METHODS AND SYSTEMS FOR MOLECULAR AND RAMAN SPECTROSCOPY ANALYSIS

CROSS-REFERENCE

[0001] This application claims the benefit of Indian Application No. 202411007607, filed February 5, 2024, which is incorporated by reference herein in its entirety.

BACKGROUND

[0002] Cancer is a serious and complex disease that may have profound effects on subjects. Cancer refers to the abnormal and uncontrolled growth and proliferation of cells, which may eventually lead to death. The exact cause for death is largely unknown, however, early detection and intervention can significantly improve the likelihood of survival of subjects who may be at risk of cancer.

SUMMARY

[0003] In an aspect, the present disclosure provides a method for determining a cancer assessment of a subject, comprising: (a) obtaining a whole blood sample from the subject; (b) performing centrifugation on the whole blood sample to obtain a plasma layer, a serum layer, a supernatant layer, or a cell pellet; (c) assaying at least a portion of the plasma layer, the serum layer, the supernatant layer, or the cell pellet, or derivatives thereof, to generate a Raman spectroscopy profile of the whole blood sample of the subject; (d) processing the Raman spectroscopy profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and (e) determining the cancer assessment of the subject, based at least in part on the processing in (d).

[0004] In some embodiments, the subject is asymptomatic for cancer.

[0005] In some embodiments, the subject has a risk factor for cancer.

[0006] In some embodiments, the risk factor for cancer comprises clinical history of cancer, family history of cancer, environmental exposure, smoking history, or genetic variation.

[0007] In some embodiments, the subject has been diagnosed with cancer.

[0008] In some embodiments, the subject is a cancer treatment naïve subject, a cancer survivor subject, a cancer subject receiving treatment, or a benign subject.

[0009] In some embodiments, the centrifugation is density gradient centrifugation, and wherein the density gradient centrifugation comprises use of Ficoll hypaque solution.

[0010] In some embodiments, (b) comprises obtaining any two of the plasma layer, the serum layer, the supernatant layer, or the cell pellet.

[0011] In some embodiments, (b) comprises obtaining any three of the plasma layer, the serum layer, the supernatant layer, or the cell pellet.

[0012] In some embodiments, (b) comprises obtaining the plasma layer, the serum layer, the supernatant layer, and the cell pellet.

[0013] In some embodiments, the method further comprises performing the centrifugation at about 1000g, for a time period of between about 5 minutes to about 20 minutes.

[0014] In some embodiments, the method further comprises extracting nucleic acids from at least a portion of the plasma layer, the serum layer, the supernatant layer, and the cell pellet, and assaying the extracted nucleic acids to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0015] In some embodiments, the nucleic acids comprise deoxyribonucleic acid (DNA).

[0016] In some embodiments, the nucleic acids comprise ribonucleic acid (RNA).

[0017] In some embodiments, the nucleic acids are extracted from the plasma layer.

[0018] In some embodiments, the nucleic acids are extracted from the serum layer.

[0019] In some embodiments, the nucleic acids are extracted from the supernatant layer.

[0020] In some embodiments, the nucleic acids are extracted from the cell pellet.

[0021] In some embodiments, the method further comprises amplifying the extracted nucleic acids.

[0022] In some embodiments, the amplifying comprises polymerase chain reaction (PCR).

[0023] In some embodiments, the method further comprises using primers or probes to selectively enrich the nucleic acids for a set of cancer-specific biomarkers.

[0024] In some embodiments, the primers or probes are nucleic acid primers or nucleic acid probes.

[0025] In some embodiments, the nucleic acid probes have sequence complementarity with nucleic acid sequences of the set of cancer-specific biomarkers.

[0026] In some embodiments, the method further comprises assaying at least a second portion of the extracted nucleic acids by deoxyribonucleic acid (DNA) sequencing, ribonucleic acid (RNA) sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, a methylation array, digital droplet polymerase chain reaction (PCR), digital polymerase chain reaction (PCR), assay for transposase-accessible chromatin (ATAC) sequencing, or methylation-specific polymerase chain reaction (PCR).

[0027] In some embodiments, the method further comprises assaying at least a portion of the plasma layer, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0028] In some embodiments, the method further comprises loading the plasma layer, or derivatives thereof, onto a substrate.

[0029] In some embodiments, the substrate comprises calcium fluoride.

[0030] In some embodiments, the method further comprises assaying at least a portion of the serum layer, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0031] In some embodiments, the method further comprises assaying at least a portion of the cell pellet, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0032] In some embodiments, the method further comprises performing time-gated Raman spectroscopy (TGRS), WITec Raman spectroscopy, or Renishaw Raman spectroscopy.

[0033] In some embodiments, the Raman spectroscopy profile comprises a spectral range for analysis between 600 cm^{-1} and 1,800 cm^{-1}.

[0034] In some embodiments, (c) is performed under light illumination.

[0035] In some embodiments, the method further comprises performing a pre-processing technique on the Raman spectroscopy profile of the whole blood sample of the subject.

[0036] In some embodiments, the pre-processing technique comprises spectral interpolation, spectral smoothing, baseline correction, or normalization, or a combination thereof.

[0037] In some embodiments, the method further comprises performing a univariate or multivariate analysis on the Raman spectroscopy profile of the whole blood sample of the subject.

[0038] In some embodiments, the method further comprises performing a dimensionality reduction on the Raman spectroscopy profile of the whole blood sample of the subject.

[0039] In some embodiments, the dimensionality reduction comprises principal component analysis (PCA) or linear discriminant analysis (LDA).

[0040] In some embodiments, the LDA comprises Principal Component Based Linear Discriminant Analysis (PC-LDA).

[0041] In some embodiments, the method further comprises extracting one or more sets of features, singularly or in combination, from the Raman spectroscopy profile of the whole blood sample of the subject.

[0042] In some embodiments, the one or more sets of features comprises an increase or a decrease in a spectral feature selected from the group consisting of:

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1087- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp
1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1160- Beta-carotene
1518-Beta carotene
1052- C-N Stretch, Phospholipid
1067- C-N Stretch, Phospholipid, C-C Stretch
1107- C-N Stretch, C-C stretch
1130- Protein and Lipid: C-N stretch and chain C-C stretch
1176- Tyr/Phe, CH2 wag, CH2 twist
1210- Tyr/Phe

[0043] In some embodiments, the one or more sets of features comprises 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 11, at least 12, at least 13, at least 14, at least 15, or at least 16 different increases or decreases in the spectral features selected from the group consisting of:

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1087- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp
1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1160- Beta-carotene
1518-Beta carotene
1052- C-N Stretch, Phospholipid
1067- C-N Stretch, Phospholipid, C-C Stretch
1107- C-N Stretch, C-C stretch
1130- Protein and Lipid: C-N stretch and chain C-C stretch
1176- Tyr/Phe, CH2 wag, CH2 twist
1210- Tyr/Phe

[0044] In some embodiments, the set of features comprises intensity values of spectral peaks.

[0045] In some embodiments, the set of features comprises an average intensity value across a plurality of spectral peaks.

[0046] In some embodiments, the set of features comprises a change of the average intensity value as compared to a reference average intensity value.

[0047] In some embodiments, the set of features comprises features that are considered in combination with each other rather than as singular peaks.

[0048] In some embodiments, the method further comprises determining a presence of cancer in the subject or a tissue or location of origin of the cancer.

[0049] In some embodiments, the method further comprises determining a presence or an absence of minimal residual disease or benign lesion in the subject.

[0050] In some embodiments, the method further comprises administering a treatment to the subject based on a detected presence of cancer in the subject.

[0051] In some embodiments, the treatment is selected from the group consisting of surgery, chemotherapy, targeted therapy, radiotherapy, or immunotherapy.

