleic acids used by methods as provided herein can be nucleic acid molecules outside of cells in a biological sample. The cell-free nucleic acid molecules can be present in various bodily fluids, e.g., blood, saliva, semen, and urine. Cell-free DNA molecules can be generated owing to cell death in various tissues that can be caused by health conditions and/or diseases, e.g., tumor invasion or growth, immunological rejection after organ transplantation.

[0089] Cell-free nucleic acid molecules, e.g., cell-free DNA, used in methods as provided herein can exist in plasma, urine, saliva, or serum. Cell-free DNA can occur naturally in the form of short fragments. Cell-free DNA fragmentation can refer to the process whereby high molecular weight DNA (such as DNA in the nucleus of a cell) are cleaved, broken, or digested to short fragments when cell-free DNA molecules are generated or released. Methods, compositions, and systems provided herein can be used to analyze cellular nucleic acid molecules in some cases, for instance, cellular DNA from a tumor tissue, or cellular DNA from white blood cells when the patient has leukemia, lymphoma, or myeloma. Sample taken from a tumor tissue can be subject to assays and analyses according to some examples of the present disclosure.

[0090] VII. Subjects

[0091] Methods, compositions, and systems provided herein can be used to analyze sample from a subject, e.g., organism, e.g., host organism. The subject can be any human patient, such as a cancer patient, a patient at risk for cancer, or a patient with a family or personal history of cancer. In some cases, the subject is in a particular stage of cancer treatment. In some cases, the subject can have or be suspected of having cancer. In some cases, whether the subject has cancer is unknown.

[0092] A subject can have any type of cancer or tumor. In an example, a subject can have colon cancer, or cancer of the large intestine. In another example, a subject can have colorectal cancer, or cancer of the colon and rectum. In another example, a subject can have liver cancer, e.g., hepatocellular carcinoma. Non-limiting examples of cancer can include, but are not limited to, adrenal cancer, anal cancer, basal cell carcinoma, bile duct cancer, bladder cancer, cancer of the blood, bone cancer, a brain tumor, breast cancer, bronchus cancer, cancer of the cardiovascular system, cervical cancer, colon cancer, colorectal cancer, cancer of the digestive

system, cancer of the endocrine system, endometrial cancer, esophageal cancer, eye cancer, gallbladder cancer, a gastrointestinal tumor, hepatocellular carcinoma, kidney cancer, hematopoietic malignancy, laryngeal cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, mesothelioma, cancer of the muscular system, Myelodysplastic Syndrome (MDS), myeloma, nasal cavity cancer, nasopharyngeal cancer, cancer of the nervous system, cancer of the lymphatic system, oral cancer, oropharyngeal cancer, osteosarcoma, ovarian cancer, pancreatic cancer, penile cancer, pituitary tumors, prostate cancer, rectal cancer, renal pelvis cancer, cancer of the reproductive system, cancer of the respiratory system, sarcoma, salivary gland cancer, skeletal system cancer, skin cancer, small intestine cancer, stomach cancer, testicular cancer, throat cancer, thymus cancer, thyroid cancer, a tumor, cancer of the urinary system, uterine cancer, vaginal cancer, or vulvar cancer. The lymphoma can be any type of lymphoma including B-cell lymphoma (e.g., diffuse large B-cell lymphoma, follicular lymphoma, small lymphocytic lymphoma, mantle cell lymphoma, marginal zone B-cell lymphoma, Burkitt lymphoma, lymphoplasmacytic lymphoma, hairy cell leukemia, or primary central nervous system lymphoma) or a T-cell lymphoma (e.g., precursor T-lymphoblastic lymphoma, or peripheral T-cell lymphoma). The leukemia can be any type of leukemia including acute leukemia or chronic leukemia. Types of leukemia include acute myeloid leukemia, chronic myeloid leukemia, acute lymphocytic leukemia, acute undifferentiated leukemia, or chronic lymphocytic leukemia. In some cases, the cancer patient does not have a particular type of cancer. For example, in some instances, the patient can have a cancer that is not breast cancer.

[0093] Examples of cancer include cancers that cause solid tumors as well as cancers that do not cause solid tumors. Furthermore, any of the cancers mentioned herein can be a primary cancer (e.g., a cancer that is named after the part of the body where it first started to grow) or a secondary or metastatic cancer (e.g., a cancer that has originated from another part of the body).

[0094] A subject diagnosed by any of the methods described herein can be of any age and can be an adult, infant or child. In some cases, the subject is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 years old, or within a range therein (e.g., between 2 and 20 years old, between 20 and 40 years old, or between 40 and 90 years old). A particular class of patients that can benefit can be patients over the age of 40. Another particular class of patients that can benefit can be pediatric patients. Furthermore, a subject diagnosed by any of the methods or compositions described herein can be male or female.

[0095] Any of the methods disclosed herein can also be performed on a non-human subject, such as a laboratory or farm animal, or a cellular sample derived from an organism disclosed herein. Non-limiting examples of a non-human subject include a dog, a goat, a guinea pig, a hamster, a mouse, a pig, a non-human primate (e.g., a gorilla, an ape, an orangutan, a lemur, or a baboon), a rat, a sheep, a cow, or a zebrafish.