[0052] In some embodiments, the trained machine learning algorithm comprises a deep learning algorithm, a linear or logistic regression, a support vector machine (SVM), a neural network, a Random Forest, or a gradient boosted algorithm.

[0053] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with an accuracy of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0054] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a sensitivity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0055] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a specificity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0056] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a positive predictive value of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0057] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a negative predictive value of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0058] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with an area under receiver operating characteristic curve (AUC) of at least about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, or at least about 0.99.

[0059] In some embodiments, the trained machine learning algorithm is trained using: a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer, wherein the subjects are cancer treatment naïve, and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer.

[0060] In some embodiments, the trained machine learning algorithm is trained using at least one of: a third set of independent training samples obtained or derived from subjects that are cancer treatment naïve; a fourth set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer; a fifth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a sixth set of independent training samples obtained or derived from subjects that are cancer survivors; a seventh set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment; and an eighth set of independent training samples obtained or derived from subjects that benign subjects.

[0061] In some embodiments, the trained machine learning algorithm is trained using at least one of: a third set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer; a fourth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a fifth set of independent training samples obtained or derived from subjects that are cancer survivors; a sixth set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment; a seventh set of independent training samples obtained or derived from subjects that benign subjects; an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions; and a ninth set of independent training samples obtained or derived from subjects known to consume tobacco.

[0062] In some embodiments, the reference is obtained or derived from at least one of: a set of subjects with a presence or an elevated susceptibility of cancer and are cancer treatment naïve; a set of subjects with an absence or a non-elevated susceptibility of cancer; a set of subjects that are cancer treatment naïve; a set of subjects that are indeterminate for a presence or an elevated susceptibility of cancer; a set of subjects that are indeterminate for an absence or a non-elevated susceptibility of cancer; a set of subjects that are cancer survivors; a set of subjects that are cancer subjects receiving treatment; a set of subjects that are benign subjects; a set of subjects that have one or more high risk conditions; and a set of subjects that are known to consume tobacco.

[0063] In some embodiments, the one or more high risk conditions comprise hypertension, diabetes, cardiovascular disease, obesity, subjects that are overweight, or any combination thereof.

[0064] In some embodiments, the cancer is selected from the group consisting of: bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and a cancer with primary unknown origin.

[0065] In some embodiments, the method further comprises using the trained machine learning algorithm to detect a presence or an absence of each of a plurality of different cancer types.

[0066] In some embodiments, the cancer assessment comprises determining a presence of a cancer, an absence of a cancer, or a likelihood or risk of a cancer.

[0067] In some embodiments, the method further comprises determining the presence of the cancer in the subject.

[0068] In some embodiments, the method further comprises determining the absence of the cancer in the subject.

[0069] In some embodiments, the method further comprises determining the likelihood or risk of the cancer in the subject.

[0070] In some embodiments, the reference is generated based on one or more of: non-cancer subjects, cancer treatment naïve subjects, cancer survivor subjects, cancer subjects receiving treatment, and benign subjects.

[0071] In another aspect, the present disclosure provides a method for determining a cancer assessment of a subject, comprising: (a) obtaining a whole blood sample from the subject; (b) performing centrifugation on the whole blood sample to obtain a cell pellet; (c) extracting nucleic acids from the cell pellet; (d) assaying the extracted nucleic acids to generate a genetic profile of the whole blood sample of the subject, a transcriptomic profile of the whole blood sample of the subject, an epigenetic profile of the whole blood sample of the subject, or a combination thereof; (e) processing the genetic profile of the whole blood sample of the subject, the transcriptomic profile of the whole blood sample of the subject, and the epigenetic profile of the whole blood

sample of the subject using a trained machine learning algorithm or against a reference; and (f) determining the cancer assessment of the subject, based at least in part on the processing in (e).

[0072] In some embodiments, the subject is asymptomatic for cancer.

[0073] In some embodiments, the subject has a risk factor for cancer.

[0074] In some embodiments, the risk factor for cancer comprises a clinical history of cancer, family history of cancer, environmental exposure, smoking history, genetic alteration or genetic variation.

[0075] In some embodiments, the subject has been diagnosed with cancer.

[0076] In some embodiments, the subject is a cancer treatment naïve subject, a cancer survivor subject, a cancer subject receiving treatment, or a benign subject.

[0077] In some embodiments, the centrifugation is density gradient centrifugation, and wherein the density gradient centrifugation comprises use of Ficoll hypaque solution.

[0078] In some embodiments, the centrifugation is density gradient centrifugation, and wherein (b) further comprises performing the density gradient centrifugation at about 1000g, for a time period of between about 5 minutes to about 20 minutes.

[0079] In some embodiments, the nucleic acids comprise deoxyribonucleic acid (DNA).

[0080] In some embodiments, the nucleic acids comprise ribonucleic acid (RNA).

[0081] In some embodiments, the method further comprises amplifying the extracted nucleic acids.

[0082] In some embodiments, the amplifying comprises polymerase chain reaction (PCR).

[0083] In some embodiments, the method further comprises using primers or probes to selectively enrich the nucleic acids for a set of cancer-specific biomarkers.

[0084] In some embodiments, the primers or probes are nucleic acid primers or nucleic acid probes.In some embodiments, the nucleic acid probes have sequence complementarity with nucleic acid sequences of the set of cancer-specific biomarkers.

[0085] In some embodiments, the method further comprises assaying at least a second portion of the extracted nucleic acids by deoxyribonucleic acid (DNA) sequencing, ribonucleic acid (RNA) sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, a methylation array, digital droplet polymerase chain reaction (PCR), digital polymerase chain reaction (PCR), assay for transposase-accessible chromatin (ATAC) sequencing, quantitative polymerase chain reaction (PCR), or methylation-specific polymerase chain reaction (PCR).

[0086] In some embodiments, the method further comprises extracting the nucleic acids from very small embryonic-like stem cells (VSELs) of the cell pellet.

[0087] In some embodiments, the method further comprises determining a presence of cancer in the subject or a tissue or location of origin of the cancer.

[0088] In some embodiments, the method further comprises determining a presence or an absence of minimal residual disease or benign lesion in the subject.

[0089] In some embodiments, the method further comprises administering a treatment to the subject based on a detected presence of cancer in the subject.

[0090] In some embodiments, the treatment is subject to a detected modulation of cancer-specific genes or pathways.

[0091] In some embodiments, the treatment is selected from the group consisting of surgery, chemotherapy, targeted therapy, radiotherapy, or immunotherapy.

[0092] In some embodiments, the trained machine learning algorithm comprises a deep learning algorithm, a linear or logistic regression, a support vector machine (SVM), a neural network, a Random Forest, or a gradient boosted algorithm.

[0093] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with an accuracy of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0094] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a sensitivity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0095] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a specificity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0096] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a positive predictive value of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0097] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a negative predictive value of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0098] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with an area under receiver operating characteristic curve (AUC) of at least

about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, or at least about 0.99.

[0099] In some embodiments, the trained machine learning algorithm is trained using: a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer; and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer.

[0100] In some embodiments, the trained machine learning algorithm is trained using at least one of: a third set of independent training samples obtained or derived from subjects that are cancer treatment naïve; a fourth set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer; a fifth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a sixth set of independent training samples obtained or derived from subjects that are cancer survivors; a seventh set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment; and an eighth set of independent training samples obtained or derived from subjects that benign subjects.

[0101] In some embodiments, the trained machine learning algorithm is trained using: a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naïve; and a second set of independent training samples obtained or derived from subjects with an absence of cancer and with a non-elevated susceptibility of cancer.

[0102] In some embodiments, the trained machine learning algorithm is trained using at least one of: a third set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer; a fourth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a fifth set of independent training samples obtained or derived from subjects that are cancer survivors; a sixth set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment; a seventh set of independent training samples obtained or derived from subjects that benign subjects; an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions; and a ninth set of independent training samples obtained or derived from subjects known to consume tobacco.