[0096] As discussed above, the subject methods, compositions, and kits can be used on subjects in various stages of cancer treatment. Results from analyses of cell-free nucleic acids in biological samples of a subject using the subject methods, compositions, and kits can be used for guiding treatment plans for the subject. In some cases, medication or therapies to treat or cure cancer in the subject can be needed. Exemplary treatment options can include chemotherapy, radiation therapy, surgical removal of tumor tissues, immunotherapy, targeted therapy, hormone therapy, and stem cell therapy. In some cases, guidance can be provided as to the choice of the different types of treatment options. In some non-limiting examples, a patient can have finished treatment of a first cancer, e.g., surgical removal of tumor tissue in affected liver leaves, and the patient can be subject to routine monitoring tests using the subject methods, compositions, or kits, in order to examine whether or not there is recurrence of the liver cancer or metastasis. In these cases, the test results can be used to provide guidance as to whether or not the patient would need further treatment of the cancer, and if recurrence of the liver cancer or metastasis into other tissue happens, what treatment options can be applied. In some cases, guidance can be provided as to the specific dosage or administration regimen of the treatment. For instance, the amount of cell-free nucleic acids from a certain tissue can be correlated with the dosage of medication that is to be administered to the patient, or with the frequency/interval of drug administration (e.g., daily, weekly, bi-weekly, or monthly). In some cases, the results from last analysis can be used as basis for evaluating and designing the treatment options and follow-on monitoring analysis.

[0097] **VIII. Computer system**

[0098] Any of the methods disclosed herein can be performed and/or controlled by one or more computer systems. In some examples, any step of the methods disclosed herein can be wholly, individually, or sequentially performed and/or controlled by one or more computer systems. Any of the computer systems mentioned herein can utilize any suitable number of subsystems. In some embodiments, a computer system includes a single computer apparatus, where the subsystems can be the components of the computer apparatus. In other embodiments, a computer system can include multiple computer apparatuses, each being a subsystem, with internal components. A computer system can include desktop and laptop computers, tablets, mobile phones and other mobile devices.

[0099] The subsystems can be interconnected via a system bus. Additional subsystems include a printer, keyboard, storage device(s), and monitor that is coupled to display adapter. Peripherals and input/output (I/O) devices, which couple to I/O controller, can be connected to the computer system by any number of connections known in the art such as an input/output (I/O) port (e.g., USB, FireWire[®]). For example, an I/O port or external interface (e.g., Ethernet, Wi-Fi, etc.) can be used to connect computer system to a wide area network such as the Internet, a mouse input device, or a scanner. The interconnection via system bus allows the central processor to communicate with each subsystem and to control the execution of a plurality of instructions from system memory or the storage device(s) (e.g., a fixed disk, such as a hard drive, or optical disk), as well as the exchange of information between subsystems. The system memory and/or the storage device(s) can embody a computer readable medium. Another subsystem is a data collection device, such as a camera, microphone, accelerometer, and the like. Any of the data mentioned herein can be output from one component to another component and can be output to the user.

[00100] A computer system can include a plurality of the same components or subsystems, e.g., connected together by external interface or by an internal interface. In some embodiments, computer systems, subsystem, or apparatuses can communicate over a network. In such instances, one computer can be considered a client and another computer a server, where each can be part of a same computer system. A client and a server can each include multiple systems, subsystems, or components.

[00101] The present disclosure provides computer control systems that are programmed to implement methods of the disclosure. **Fig. 10** shows a computer system 101 that is programmed or otherwise configured to determine an absolute amount of cell-free nucleic acid molecules from a tissue of an organism as described herein. The computer system 101 can implement and/or regulate various aspects of the methods provided in the present disclosure, such as, for example, controlling sequencing of the nucleic acid molecules from a biological sample, performing various steps of the bioinformatics analyses of sequencing data as described herein, integrating data collection, analysis and result reporting, and data management. The computer system 101 can be an electronic device of a user or a computer system that is remotely located with respect to the electronic device. The electronic device can be a mobile electronic device.

[00102] The computer system 101 includes a central processing unit (CPU, also “processor” and “computer processor” herein) 105, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system 101 also includes memory or memory location 110 (e.g., random-access memory, read-only memory, flash memory), electronic storage unit 115 (e.g., hard disk), communication interface 120 (e.g., network adapter)

for communicating with one or more other systems, and peripheral devices 125, such as cache, other memory, data storage and/or electronic display adapters. The memory 110, storage unit 115, interface 120 and peripheral devices 125 are in communication with the CPU 105 through a communication bus (solid lines), such as a motherboard. The storage unit 115 can be a data storage unit (or data repository) for storing data. The computer system 101 can be operatively coupled to a computer network (“network”) 130 with the aid of the communication interface 120. The network 130 can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network 130 in some cases is a telecommunication and/or data network. The network 130 can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network 130, in some cases with the aid of the computer system 101, can implement a peer-to-peer network, which can enable devices coupled to the computer system 101 to behave as a client or a server.

[00103] The CPU 105 can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions can be stored in a memory location, such as the memory 110. The instructions can be directed to the CPU 105, which can subsequently program or otherwise configure the CPU 105 to implement methods of the present disclosure. Examples of operations performed by the CPU 105 can include fetch, decode, execute, and writeback.

[00104] The CPU 105 can be part of a circuit, such as an integrated circuit. One or more other components of the system 101 can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

[00105] The storage unit 115 can store files, such as drivers, libraries and saved programs. The storage unit 115 can store user data, e.g., user preferences and user programs. The computer system 101 in some cases can include one or more additional data storage units that are external to the computer system 101, such as located on a remote server that is in communication with the computer system 101 through an intranet or the Internet.

[00106] The computer system 101 can communicate with one or more remote computer systems through the network 130. For instance, the computer system 101 can communicate with a remote computer system of a user (e.g., a Smart phone installed with application that receives and displays results of sample analysis sent from the computer system 101). Examples of remote computer systems include personal computers (e.g., portable PC), slate or tablet PC’s (e.g., Apple® iPad, Samsung® Galaxy Tab), telephones, Smart phones (e.g., Apple® iPhone, Android-enabled device, Blackberry®), or personal digital assistants. The user can access the computer system 101 via the network 130.

[00107] Methods as described herein can be implemented by way of machine (e.g., computer processor) executable code stored on an electronic storage location of the computer system 101, such as, for example, on the memory 110 or electronic storage unit 115. The machine executable or machine readable code can be provided in the form of software. During use, the code can be executed by the processor 105. In some cases, the code can be retrieved from the storage unit 115 and stored on the memory 110 for ready access by the processor 105. In some situations, the electronic storage unit 115 can be precluded, and machine-executable instructions are stored on memory 110.

[00108] The code can be pre-compiled and configured for use with a machine having a processor adapted to execute the code, or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

[00109] Aspects of the systems and methods provided herein, such as the computer system 101, can be embodied in programming. Various aspects of the technology can be thought of as “products” or “articles of manufacture” typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such as memory (e.g., read-only memory, random-access memory, flash memory) or a hard disk. “Storage” type media can include any or all of the tangible memory of the computers, processors or the like, or associated modules thereof, such as various semiconductor memories, tape drives, disk drives and the like, which can provide non-transitory storage at any time for the software programming. All or portions of the software can at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, can enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that can bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also can be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible “storage” media, terms such as computer or machine “readable medium” refer to any medium that participates in providing instructions to a processor for execution.

[00110] Hence, a machine readable medium, such as computer-executable code, can take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or

magnetic disks, such as any of the storage devices in any computer(s) or the like, such as can be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media can take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer can read programming code and/or data. Many of these forms of computer readable media can be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[00111] The computer system 101 can include or be in communication with an electronic display 135 that comprises a user interface (UI) 140 for providing, for example, results of sample analysis, such as, but not limited to graphic showings of relative and/or absolute amounts of cell-free nucleic acids from different tissues, control or reference amount of cell-free nucleic acids from certain tissues, comparison between detected and reference amounts, and readout of presence or absence of cancer metastasis. Examples of UI's include, without limitation, a graphical user interface (GUI) and web-based user interface.

[00112] Methods and systems of the present disclosure can be implemented by way of one or more algorithms. An algorithm can be implemented by way of software upon execution by the central processing unit 105. The algorithm can, for example, control sequencing of the nucleic acid molecules from a sample, direct collection of sequencing data, analyzing the sequencing data, or determining a classification of pathology based on the analyses of the sequencing data.

[00113] In some cases, as shown in **Fig. 11**, a sample 202 can be obtained from a subject 201, such as a human subject. A sample 202 can be subjected to one or more methods as described herein, such as performing an assay. In some cases, an assay can comprise hybridization, amplification, sequencing, labeling, epigenetically modifying a base, or any combination thereof. One or more results from a method can be input into a processor 204. One or more input parameters such as sample identification, subject identification, sample type, a reference, or other information can be input into a processor 204. One or more metrics from an assay can be input into a processor 204 such that the processor can produce a result, such as a classification of

pathology (e.g., diagnosis) or a recommendation for a treatment. A processor can send a result, an input parameter, a metric, a reference, or any combination thereof to a display 205, such as a visual display or graphical user interface. A processor 204 can (i) send a result, an input parameter, a metric, or any combination thereof to a server 207, (ii) receive a result, an input parameter, a metric, or any combination thereof from a server 207, (iii) or a combination thereof.

[00114] Aspects of the present disclosure can be implemented in the form of control logic using hardware (e.g., an application specific integrated circuit or field programmable gate array) and/or using computer software with a generally programmable processor in a modular or integrated manner. As used herein, a processor includes a single-core processor, multi-core processor on a same integrated chip, or multiple processing units on a single circuit board or networked. Based on the disclosure and teachings provided herein, a person of ordinary skill in the art will know and appreciate other ways and/or methods to implement embodiments described herein using hardware and a combination of hardware and software.