[0103] In some embodiments, the reference is obtained or derived from at least one of: a set of subjects with a presence or elevated susceptibility of cancer; a set of subjects with an absence or non-elevated susceptibility of cancer; a set of subjects that are cancer treatment naïve; a set of

subjects that are indeterminate for presence or elevated susceptibility of cancer; a set of subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a set of subjects that are cancer survivors; a set of subjects that are cancer subjects receiving treatment; a set of subjects that are benign subjects; a set of subjects that have one or more high risk conditions; and a set of subjects known to consume tobacco.

[0104] In some embodiments, the one or more high risk conditions comprise hypertension, diabetes, cardiovascular disease, obesity, or subjects that are overweight, or any combination thereof.

[0105] In some embodiments, the cancer is selected from the group consisting of: bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and a cancer with primary unknown origin.

[0106] In some embodiments, the method further comprises using the trained machine learning algorithm to detect a presence or an absence of each of a plurality of different cancer types.

[0107] In some embodiments, the cancer assessment comprises determining a presence of cancer, an absence of a cancer, or a likelihood or risk of a cancer.

[0108] In some embodiments, the method further comprises determining the presence of the cancer in the subject.

[0109] In some embodiments, the method further comprises determining the absence of the cancer in the subject.

[0110] In some embodiments, the method further comprises determining the likelihood or risk of the cancer in the subject.

[0111] In some embodiments, the reference is generated based on one or more of: non-cancer subjects, cancer treatment naïve subjects, cancer survivor subjects, cancer subjects receiving treatment, or benign subjects.

[0112] In another aspect, the present disclosure provides a method for evaluating or monitoring a therapy response of a subject with cancer, comprising: (a) obtaining a whole blood sample from the subject, subsequent to the subject being administered a therapy; (b) performing centrifugation on the whole blood sample of the subject to obtain a plasma layer, a serum layer, or a cell pellet;

(c) assaying at least a portion of the plasma layer, the serum layer, or the cell pellet, or derivatives thereof, to generate a Raman spectroscopy profile of the whole blood sample of the subject; (d) processing the Raman spectroscopy profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and (e) evaluating the therapy response of the subject, based at least in part on the processing in (d).

[0113] In some embodiments, the method further comprises obtaining any two of the plasma layer, the serum layer, or the cell pellet.

[0114] In some embodiments, the method further comprises obtaining the plasma layer, the serum layer, and the cell pellet.

[0115] In another aspect, the present disclosure provides a method for determining a cancer assessment of a subject, comprising: (a) obtaining a whole blood sample from the subject; (b) performing centrifugation on the whole blood sample of the subject to obtain a supernatant or a cell pellet; (c) assaying at least a portion of the supernatant or the cell pellet, or derivatives thereof, to generate a Raman spectroscopy profile of the whole blood sample of the subject; (d) processing the Raman spectroscopy profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and (e) determining the cancer assessment of the subject, based at least in part on the processing in (d).

[0116] In some embodiments, the method further comprises obtaining the supernatant and the cell pellet.

[0117] In some embodiments, the supernatant, or derivatives thereof, is assayed to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0118] In some embodiments, the subject is asymptomatic for cancer.

[0119] In some embodiments, the subject has a risk factor for cancer.

[0120] In some embodiments, the risk factor for cancer comprises clinical history of cancer, family history of cancer, environmental exposure, smoking history, or genetic variation.

[0121] In some embodiments, the subject has been diagnosed with cancer.

[0122] In some embodiments, the subject is a cancer treatment naïve subject, a cancer survivor subject, a cancer subject receiving treatment, or a benign subject.

[0123] In some embodiments, the supernatant is incubated one or more times prior to the assaying of (c).

[0124] In some embodiments, the method further comprises performing the centrifugation at about 1000g, for a time period of between about 5 minutes to about 20 minutes.

[0125] In some embodiments, the method further comprises extracting nucleic acids from the supernatant or the cell pellet, and assaying the extracted nucleic acids to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0126] In some embodiments, the nucleic acids comprise deoxyribonucleic acid (DNA).

[0127] In some embodiments, the nucleic acids comprise ribonucleic acid (RNA).

[0128] In some embodiments, the nucleic acids are extracted from the supernatant.

[0129] In some embodiments, the nucleic acids are extracted from the cell pellet.

[0130] In some embodiments, the method further comprises amplifying the extracted nucleic acids.

[0131] In some embodiments, the amplifying comprises polymerase chain reaction (PCR).

[0132] In some embodiments, the method further comprises using primers or probes to selectively enrich the nucleic acids for a set of cancer-specific biomarkers.

[0133] In some embodiments, the primers or probes are nucleic acid primers or nucleic acid probes.

[0134] In some embodiments, the nucleic acid probes have sequence complementarity with nucleic acid sequences of the set of cancer-specific biomarkers.

[0135] In some embodiments, the method further comprises assaying at least a second portion of the extracted nucleic acids by deoxyribonucleic acid (DNA) sequencing, ribonucleic acid (RNA) sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, a methylation array, digital droplet polymerase chain reaction (PCR), digital polymerase chain reaction (PCR), assay for transposase-accessible chromatin (ATAC) sequencing, quantitative polymerase chain reaction (PCR), or methylation-specific polymerase chain reaction (PCR).

[0136] In some embodiments, the method further comprises assaying at least a portion of the plasma layer, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0137] In some embodiments, the method further comprises loading the plasma layer, or derivatives thereof, onto a substrate.

[0138] In some embodiments, the substrate comprises calcium fluoride.

[0139] In some embodiments, the method further comprises assaying at least a portion of the supernatant, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0140] In some embodiments, the method further comprises assaying at least a portion of the cell pellet, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

[0141] In some embodiments, the method further comprises performing time-gated Raman spectroscopy (TGRS), WITec Raman spectroscopy, or Renishaw Raman spectroscopy.

[0142] In some embodiments, the Raman spectroscopy profile comprises a spectral analysis range between 600 cm^{-1} and 1800 cm^{-1}.

[0143] In some embodiments, (c) is performed under light illumination.

[0144] In some embodiments, the method further comprises performing a pre-processing technique on said Raman spectroscopy profile of the whole blood sample of the subject.

[0145] In some embodiments, the pre-processing technique comprises spectral interpolation, spectral smoothing, baseline correction, or normalization, or a combination thereof.

[0146] In some embodiments, the method further comprises performing a univariate or multivariate analysis on the Raman spectroscopy profile of the whole blood sample of the subject.

[0147] In some embodiments, the method further comprises performing a dimensionality reduction on the Raman spectroscopy profile of the whole blood sample of the subject.

[0148] In some embodiments, the dimensionality reduction comprises principal component analysis (PCA) or linear discriminant analysis (LDA).

[0149] In some embodiments, the LDA comprises Principal Component Based Linear Discriminant Analysis (PC-LDA).

[0150] In some embodiments, the method further comprises extracting one or more sets of features, singularly or in combination, from the Raman spectroscopy profile of the whole blood sample of the subject.

[0151] In some embodiments, the one or more sets of features comprises an increase or a decrease in a spectral feature selected from the group consisting of:

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1087- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp

1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1160- Beta-carotene
1518-Beta carotene
1052- C-N Stretch, Phospholipid
1067- C-N Stretch, Phospholipid, C-C Stretch
1107- C-N Stretch, C-C stretch
1130- Protein and Lipid: C-N stretch and chain C-C stretch
1176- Tyr/Phe, CH2 wag, CH2 twist
1210- Tyr/Phe

[0152] In some embodiments, the one or more sets of features comprises 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 11, at least 12, at least 13, at least 14, at least 15, or at least 16 different increases or decreases in the spectral features selected from the group consisting of:

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1087- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp
1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1160- Beta-carotene
1518-Beta carotene

1052- C-N Stretch, Phospholipid
1067- C-N Stretch, Phospholipid, C-C Stretch
1107- C-N Stretch, C-C stretch
1130- Protein and Lipid: C-N stretch and chain C-C stretch
1176- Tyr/Phe, CH2 wag, CH2 twist
1210- Tyr/Phe

[0153] In some embodiments, the one or more sets of features comprises intensity values of spectral peaks.