[00115] Any of the software components or functions described in this application can be implemented as software code to be executed by a processor using any suitable computer language such as, for example, Java, C, C++, C#, Objective-C, Swift, or scripting language such as Perl or Python using, for example, conventional or object-oriented techniques. The software code can be stored as a series of instructions or commands on a computer readable medium for storage and/or transmission. A suitable non-transitory computer readable medium can include random access memory (RAM), a read only memory (ROM), a magnetic medium such as a hard drive or a floppy disk, or an optical medium such as a compact disk (CD) or DVD (digital versatile disk), flash memory, and the like. The computer readable medium can be any combination of such storage or transmission devices.

[00116] Such programs can also be encoded and transmitted using carrier signals adapted for transmission via wired, optical, and/or wireless networks conforming to a variety of protocols, including the Internet. As such, a computer readable medium can be created using a data signal encoded with such programs. Computer readable media encoded with the program code can be packaged with a compatible device or provided separately from other devices (e.g., via Internet download). Any such computer readable medium can reside on or within a single computer product (e.g., a hard drive, a CD, or an entire computer system), and can be present on or within different computer products within a system or network. A computer system can include a monitor, printer, or other suitable display for providing any of the results mentioned herein to a user.

[00117] Any of the methods described herein can be totally or partially performed with a computer system including one or more processors, which can be configured to perform the

steps. Thus, embodiments can be directed to computer systems configured to perform the steps of any of the methods described herein, with different components performing a respective steps or a respective group of steps. Although presented as numbered steps, steps of methods herein can be performed at a same time or in a different order. Additionally, portions of these steps can be used with portions of other steps from other methods. Also, all or portions of a step can be optional. Additionally, any of the steps of any of the methods can be performed with modules, units, circuits, or other approaches for performing these steps.

EXAMPLES

[00118] The examples below further illustrate the described embodiments without limiting the scope of this disclosure.

[00119] Example 1. Materials and Methods

[00120] This example describes several methods utilized for Examples 2-5.

[00121] Subjects

[00122] Patients who previously underwent liver transplantation were recruited during their visits to the liver transplantation clinic from the Department of Surgery of the Prince of Wales Hospital in Hong Kong. Patients with chronic hepatitis B virus (HBV) infection and cirrhosis were recruited from the Department of Medicine and Therapeutics of the Prince of Wales Hospital in Hong Kong. Patients with hepatocellular carcinoma (HCC) and CRC were recruited from the Department of Surgery and the Department of Clinical Oncology of the Prince of Wales Hospital in Hong Kong. The demographics of the recruited subjects are shown in the **Table 2**. All recruited subjects gave written consents. The study was approved by the Joint Hospital Authority New Territories East Cluster - The Chinese University of Hong Kong Clinical Research Ethics Committee.

Table 2. Demographics of subjects analyzed in the study.

Health and disease status	Number	Sex (M:F)	Age median (range)
Healthy subjects	30	16 : 14	57 (26 – 73)
Liver transplant recipients	13	12 : 1	67 (54 – 73)
HCC patients	40	36 : 4	60 (39 – 82)
HBV carriers	20	20 : 0	52 (43 – 77)

Cirrhosis patients	9	9 : 0	62 (53 – 75)
CRC patients with liver metastasis	27	20 : 7	62 (34 – 78)
CRC patients without liver metastasis	35	18 : 17	67 (51 – 83)

[00123] Sample preparation

[00124] For each subject, 10 mL peripheral blood was collected into EDTA-containing tubes. The blood sample was processed within 6 h after blood withdrawal for the separation of plasma and buffy coat. DNA was extracted from plasma using the QIAamp DSP DNA Mini Kit (Qiagen) following the manufacturer's protocol. The DNA extracted from 2 to 4 mL of plasma was subjected to two rounds of bisulfite treatment using an Epitect Plus Bisulfite Kit (Qiagen). The bisulfite converted DNA was eluted in 50 μ L of water for downstream analysis.

[00125] Identification of liver-specific and colon-specific methylation markers

[00126] The methylation profiles of the tissue of interest (i.e. the liver or the colon) were compared with those of other blood cells and tissues to mine a tissue-specific methylation marker. The methylation profiles of different cell types were retrieved from the database of the RoadMap Epigenomics Project for the lung, esophagus, small intestines, colon, pancreas, urinary bladder, heart, and liver, and the database of the BLUEPRINT project for erythroblasts, neutrophils, B-lymphocytes and T-lymphocytes.

[00127] The following criteria for a methylation marker were established.

[00128] 1. A CpG site was defined as hypermethylated in a target tissue (i.e. liver or colon) if the methylation density of the CpG site was >50% in the target tissue and <20% in other blood cells and tissues.

[00129] 2. A stretch of at least 5 hypermethylated CpG sites would be within the differentiated methylated region (DMR) so as to improve the latter's signal-to-noise ratio and analytical specificity of the methylation marker.

[00130] 3. The DMR should be shorter than 166 bp because most of the circulating DNA molecules are short fragments and the peak size is 166 bp.