[0154] In some embodiments, the one or more sets of features comprises an average intensity value across a plurality of spectral peaks.

[0155] In some embodiments, the one or more sets of features comprises a change of an average intensity value as compared to a reference average intensity value.

[0156] In some embodiments, the set of features comprises features that are considered in combination with each other rather than as singular peaks.

[0157] In some embodiments, the method further comprises determining a presence of cancer in the subject or a tissue or location of origin of the cancer.

[0158] In some embodiments, the method further comprises determining a presence or an absence of minimal residual disease or benign lesion in the subject.

[0159] In some embodiments, the method further comprises administering a treatment to the subject based on a detected presence of cancer in the subject.

[0160] In some embodiments, the treatment is selected from the group consisting of surgery, chemotherapy, targeted therapy, radiotherapy, or immunotherapy.

[0161] In some embodiments, the trained machine learning algorithm comprises a deep learning algorithm, a linear or logistic regression, a support vector machine (SVM), a neural network, a Random Forest, or a gradient boosted algorithm.

[0162] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with an accuracy of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0163] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a sensitivity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0164] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a specificity of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0165] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a positive predictive value of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0166] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with a negative predictive value of at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, or at least about 99%.

[0167] In some embodiments, the trained machine learning algorithm detects a presence or an absence of cancer with an area under receiver operating characteristic curve (AUC) of at least about 0.50, at least about 0.55, at least about 0.60, at least about 0.65, at least about 0.70, at least about 0.75, at least about 0.80, at least about 0.85, at least about 0.90, at least about 0.95, or at least about 0.99.

[0168] In some embodiments, the trained machine learning algorithm is trained using: a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer; and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer.

[0169] In some embodiments, the trained machine learning algorithm is trained using at least one of: a third set of independent training samples obtained or derived from subjects that are cancer treatment naïve; a fourth set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer; a fifth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a sixth set of independent training samples obtained or derived from subjects that are cancer survivors; a seventh set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment; and an eighth set of independent training samples obtained or derived from subjects that benign subjects.

[0170] In some embodiments, the trained machine learning algorithm is trained using: a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naïve; and a second set of independent training samples obtained or derived from subjects with an absence of cancer and with a non-elevated susceptibility of cancer.

[0171] In some embodiments, the trained machine learning algorithm is trained using at least one of a third set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer; a fourth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a fifth set of independent training samples obtained or derived from subjects that are cancer survivors; a sixth set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment; a seventh set of independent training samples obtained or derived from subjects that benign subjects; an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions; and a ninth set of independent training samples obtained or derived from subjects known to consume tobacco.

[0172] In some embodiments, the reference is obtained or derived from at least one of: a set of subjects with a presence or elevated susceptibility of cancer; a set of subjects with an absence or non-elevated susceptibility of cancer; a set of subjects that are cancer treatment naïve; a set of subjects that are indeterminate for presence or elevated susceptibility of cancer; a set of subjects that are indeterminate for absence or non-elevated susceptibility of cancer; a set of subjects that are cancer survivors; a set of subjects that are cancer subjects receiving treatment; a set of subjects that are benign subjects; a set of subjects that have one or more high risk conditions; and a set of subjects known to consume tobacco.

[0173] In some embodiments, the one or more high risk conditions comprise hypertension, diabetes, cardiovascular disease, obesity, or subjects that are overweight, or any combination thereof.

[0174] In some embodiments, the cancer is selected from the group consisting of: bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and a cancer with primary unknown origin.

[0175] In some embodiments, the method further comprises using the trained machine learning algorithm to detect a presence or an absence of each of a plurality of different cancer types.

[0176] In some embodiments, the cancer assessment comprises determining a presence of a cancer, an absence of a cancer, or a likelihood or risk of a cancer.

[0177] In some embodiments, the method further comprises determining the presence of the cancer in the subject.

[0178] In some embodiments, the method further comprises determining the absence of the cancer in the subject.

[0179] In some embodiments, the method further comprises determining the likelihood or risk of the cancer in the subject.

[0180] In some embodiments, the reference is generated based on one or more of: non-cancer subjects, cancer treatment naïve subjects, cancer survivor subjects, cancer subjects receiving treatment, or benign subjects.

[0181] In another aspect, the present disclosure provides a system comprising one or more processors and a memory operatively coupled to the one or more processors, wherein the one or more processors are individually or collectively programmed to perform any one of the methods herein.

[0182] Another aspect of the present disclosure provides a system comprising one or more computer processors and computer memory coupled thereto. The computer memory comprises machine executable code that, upon execution by the one or more computer processors, implements any of the methods above or elsewhere herein.

[0183] Additional aspects and advantages of the present disclosure will become readily apparent to those skilled in this art from the following detailed description, wherein only illustrative embodiments of the present disclosure are shown and described. As will be realized, the present disclosure is capable of other and different embodiments, and its several details are capable of modifications in various obvious respects, all without departing from the disclosure. Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0184] 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. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0185] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also "Figure" and "FIG." herein), of which:

[0186] FIG. 1 shows a flowchart of a method 100 for detecting a presence, an absence, or a risk of a cancer in a subject.

[0187] FIG. 2 shows a flowchart of a method 200 for detecting a presence, an absence, or a risk of a cancer in a subject.

[0188] FIG. 3 shows a flowchart for a method 300 for detecting a presence, an absence, or a risk of a cancer in a subject.

[0189] FIG. 4 shows a computer system that is programmed or otherwise configured to implement methods provided herein.

[0190] FIG. 5 shows molecular vibrations modes relevant to Raman scattering.

[0191] FIG. 6 shows Raman scattering in time-gated Raman spectroscopy (TGRS) vs. conventional Raman spectroscopy.

[0192] FIG. 7 shows spectra comparison between plasma sample obtained following centrifugation at 400g and 1000g.

[0193] FIG. 8 shows that Raman scattering occurs almost instantly whereas fluorescence emission is spread over a time range determined by its lifetime.

[0194] FIG. 9 shows a representative Raman spectra of control and cancer plasma samples.

[0195] FIG. 10 shows a comparison of plasma and serum Raman spectra.

DETAILED DESCRIPTION

[0196] While various embodiments of the invention 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 may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

[0197] Whenever the term "at least," "greater than," or "greater than or equal to" precedes the first numerical value in a series of two or more numerical values, the term "at least," "greater than" or "greater than or equal to" applies to each of the numerical values in that series of numerical values. For example, greater than or equal to 1, 2, or 3 is equivalent to greater than or equal to 1, greater than or equal to 2, or greater than or equal to 3.

[0198] Whenever the term "no more than," "less than," or "less than or equal to" precedes the first numerical value in a series of two or more numerical values, the term "no more than," "less than," or "less than or equal to" applies to each of the numerical values in that series of numerical values. For example, less than or equal to 3, 2, or 1 is equivalent to less than or equal to 3, less than or equal to 2, or less than or equal to 1.

[0199] The term "HrC", as used herein, generally refers to a quantitative measure of fold-change of one or more biomarkers and/or a quantitative measure of intensity shifts within a molecular profile. For example, an HrC score may be compared to one or more ranges to determine a presence, absence, or risk of cancer in a subject.

[0200] Though described herein with respect to determining a presence, absence, or risk of a cancer in a subject, the methods and systems of the present disclosure can be used to determine a presence, absence, or risk of other conditions or diseases (e.g., other abnormal conditions or disorders of a biological function or a biological structure such as an organ, that affects part or all of a subject). The methods of the present disclosure may be combined to improve the accuracy, specificity, sensitivity, or the like. For example, a Raman based determination of a presence or absence of a cancer can be combined with a PCR based determination of a presence or absence of a cancer for a single subject, thereby enhancing the predictive power of the method.