[00131] DMRs for the liver and colon

[00132] Using the criteria above, one liver-specific marker and one colon-specific marker were identified. The liver-specific DMR was located in the exonic region of the protein tyrosine kinase 2 beta (*PTK2B*) gene on chromosome 8. The eight CpG sites within the liver-specific DMR were hypermethylated in the liver but hypomethylated in other tissues and blood cells (Fig. 7). The *PTK2B* gene is located on chromosome 8 and the genomic coordinates of the CpG sites are shown on the X-axis of Fig. 7. All the eight CpG sites located within the DMR (region

between the two vertical dotted lines) are hypermethylated for the liver compared with other tissues. The region highlighted in yellow contains three CpG sites within the fluorescent probes of the digital PCR assays (see **Table 3**). The other CpG sites on each side of the highlighted region within the DMR were covered by the primers of the digital PCR assays. The colon-specific DMR was within the exonic region of the Sestrin 3 (SESN3) gene on chromosome 11. All the six CpG sites within the colon-specific DMR were hypermethylated in colon tissues (Fig. 8) but hypomethylated in other tissues. The SESN3 is located on chromosome 11 and the genomic coordinates of the CpG sites are shown on the X-axis in Fig. 8. All the six CpG sites located within the DMR (region between the two vertical dotted lines) are hypermethylated for the colon compared with other tissues. The three CpG sites located within the region highlighted in yellow were covered by the fluorescent probes of the colon-specific methylation assay (see **Table 3**). The other CpG sites on each side of the highlighted region within the DMR were covered by the primers of the digital PCR assays. For illustration purpose, individual results in Figs. 7 and 8 for liver, urinary bladder, esophagus, heart, lung, pancreas, and small intestines are not shown, their mean values are represented by “Other tissues”.

[00133] Droplet digital PCR design and performance

[00134] Two droplet digital PCR assays were developed to quantify the methylated and unmethylated DNA molecules in each of liver-specific and colon-specific methylation markers. The sequences of primers and probes for the assays are listed in the **Table 3** (The underlined nucleotides in the primers and probes were the differentially methylated cytosines at the CpG sites). The two droplet digital PCR assays can quantify methylated (from target tissue) and unmethylated (from non-target tissue) using probes labeled with FAM and VIC, respectively. The liver-specific marker was PTK2B gene marker site (chr8: 27,183,116-27,183,176), and the colon-specific marker was SESN3 gene marker site (chr11:94,965,508-94,965,567).

[00135] For each sample, digital PCR analysis was run in duplicates. A total volume of 20 μ L of reaction mix was prepared, containing 8 μ L of bisulfite converted DNA, a final concentration of 450 nM of each of forward primers and reverse primers, 250 nM of unmethylation-specific probe, and 350 nM (liver assay) or 250 nM (colon assay) of methylation-specific probe for the colon assay. The reaction mix was submitted to droplets generation before the PCR reaction using the BioRad QX200 ddPCR droplet generator. Universally methylated DNA (CpGenome Human Methylated DNA from EMD Millipore) and universally unmethylated DNA (EpiTect Unmethylated Human Control DNA from Qiagen) were run on each plate as positive and negative controls. The thermal profile was: 95°C for 10 minutes followed by 45 cycles of 94°C for 15 seconds and 60°C (liver assay) or 56°C (colon assay) for 1 minute and a final incubation at 98°C for 10 minutes. After the PCR, droplets from each sample were analyzed by the QX200

droplet reader and the results were interpreted using the QuantaSoft (version 1.7) software. The cut-off values for positive fluorescence signals were determined with reference to the controls. The numbers of methylated and unmethylated DNA sequences in each sample was calculated using combined counts from duplicate wells followed by Poisson correction. The calculation of the concentration of the methylated or unmethylated DNA sequences in plasma is as follows:

$$Cr = -\ln\left(1 - \frac{P}{R}\right) \times \left(\frac{Ve}{Vd}\right)$$

where Cr represents the concentration of the target molecules in plasma (i.e. the methylated or unmethylated DNA sequences), P represents the number of droplets containing amplified signal for the targeted molecules (either the methylated or unmethylated DNA sequences), R represents the number of total droplets analyzed (with and without amplified signals), V_d represents the mean volume of a droplet (i.e. 0.9×10^{-3} μ L in the current example) and V_e represents the volume of plasma used for the experiment (i.e. 320 μ L in the current example).