[0201] **FIG. 1** shows a flowchart of a method 100 for determining a cancer assessment (e.g., detecting a presence or an absence of a cancer) of a subject. In an operation 110, the method 100 may comprise obtaining a whole blood sample from the subject. The whole blood sample may not be processed prior to the method 100. For example, the whole blood sample may be collected from the subject and subjected to the method 100 without fractionalization or other pre-treatments. In some embodiments, a urine sample or a saliva sample may be obtained and analyzed in place of whole blood.

[0202] The subject may be, for example, a human or non-human mammal. A subject may be afflicted with a cancer or suspected of being afflicted with or having a cancer. The subject may not be suspected of being afflicted with or having the cancer. The subject may be symptomatic (e.g., symptomatic for the cancer). Alternatively, the subject may be asymptomatic (e.g., asymptomatic for the cancer). In some cases, the subject may be treated to alleviate the symptoms of the cancer or cure the subject of the cancer. The subject may have a risk factor for cancer. For example, the subject may have a genetic profile indicating a predisposition to having a cancer. In another example, the subject can have a family history of cancer. Examples of risk factors include, but are not limited to, a clinical history of cancer in the subject (e.g., the subject previously had cancer or pre-cancer), a family history of cancer (e.g., related members of the subject's family had cancer or pre-cancer), environmental exposure (e.g., acute and/or chronic

exposure to one or more carcinogens), smoking history (e.g., a history of the subject or a person close to the subject smoking), genetic variation (e.g., a presence of a gene indicative of a predisposition for cancer), or the like, or any combination thereof. In some cases, the subject has been diagnosed with cancer (e.g., prior to the operation of method 100). For example, the subject can be diagnosed with a cancer and the method 100 can confirm the presence of the cancer in the subject. In another example, the subject can be diagnosed with a first cancer and the method 100 can detect a presence of a second metastasized cancer within the subject. The cancer may be, but is not limited to, for example, bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and/or cancer with primary unknown.

[0203] The cancer may comprise one or more of: Breast - Infiltrating Ductal Carcinoma (NOS), Mucinous Adenocarcinoma, Lobular Carcinoma (NOS), DCIS (Ductal Carcinoma In Situ); Cervical - Adenocarcinoma, Squamous cell carcinoma; Uterine - Endometrioid adenocarcinoma, Malignant mesothelioma, Squamous cell carcinoma, Serous carcinoma, Clear cell adenocarcinoma; Ovarian - Serous tumors, Mucinous tumors, Leiomyosarcoma, Krukenberg tumor, Granulosa cell carcinoma, Serous cyst adenocarcinoma, Adenocarcinoma, Sexcord stromal tumor, Serous papillary tumor; Vulvar and Vaginal - Squamous cell carcinoma; Oesophageal - Squamous cell carcinoma, Adenocarcinoma; Stomach - Adenocarcinoma, Mucinous carcinoma, Neuroendocrine tumors, Signet ring carcinoma, Diffused large B cell lymphoma, Desmoplastic round cell tumor; Duodenum – Adenocarcinoma; Ileo-caecal - Neuroendocrine tumor; Colorectal - Adenocarcinoma, Squamous cell carcinoma, Mucinous carcinoma, Signet ring carcinoma; Appendix - Carcinoid tumor, Adenocarcinoma; Liver - Hepatocellular carcinoma, Adenocarcinoma; Gall Bladder - Adenocarcinoma, Papillary Carcinoma; Bile duct (Ampulla of Vater) – Adenocarcinoma; Pancreatic (Periampullary cancer/Head of pancreas) - Leiomyosarcoma, Cholangiocarcinoma; Oral cavity - Adenocarcinoma, Squamous cell carcinoma, Adenoid cystic carcinoma, Mucoepidermoid carcinoma; Tongue - Pleomorphic adenoma, Squamous cell carcinoma, Adenocarcinoma; Maxilla and Mandible - Follicular Ameloblastoma; Salivary gland - Squamous Cell Carcinoma, Acinic cell carcinoma, Adenocarcinoma, Adenoid cystic carcinoma, Squamous cell carcinoma,

Mucoepidermoid carcinoma (salivary); Oro-pharyngeal, Nasal Cavity and Nasopharyngeal - Squamous Cell Carcinoma, Adenocarcinoma, Tonsil-Squamous cell carcinoma; Larynx cancer - Vocal cord carcinoma-Squamous cell carcinoma; Thyroid - Medullary Carcinoma, Papillary carcinoma, Follicular adenoma, Follicular carcinoma, Anaplastic Thyroid Carcinoma; Hematological and Lymphoid - Acute Lymphoblastic Leukemia, Acute Myeloid Leukemia, Chronic Lymphocytic Leukemia, Angio-immunoblastic T cell lymphoma, Anaplastic large cell lymphoma, Follicular Lymphoma, Hodgkin's Lymphoma, Non-Hodgkin's Lymphoma, Lymphoblastic lymphoma, Multiple myeloma, Diffuse large B cell lymphoma, T lymphoblastic lymphoma, Squamous cell carcinoma, B Cell Lymphoma, Mantle Cell Lymphoma, Lymphocytic NHL, Follicular NHL, Hairy cell leukemia, Multiple myeloma, Lymphocytic Lymphoma; Lung and Mediastinum - Squamous cell carcinoma, Adenocarcinoma, Small cell carcinoma, Non-small cell carcinoma, Large cell carcinoma, Neuroendocrine tumors, Lymphoma, Adenosquamous carcinoma, Mesothelioma; Testicular - Germ cell tumors (Seminoma), Leydig cell tumor, Lymphoma, Teratoma; Penile - Squamous cell carcinoma; Prostate – Adenocarcinoma; Renal - Acinar adenocarcinoma, Renal cell carcinoma, Sarcoma; Bladder - Urothelial carcinoma, Transitional cell carcinoma, Lymphoma; Adrenal – Pheochromocytoma; Bone - Osteosarcoma, Ewings sarcoma, Plasmacytoma; Soft tissue - Leiomyosarcoma, Synovial sarcoma, Dermatofibrosarcoma; Skin and Adnexal tumors - Squamous Cell Carcinoma, Basal Cell Carcinoma, or Melanoma. The cancer may be, but is not limited to, Fibroadenoma, Leiomyoma, Cysts, Benign thyroid nodule, Capillary hemangioma, Polyps, Pleomorphic adenoma, BPH (Benign prostatic hyperplasia), Lipoma, Leukoplakia, Fibroadenosis, Hemangioma, Seborrheic keratosis, Adrenal adenoma, Angiomyolipoma, or Follicular adenoma.

[0204] In an operation 120, the method 100 may comprise performing centrifugation (e.g., density gradient centrifugation) on the whole blood sample to obtain at least one of a plasma layer, a serum layer, a supernatant layer, and a cell pellet.

[0205] The centrifugation may comprise the use of density gradient centrifugation, and may comprise the use of a solution configured for the processing of whole blood by density gradient centrifugation. For example, a Ficoll hypaque solution can be used in the density gradient centrifugation. The operation 120 may comprise performing the density gradient centrifugation at a g-force of at least about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, or more g. The operation 120 may comprise performing the density gradient centrifugation at a g-force of at most about 5,000, 4,500, 4,000, 3,500, 3,000, 2,500, 2,000, 1,500, 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100, or fewer g. The operation 102 may comprise performing the density gradient centrifugation at a g-force in a range as defined by any two of the preceding values. The density gradient centrifugation may be

for at least about 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, or more minutes. The density gradient centrifugation may be for at most about 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, or fewer minutes. The density gradient centrifugation may be for a time in a range as defined by any two of the preceding values. For example, the density gradient centrifugation may be for a time period of from about 5 minutes to about 20 minutes. The density gradient centrifugation may comprise a differential centrifugation, a rate-zonal centrifugation, an isopycnic centrifugation, or the like, or any combination thereof. The density gradient centrifugation may be repeated one or more times.

[0206] In an operation 130, the method 100 may comprise assaying at least a portion of the at least one of the plasma layer, the serum layer, the supernatant layer, the cell pellet, or derivatives thereof to generate a Raman spectroscopy profile of the subject.

[0207] In some cases, operation 130 can comprise extracting nucleic acids from at least a portion of the at least one of the plasma layer, the serum layer, the cell pellet, or any combination thereof. Examples of nucleic acid extraction techniques include, but are not limited to, organic extraction (e.g., use of one or more organic solvents to extract the nucleic acid molecules), solid phase extraction (e.g., use of a binding substrate to bind the nucleic acid, thereby separating it from a sample), chelex extraction (e.g., use of a chelex resin), cetyltrimethylammonium (CTAB) extraction, or the like, or any combination thereof.

[0208] The nucleic acids may be a molecule that comprises one or more nucleotide bases in one or more polynucleotides. The polynucleotides can be, for example, nucleic acid molecules such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), including variants or derivatives thereof (e.g., single stranded DNA). Operation 130 may comprise amplifying the extracted nucleic acids. Examples of nucleic acid amplification processes include, but are not limited to, polymerase chain reaction (PCR) amplification (e.g., digital PCR, quantitative PCR, real time PCR, etc.), isothermal amplification, or the like.

[0209] Operation 130 may comprise use of one or more primers or probes to selectively enrich the nucleic acids. The selective enrichment may be, for example, a set of biomarkers specific to cancer (or a particular cancer). The selective enrichment may enhance a signal generated by the nucleic acids by improving the concentration of the nucleic acids as compared to a sample that has not been enriched. The primers or probes may comprise nucleic acid primers or nucleic acid probes. For example, the nucleic acid primers or nucleic acid probes can comprise sequences complementary to the nucleic acid sequences of the set of specific biomarkers. In this example, the biomarkers can hybridize to the nucleic acid primers or nucleic acid probes, thereby enriching the nucleic acids.

[0210] Operation 130 may comprise assaying at least a second portion of the extracted nucleic acids. The assaying may provide complementary data to the Raman spectroscopy profile. Examples of assays include, but are not limited to, DNA sequencing, RNA sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, use of a methylation array, digital droplet PCR, digital PCR, ATAC sequencing, quantitative PCR, methylation-specific PCR, or the like, or any combination thereof.

[0211] In some cases, operation 130 may comprise assaying at least a portion of the plasma layer, the serum layer, the supernatant layer, or the cell pellet, or derivatives thereof to generate the Raman spectroscopy profile of the subject. For example, the extracted nucleic acids may be assayed to generate the Raman spectroscopy profile of the subject. The assaying may comprise loading at least a portion of the plasma layer or a derivative thereof onto a substrate. For example, a portion of the plasma layer can be removed from a centrifugation tube and applied to a substrate for further assay. In another example, the layer can be removed from a centrifuge tube, amplified to generate an amplification product of the nucleic acid present in the layer, and the amplification product can be loaded onto the substrate. Examples of substrates include, but are not limited to, silicon, silicon oxide, sodium chloride, sodium fluoride, calcium fluoride, metals (e.g., gold, silver, etc.), graphite, glass (e.g., glass slides), or the like, or any combination thereof. Similar to the assaying of the layer, at least a portion of the serum layer, the supernatant, the cell pellet, or derivatives thereof can be assayed by Raman spectroscopy to generate the Raman spectroscopy profile of the subject. In some cases, very small embryonic-like stem cells (VSELs) can be enriched from the cell pellet. Nucleic acids from the VSELs can be extracted and processed as described elsewhere herein.

[0212] The assaying may comprise use of one or more Raman spectroscopy instruments. The Raman spectroscopy instrument may be configured with a laser system configured to expose the sample to an excitation light. The laser system may comprise a pulsed laser. The Raman spectroscopy instrument may comprise a detector configured to record a spectrum of a light generated from the interaction of the excitation light with the sample. For example, the excitation light can be exposed to the sample, the light can interact with various vibrational modes of molecules of the sample to generate an emitted light, and the emitted light can be detected by the detector. The Raman spectroscopy instrument may be configured to record Raman signals with energies of at least about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,100, 1,200, 1,300, 1,400, 1,500, 1,600, 1,700, 1,800, 1,900, 2,000, 2,100, 2,200, 2,300, 2,400, 2,500, 2,600, 2,700, 2,800, 2,900, 3,000, 3,100, 3,200, 3,300, 3,400, 3,500, 3,600, 3,700, 3,800, 3,900, or 4,000 or more wavenumber (cm^{-1}). The Raman spectroscopy instrument may be configured to record Raman signals with energies of at most about 4,000, 3,900, 3,800, 3,700, 3,600, 3,500, 3,400,

3,300, 3,200, 3,100, 3,000, 2,900, 2,800, 2,700, 2,600, 2,500, 2,400, 2,300, 2,200, 2,100, 2,000, 1,900, 1,800, 1,700, 1,600, 1,500, 1,400, 1,300, 1,200, 1,100, 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 50, or fewer wavenumber. The Raman spectroscopy instrument may be configured to record Raman signals with energies in a range as defined by any two of the preceding numbers. For example, the Raman spectroscopy instrument may be configured to record Raman signals with an energy of about 1,000 to about 1,400 wavenumber. In some cases, operation 130 may be performed under light illumination (e.g., illumination by a light other than the excitation light). For example, light biasing may be performed on the sample during recordation of the Raman spectrum. The excitation illumination may be at, for example, 532 nanometers (nm), 785 nm, 1064 nm, or the like.

[0213] The Raman spectroscopy may comprise time-gated Raman spectroscopy (TGRS), WITec Raman spectroscopy, or Renishaw Raman spectroscopy. The TGRS may comprise recording a time resolved Raman spectrum. For example, a pulse of light from a light source can generate a Raman signal upon interaction with the sample, and the TGRS can record a portion of the Raman signal received at a particular time. The TGRS may improve signal by reducing the amount of background signal (e.g., fluorescence signal, phosphorescence signal, etc.). For example, the Raman signal may be on a timescale faster than a fluorescence signal, so by gating the detector to a short time after the pulse of excitation light is delivered to the sample, the longer timescale background processes can be excluded from the signal. The TGRS may comprise gating the Raman signal with a time window of at most about 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, or fewer picoseconds.

[0214] In an operation 140, the method 100 may comprise processing (e.g., computer processing) the Raman spectroscopy profile of the subject using a trained machine learning algorithm and/or against a reference.

[0215] In some cases, the computer processing may comprise performing a pre-processing technique on the Raman spectroscopy profile of the subject. The pre-processing technique may be performed prior to other processing on the Raman spectroscopy profile. For example, the pre-processing technique can be performed on an unprocessed Raman profile. The pre-processing technique may comprise one or more of spectral axis alignment, cosmic ray removal, background correction, spectral interpolation (e.g., interpolation of data points between acquired data points), spectral smoothing (e.g., application of one or more smoothing algorithms (e.g., local regression algorithms, low pass filtering algorithms, moving averages, smoothing splines, etc.), baseline correction (e.g., removal of a baseline signal (e.g., via subtraction)), normalization (e.g., normalization of the data using either maximum value, vector, min-max, standard normal variate and peak normalisations, mean, range, area, etc.), or any combination thereof.