Table 3. Oligonucleotide sequences for the digital PCR assays

Primers/probes/marker	Sequence (5'-3')	SEQ ID NO
<i>PTK2B</i> gene marker site	CGGCCGACTTACCTGTACTTGCCGCCGTCCCGGCTCACCTGGCG GTGCCCAGGAGTAGTC	1
Methylated assay (for liver-derived DNA)		
Forward primer (F1)	TTTATTGTT <u>CGGT</u> CGATTATTTGTA	2
Reverse primer (R1)	AACG <u>ACTACTCCTCG</u> AACACCG	3
Fluorescence probe (P1)	5'-FAM-TTGT <u>CGT</u> CGTTT <u>CGGT</u> -MGB-3'	15
Fluorescence probe (P1) sequence	TTGT <u>CGT</u> CGTTT <u>CGGT</u>	4
Unmethylated assay (for non-liver-derived DNA)		
Forward primer (F2)	TGTATATTTATTTGTT <u>GGT</u> IGATTATTTGTA	5
Reverse primer (R2)	CCAACA <u>ACTACTCCTCA</u> AACACCA	6
Fluorescence probe (P2)	5'-VIC-TTTGT <u>IGT</u> IGTTT <u>GGTTA</u> -MGB-3'	16
Fluorescence probe (P2) sequence	TTTGT <u>IGT</u> IGTTT <u>GGTTA</u>	7
<i>SESN3</i> gene marker site	CGCAACCTGCTCCCGAGTGAGAACAAAGGCCGCGCAGAGCGA GAACCTGATTGGTGCCC	8
Methylated assay (for colon-derived DNA)		

Forward primer (F1)	<u>C</u> GTAATTTGTTTT <u>C</u> GAGTGAGAATAA	9
Reverse primer (R1)	AAAAAC <u>C</u> AACACCAATCAAATTC	10
Fluorescence probe (P1)	5'-FAM-T <u>C</u> GG <u>C</u> GTAGAG <u>C</u> GA-MGB-3'	17
Fluorescence probe (P1) sequence	T <u>C</u> GG <u>C</u> GTAGAG <u>C</u> GA	11
Unmethylated assay (for non-colon-derived DNA)		
Forward primer (F2)	GTT <u>I</u> GTAATTTGTTTT <u>I</u> GAGTGAGAATAA	12
Reverse primer (R2)	AAAAAC <u>C</u> AACACCAATCAAATTC	13
Fluorescence probe (P2)	5'-VIC-AGGT <u>I</u> GG <u>I</u> GTAGAG <u>I</u> GA-MGB-3'	18
Fluorescence probe (P2) sequence	AGGT <u>I</u> GG <u>I</u> GTAGAG <u>I</u> GA	14

[00136] Analysis of DNA from different types of samples

[00137] Formalin-fixed paraffin embedded (FFPE) samples of 10 types of tissue (e.g. liver, lung, esophagus, stomach, small intestines, colon, pancreas, urinary bladder, heart, and brain) were retrieved from the Department of Anatomical and Cellular Pathology of the Prince of Wales Hospital in Hong Kong. These tissues were confirmed to be normal on histological examination. Buffy coat samples were collected from healthy subjects. DNA was extracted from FFPE tissues using a QIAamp DNA Mini Kit (Qiagen). DNA from buffy coat was extracted using the QIAamp DNA Blood Mini Kit (Qiagen). Bisulfite conversion was performed using 1 µg of cellular DNA. The converted DNA was eluted in 20 µL water and then diluted 50 fold for downstream analysis.

[00138] Measurement of donor-derived DNA in plasma of liver transplant recipients

[00139] DNA extracted from the liver tissue of donors and the buffy coat of recipients was analyzed to determine genotype information of donors and recipients using the Illumina iScan system. DNA extracted from 4 mL of plasma for each recipient was used for sequencing library preparation. Plasma DNA sequencing libraries were prepared with a KAPA Library Preparation Kit (KAPA Biosystems) following the manufacturer's instructions. The indexed libraries were then multiplexed and sequenced using the Illumina HiSeq 2500 platform (75 × 2 cycles). At least 20 million paired-end reads were obtained for each sample. Paired-end reads were aligned to the non-repeat-masked human reference genome (GRCh 37/hg 19) using the Short Oligonucleotide Alignment Program 2 (SOAP2). Only paired-end reads with both ends aligned to the same chromosome with the correct orientation and aligned to a single location in the human genome were included. The paired-end reads spanning an insert size ≤ 600 bp were retrieved for analysis.

If more than one pair of reads was mapped to the same genome location (i.e. duplicated reads), only one pair of reads was retained for subsequent analysis. Up to two nucleotide mismatches were allowed for either member of the paired-end reads. The fractional concentration of donor-specific DNA in the circulation was determined by counting sequencing reads with single nucleotide polymorphism (SNP) alleles that were homozygous in the recipient and heterozygous in the donor.

[00140] Example 2. Tissue-specificity of the liver and colon markers

[00141] For both the liver-specific and colon-specific markers, the DNA molecules derived from the target tissue would be hypermethylated and those from the non-target tissue would be hypomethylated. Therefore, the percentage of total molecules was denoted being methylated in the liver assay as L% and the percentage of total molecules was denoted being methylated in the colon assay as C%. To confirm the specificity of the liver and colon markers, the DNA extracted from buffy coat samples and 10 types of normal tissues were analyzed using these two digital PCR assay sets. For each type of tissue, 4 samples from different individuals were included.

[00142] The mean L% for the liver tissues was 67% (range: 57% – 76%) and the mean L% for other tissue types were 0.6% (range: 0.0% – 2.2%). The results for each tissue type are summarized in the **Table 4**. These results suggested that the liver assay was able to specifically detect liver-derived DNA.

Table 4. Mean fractional concentration of liver-derived DNA (L%) and colon-derived DNA (C%)

Tissue	Mean L%	Mean C%
Liver	67	0.8
Colon	0.4	22
Brain	1.9	0.8
Esophagus	0.3	1.4
Heart	0.6	0.2
Lung	0.3	1.4
Pancreas	0.3	0.9
Small bowel	0.3	4.1
Stomach	0.1	1.3
Urinary bladder	2.2	1.3
Buffy coat	0.0	0.1

[00143] The mean C% of colon tissues were 22% (range: 17% – 33%). The mean C% for all other tissues was 1.2% (range: 0.1% – 4.1%) indicating the specificity of the methylated sequences as being colon derived. The relatively low C% in colon tissues was probably due to the non-homogenous cellular compositions of the colon tissues. This relatively low C% in colon tissues would not significantly hamper its clinical application when the same assay would be used for comparing the levels of subjects with different disease status.

[00144] Example 3. Plasma concentrations of liver-derived DNA in liver transplant recipients

[00145] The quantitative accuracy of the liver-specific assay was validated through the analysis of the plasma of liver transplant recipients. In these subjects, the fractional concentrations of DNA derived from the transplanted liver could be accurately determined from the proportion of plasma DNA molecules carrying the donor-specific alleles using next-generation sequencing. 14 plasma samples collected from 13 patients who had received liver transplants were analyzed by both the liver-specific methylation marker and sequencing. A positive linear relationship was observed between the concentrations determined by these two methods ($R = 0.99$, $P < 0.0001$, Pearson correlation, Fig. 2) indicating that the liver-specific methylation marker can accurately reflect the concentration of liver-derived DNA in plasma.

[00146] As demonstrated here, the measurement of the percentage contribution of liver DNA concentration by the liver-specific methylation marker correlated well with the results based on the measurement of donor-specific alleles. These results confirmed the accuracy of the liver-specific marker in reflecting the concentration of liver-derived DNA in plasma.

[00147] Example 4. Increased liver-derived DNA in the plasma of HCC patients

[00148] The absolute and fractional concentrations of liver-derived DNA were determined by the digital PCR targeting the sequences with liver-specific methylation pattern in 40 HCC patients, 9 cirrhotic patients, 20 chronic HBV carriers and 30 healthy subjects.

[00149] The median concentration of liver-derived methylated sequences for the healthy subjects, chronic HBV carriers, cirrhotic patients and HCC patients were 40 copies/mL (interquartile range (IQR): 18 – 86), 122 copies/mL (IQR: 47 – 185), 118 copies/mL (IQR: 86 – 159), and 487 (IQR: 138 – 1151), respectively (Fig. 3A). The concentrations were significantly different across the four groups ($P < 0.001$, Kruskal Wallis test). In the post-hoc analysis, the plasma concentrations of liver-derived DNA of the HCC patients were significantly higher than the healthy subjects ($P < 0.001$, Dunn's test) and the chronic HBV carriers ($P = 0.015$, Dunn's test) but not the cirrhotic patient groups ($P = 0.248$, Dunn's test). The concentrations of the

healthy subjects, chronic HBV carriers and cirrhotic patients were not statistically different ($P > 0.05$, Dunn's test).

[00150] The median fractional concentrations of the liver-derived DNA in plasma for the healthy subjects, chronic HBV carriers, cirrhotic patients and HCC patients were 1.4% (IQR: 0.94% – 3.2%), 4.6% (IQR: 1.7% – 6.0%), 3.0% (interquartile range: 1.8% – 7.3%) and 9.4% (IQR: 4.1% – 16.0%), respectively (Fig. 3B). The fractional concentrations were significantly different across the four groups ($P < 0.001$, Kruskal Wallis test). In the post-hoc analysis, the plasma concentrations of liver-derived DNA of the HCC patients were significantly higher than the healthy subjects ($P < 0.001$, Dunn's test), but not the chronic HBV carriers ($P = 0.129$, Dunn's test) and the cirrhotic patients ($P = 0.592$, Dunn's test). The concentrations of the healthy subjects, chronic HBV carriers and cirrhotic patients were not statistically different.

[00151] These results demonstrated that the analysis of both the absolute and fractional concentrations of liver-derived DNA in plasma could differentiate HCC patients from non-HCC subjects (including healthy subjects, chronic HBV carriers and cirrhotic patients). To further determine if absolute or fractional concentration would be better for differentiating HCC and non-HCC subjects, receiver operating characteristic (ROC) curve analysis was performed (Fig. 9). The areas under curve (AUC) were 0.82 and 0.78, for absolute and fractional concentrations, respectively. The difference in the AUCs was statistically significant ($P = 0.022$, DeLong test).

[00152] The correlation between the concentrations (absolute and fractional) of liver-derived DNA in plasma and the maximum dimensions of the tumors (determined either by computed tomography scan or measured after tumor resection) in the HCC patients was further analyzed. Interestingly, the maximal dimensions of the tumors showed a stronger positive correlation with the absolute concentrations ($R = 0.74$, $P < 0.0001$, Spearman correlation) than with the fractional concentration ($R = 0.56$, $P = 0.0002$, Spearman correlation) (Figs. 4A and 4B). The concentration of the liver-derived DNA showed a positive correlation with the maximum dimensions of the tumors in HCC patients indicating that the amount of DNA released from the liver would be reflective of the tumor load.