[0216] In some cases, operation 140 may comprise performing one or more of a multivariate and/or univariate analysis, a dimensionality reduction, or the like, or any combination thereof on the Raman spectroscopy profile of the subject. The multivariate analysis may comprise, for example, principal component analysis, multivariate analysis of variance, canonical correspondence analysis, discriminant analysis, clustering systems, or the like, or any combination thereof. The dimensionality reduction analysis may comprise reducing the dimensionality of the Raman spectroscopy profile (e.g., reducing the number of dimensions in the Raman spectroscopy profile). The dimensionality reduction analysis may provide improvements to the computational tractability of the Raman spectroscopy profile (e.g., reduce the amount of computer processing power used to process the Raman spectroscopy profile). The dimensionality reduction analysis may comprise one or more of principal component analysis, linear discriminant analysis (e.g., principal component based linear discriminant analysis, etc.), or the like, or any combination thereof. Unsupervised univariate analysis may include k-means, Hierarchical Cluster Analysis (HCA), PCA or the like or any combination thereof. Supervised multivariate analysis may include Multivariate Curve Resolution (MCR), MCR – Alternating Least Squares (MCR-ALS), Partial Least Square Regression (PLSR), Linear PLS, (L-PLS), LDA, Support Vector Machine (SVM), Partial Least Squares (PLS), Principal Component Regression (PCR), Multiple Linear Regression (MLR), or any combination thereof.

[0217] In some cases, operation 140 may comprise extracting one or more sets of features from the Raman spectroscopy profile of the subject. The features may be related to the presence or absence of the cancer in the subject. The features may comprise peaks in the Raman spectrum of the sample. For example, the features can be wavenumber values and intensity values of an unprocessed Raman spectrum. The features may comprise an average intensity value for a plurality of peaks of the Raman spectrum (e.g., an average value calculated across the plurality of peaks). The features may comprise a comparison of the average intensity value to a reference average intensity value. For example, the features can comprise a comparison of the fold difference of the average intensity value of peaks in a sample, singularly or in combination, as compared to an average intensity value of a sample from a subject known to have or not have cancer. The features may be features derived from the Raman spectrum. For example, the features can comprise identities and concentrations of molecules present in the sample as determined by the Raman spectrum. The features may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, or more of increases or decreases in 622.1-Nucleotide conformation/Phenylalanine, 644-tyrosine doublet/C-S twisting, 757-Guanine/Cytochrome/Protein/Tryptophan, d-ring, 881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein), 942.2-Phenylalanine C-C stretch/ (Amide III helix), 1006-phenylalanine,

1034-Protein: Phenylalanine, 1052 C-N Stretch, Phospholipid, 1067 C-N Stretch, Phospholipid, C-C Stretch, 1087- DNA/RNA symmetric stretching because of parallel dichroism, 1107 C-N Stretch, C-C stretch, 1130 Protein and Lipid: C-N stretch and chain C-C stretch, 1176 Tyr/Phe, CH2 wag, CH2 twist, 1210 Tyr/Phe, 1343-C-H bend/trp, 1407 - ns COO2 (IgG), 1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III, 1558-tyr doublet/trp, 1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch, 1735-Phospholipid: ester carbonyl, 1160- Beta-carotene, 1518-Beta carotene.

[0218] In some cases, the methods described herein may comprise extracting a set of features. The set of features may comprise an increase or a decrease in a spectral feature. The set of features may comprise more than or equal to 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 2,000, 3,000, 4,000, or 5,000 features. The set of features may comprise less than or equal to 5,000, 4,000, 3,000, 2,000, 1,000, 950, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, 100, 50, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, or 2 features. The set of features may comprise about 8 features. The set of features may comprise about 16 features. The set of features may comprise about 1006 features. The set of features may be selected from a group of more than or equal to 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1,000, 2,000, 3,000, 4,000, or 5,000 features. The set of features may be selected from a group of less than or equal to 5,000, 4,000, 3,000, 2,000, 1,000, 950, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150, or 100. The set of features may be selected from a group of about 1006 features.

[0219] In an operation 150, the method 100 may comprise performing the cancer assessment (e.g., detecting the presence or the absence of the cancer) of the subject based at least in part on the processing (e.g., computer processing) of operation 140.

[0220] In some cases, operation 150 comprises determining the presence of the cancer in the subject and/or a tissue or location of origin of the cancer. For example, not only can the presence of the cancer be determined, but the tissue where the cancer originated can also be determined. In some cases, a presence or an absence of a minimal residual disease or benign lesion in the subject can be determined. The minimal residual disease may be a presence of a number of cancer cells remaining in the body of the subject after treatment. The determination of the presence of a minimal residual disease in a subject may improve outcomes for the subject by indicating that additional treatment may be needed to fully remove the cancer of the subject. The trained machine learning algorithm may be configured to detect a presence or absence of each of a plurality of different cancer types. For example, a single trained machine learning algorithm can

be configured to detect a presence or absence of lung cancer, throat cancer, and mouth cancer. In some cases, other conditions such as, for example, a non-cancer disease state, presence of benign lesions, surgery, conditions resulting in organ inflammation, or the like, or any combination thereof may be determined using the Raman spectroscopy profile of the subject.

[0221] In some cases, a treatment (e.g., surgery, chemotherapy, targeted therapy, radiotherapy, immunotherapy, etc.) can be administered to the subject based on the detected presence of the cancer in the subject. For example, the presence of the cancer can lead to prescription and administration of the treatment to address the cancer. In some cases, the determination of the presence of the cancer, as well as other features of the Raman spectroscopy profile of the subject, can inform the types of treatment that are administered. For example, the Raman spectroscopy profile of the subject can inform the subtype of cancer present in the subject which can, in turn, inform the type of treatment provided to the subject.

[0222] The trained machine learning algorithm may comprise one or more of a deep learning algorithm, a linear regression, a logistic regression, a support vector machine, a neural network, a random forest, a gradient boosted algorithm, or the like, or any combination thereof. The trained machine learning algorithm may be trained using a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer. For example, data from a cohort of subjects can be separated into a first set of samples and a second set of samples, and the machine learning algorithm can be trained on the two sets of samples to generate the trained machine learning algorithm. In some cases, the trained machine learning algorithm can be trained using at least one of a third set of independent training samples obtained or derived from subjects that are cancer treatment naïve, a fourth set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer, a fifth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer, a sixth set of independent training samples obtained or derived from subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample), a seventh set of independent training samples obtained or derived from subjects currently undergoing treatment, an eighth set of independent training samples obtained or derived from subjects without cancer but who have another pathology (e.g., organ inflammation, etc.), and a ninth set of independent training samples obtained or derived from subjects with the presence of a benign lesion, or the like, or any combination thereof.

[0223] In some cases, the data from a cohort of subjects can be separated into a third set, a fourth set, a fifth set, a sixth set, a seventh set, an eighth set, a ninth set and/or more of samples, and the machine learning algorithm can be trained on two or more sets of samples to generate the trained machine learning algorithm. In some cases, the trained machine learning algorithm can be trained using at least one of a first set of independent training samples obtained or derived from subjects with an absence of cancer and with a non-elevated susceptibility of cancer, a second set of independent training samples obtained or derived from subjects that have not been diagnosed with cancer but who have another pathology (e.g., organ inflammation) and are indeterminate for the presence or elevated susceptibility of cancer, a third set of independent training samples obtained or derived from subjects that are indeterminate for the absence and/or non-elevated susceptibility of cancer, a fourth set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naïve, a fifth set of independent training samples obtained or derived from subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample) and are radiologically confirmed to be cancer-free, a sixth set of independent training samples obtained or derived from subjects currently undergoing treatment, a seventh set of independent training samples obtained or derived from subjects with the presence of a benign lesion, an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions, and a ninth set of independent training samples obtained or derived from subjects known to consume tobacco in any form (e.g., including smoking tobacco and chewing tobacco), or the like, or any combination thereof. The reference may be obtained or derived from at least one of a set of subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naive, a set of subjects with an absence of cancer and a non-elevated susceptibility of cancer, a set of subjects that are without cancer but who have another pathology (e.g., organ inflammation, etc.) and are indeterminate for presence or elevated susceptibility of cancer, and a set of subjects that are indeterminate for absence and/or non-elevated susceptibility of cancer. Other sets from which the reference may be obtained or derived may be, for example, subjects who have completed treatment for cancer (e.g., at least about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, about 10, about 11, about 12, or more months prior to obtaining the sample) and are radiologically confirmed to be cancer-free, subjects currently undergoing treatment, subjects with a presence of a benign lesion, subjects who have one or more high risk conditions (e.g., hypertension, diabetes, cardiovascular disease, obesity, subjects who are overweight, or the like), subjects who consume tobacco in any form, including smoking tobacco and chewing tobacco, or the like, or any combination thereof. The one or more high risk conditions may comprise a subject having

hypertension, a subject with diabetes, a subject with a cardiovascular disease, a subject with obesity, subjects who are overweight, or the like.