[00153] Example 5. Analysis of liver- and colon-derived DNA in CRC patients with and without liver metastasis

[00154] The plasma concentrations of liver-derived and colon-derived DNA were measured in 30 healthy subjects, 35 CRC patients without liver metastasis, and 27 CRC patients with liver metastasis. The median plasma concentrations of colon-derived DNA were 0 copy/mL (IQR: 0 – 0), 4 copies/mL (IQR: 0 – 31), and 138 copies/mL (IQR: 0 – 6850) for the three groups, respectively (Fig. 5A). The concentrations were significantly different between the three groups ($P < 0.001$, Kruskal Wallis test). In the post-hoc analysis, the concentrations of the CRC patients

with and without liver metastasis were significantly higher than the healthy subjects ($P < 0.001$ and $P = 0.042$, respectively, Dunn's test). The difference between the CRC patients with and without liver metastasis were not statistically significant ($P = 0.079$, Dunn's test).

[00155] The median fractional concentrations of colon-derived DNA in plasma for the healthy control subjects, CRC patients without liver metastasis, and CRC patients with liver metastasis were 0% (IQR: 0% – 0%), 0.09% (IQR: 0% – 1.1%), and 0.84% (IQR: 0% – 49.5%), respectively (Fig. 5B). The fractional concentrations were significantly different between the three groups ($P < 0.001$, Kruskal Wallis test). In the post-hoc analysis, the concentrations of the CRC patients with and without liver metastasis were significantly higher than the healthy subjects ($P < 0.001$ and $P = 0.041$, respectively, Dunn's test). The difference between the CRC patients with and without liver metastasis was not statistically significant ($P = 0.084$, Dunn's test).

[00156] The median concentrations of liver-derived DNA in plasma for the healthy control subjects, CRC patients without liver metastasis, and CRC patients with liver metastasis were 40 copies/mL (IQR: 18 – 86), 23 copies/mL (IQR: 13 – 108), and 233 copies/mL (IQR: 56 – 2290), respectively (Fig. 5C). The fractional concentrations were significantly different between the 3 groups ($P < 0.001$; Kruskal–Wallis test). In the post hoc analysis, the concentrations of the CRC patients with liver metastases were significantly higher than those without liver metastases and the healthy controls ($P < 0.001$ and $P < 0.001$, respectively; Dunn test). Interestingly, there was no significant difference between patients without liver metastases and healthy controls ($P = 1.0$; Dunn test).

[00157] The median fractional concentrations of liver-derived DNA in plasma for the healthy control subjects, CRC patients without liver metastasis, and CRC patients with liver metastasis were 0.8% (IQR: 0.3% – 2.8%), 1.4% (IQR: 0.9% – 3.3%), and 3.1% (IQR: 1.5% – 5.3%), respectively (Fig. 5D). The fractional concentrations were significantly different between the three groups ($P < 0.001$, Kruskal-Wallis test). In the post-hoc analysis, the concentrations of the CRC patients with liver metastasis were significantly higher than those without liver metastasis ($P < 0.003$, Dunn's test). The fractional concentrations of the healthy subjects were not statistically significantly different from the CRC patients with and without liver metastasis ($P = 0.114$ and $P = 0.717$, respectively).

[00158] As significant difference was observed between the CRC patients with and without liver metastasis in the absolute and fractional concentrations of liver- and colon-derived DNA in plasma, ROC curve analysis was used to determine which parameter would be most useful for differentiating the two groups. The AUCs for the absolute and fractional concentrations of liver-derived DNA were 0.85 and 0.75 ($P = 0.01$, Delong test), respectively, and AUCs for the

absolute and fractional concentrations of colon-derived DNA were 0.69 and 0.69 ($P = 0.75$, Delong test), respectively (Fig. 6).

[00159] The analysis of the absolute concentrations of liver-derived DNA was better than the fractional concentration in differentiating the CRC patients with and without liver metastasis in the ROC analysis (AUC: 0.85 vs 0.75, $P = 0.01$, Fig. 6). Without being bound by theory, the likely explanation is that in some patients with metastatic CRC to the liver, the absolute concentrations of both the liver- and colon-derived DNA would be increased. In some patients, although the absolute concentration of liver-derived DNA was increased, the fractional concentration of the liver remains unchanged or reduced due to a greater increase in the colon-derived DNA. Similarly, it was also shown that the absolute concentration of liver-derived DNA had a better correlation with tumor size compared with the fractional concentration ($R = 0.74$ vs 0.56, Spearman correlation, Fig. 4).

[00160] While preferred embodiments of the present disclosure have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the disclosure. It should be understood that various alternatives to the embodiments of the disclosure described herein can be employed in practicing the disclosure. It is intended that the following claims define the scope of the disclosure and that methods and structures within the scope of these claims and their equivalents be covered thereby.