[0224] The inclusion of the additional training sets may improve one or more of the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the trained machine learning algorithm. The reference may be obtained or derived from at least one of a set of subjects with a presence or elevated susceptibility of cancer, a set of subjects with an absence or non-elevated susceptibility of cancer, a set of subjects that are cancer treatment naïve, a set of subjects that are indeterminate for presence or elevated susceptibility of cancer, and a set of subjects that are indeterminate for absence or non-elevated susceptibility of cancer. Other sets of training samples can come from, for example, subjects without cancer but who have another pathology (e.g., organ inflammation, etc.), subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample), subject who are radiologically confirmed to be cancer-free, subjects currently undergoing treatment, subjects with the presence of a benign lesion, subjects who are known to consume tobacco in any form (e.g., including smoking tobacco and chewing tobacco), subjects who have one or more high risk conditions (e.g., hypertension, diabetes, cardiovascular disease, obesity, subjects who are overweight, etc.), or the like, or any combination thereof.

[0225] The trained machine learning algorithm may detects the presence or the absence of cancer with an accuracy, sensitivity, specificity, positive predictive value, negative predictive value, or any combination thereof of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, or more. The trained machine learning algorithm may detect the presence or the absence of cancer with an area under receiver operating characteristic curve of at least about 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.999, or more.

[0226] **FIG. 2** shows a flowchart of a method 200 for detecting a presence or an absence of a cancer in a subject. In an operation 210, the method 200 may comprise obtaining a whole blood sample from the subject. The whole blood sample may not be processed prior to the method 200. For example, the whole blood sample may be collected from the subject and subjected to the method 200 without fractionalization or other pre-treatments. In some embodiments, a urine sample or a saliva sample may be obtained and analyzed in place of whole blood.

[0227] The subject may be, for example, a human or non-human mammal. A subject may be afflicted with a cancer or suspected of being afflicted with or having a cancer. The subject may not be suspected of being afflicted with or having the cancer. The subject may be symptomatic (e.g., symptomatic for the cancer). Alternatively, the subject may be asymptomatic (e.g., asymptomatic for the cancer). In some cases, the subject may be treated to alleviate the

symptoms of the cancer or cure the subject of the cancer. The subject may have a risk factor for cancer. For example, the subject may have a genetic profile indicating a predisposition to having a cancer. In another example, the subject can have a family history of cancer. Examples of risk factors include, but are not limited to, a clinical history of cancer in the subject (e.g., the subject previously had cancer or pre-cancer), a family history of cancer (e.g., related members of the subject's family had cancer or pre-cancer), environmental exposure (e.g., acute and/or chronic exposure to one or more carcinogens), smoking history (e.g., a history of the subject or a person close to the subject smoking), genetic variation (e.g., a presence of a gene indicative of a predisposition for cancer), or the like, or any combination thereof. In some cases, the subject has been diagnosed with cancer (e.g., prior to the operation of method 200). For example, the subject can be diagnosed with a cancer and the method 200 can confirm the presence of the cancer in the subject. In another example, the subject can be diagnosed with a first cancer and the method 200 can detect a presence of a second metastasized cancer within the subject. The cancer may be, for example, bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and/or cancer with primary unknown.

[0228] The cancer may comprise one or more of: Breast - Infiltrating Ductal Carcinoma (NOS), Mucinous Adenocarcinoma, Lobular Carcinoma (NOS), DCIS (Ductal Carcinoma In Situ); Cervical - Adenocarcinoma, Squamous cell carcinoma; Uterine - Endometrioid adenocarcinoma, Malignant mesothelioma, Squamous cell carcinoma, Serous carcinoma, Clear cell adenocarcinoma; Ovarian - Serous tumors, Mucinous tumors, Leiomyosarcoma, Krukenberg tumor, Granulosa cell carcinoma, Serous cyst adenocarcinoma, Adenocarcinoma, Sexcord stromal tumor, Serous papillary tumor; Vulvar and Vaginal - Squamous cell carcinoma; Oesophageal - Squamous cell carcinoma, Adenocarcinoma; Stomach - Adenocarcinoma, Mucinous carcinoma, Neuroendocrine tumors, Signet ring carcinoma, Diffused large B cell lymphoma, Desmoplastic round cell tumor; Duodenum – Adenocarcinoma; Ileo-caecal - Neuroendocrine tumor; Colorectal - Adenocarcinoma, Squamous cell carcinoma, Mucinous carcinoma, Signet ring carcinoma; Appendix - Carcinoid tumor, Adenocarcinoma; Liver - Hepatocellular carcinoma, Adenocarcinoma; Gall Bladder - Adenocarcinoma, Papillary

Carcinoma; Bile duct (Ampulla of Vater) – Adenocarcinoma; Pancreatic (Periampullary cancer/Head of pancreas) - Leiomyosarcoma, Cholangiocarcinoma; Oral cavity - Adenocarcinoma, Squamous cell carcinoma, Adenoid cystic carcinoma, Mucoepidermoid carcinoma; Tongue - Pleomorphic adenoma, Squamous cell carcinoma, Adenocarcinoma; Maxilla and Mandible - Follicular Ameloblastoma; Salivary gland - Squamous Cell Carcinoma, Acinic cell carcinoma, Adenocarcinoma, Adenoid cystic carcinoma, Squamous cell carcinoma, Mucoepidermoid carcinoma (salivary); Oro-pharyngeal, Nasal Cavity and Nasopharyngeal - Squamous Cell Carcinoma, Adenocarcinoma, Tonsil-Squamous cell carcinoma; Larynx cancer - Vocal cord carcinoma-Squamous cell carcinoma; Thyroid - Medullary Carcinoma, Papillary carcinoma, Follicular adenoma, Follicular carcinoma, Anaplastic Thyroid Carcinoma; Hematological and Lymphoid - Acute Lymphoblastic Leukemia, Acute Myeloid Leukemia, Chronic Lymphocytic Leukemia, Angio-immunoblastic T cell lymphoma, Anaplastic large cell lymphoma, Follicular Lymphoma, Hodgkin's Lymphoma, Non-Hodgkin's Lymphoma, Lymphoblastic lymphoma, Multiple myeloma, Diffuse large B cell lymphoma, T lymphoblastic lymphoma, Squamous cell carcinoma, B Cell Lymphoma, Mantle Cell Lymphoma, Lymphocytic NHL, Follicular NHL, Hairy cell leukemia, Multiple myeloma, Lymphocytic Lymphoma; Lung and Mediastinum - Squamous cell carcinoma, Adenocarcinoma, Small cell carcinoma, Non-small cell carcinoma, Large cell carcinoma, Neuroendocrine tumors, Lymphoma, Adenosquamous carcinoma, Mesothelioma; Testicular - Germ cell tumors (Seminoma), Leydig cell tumor, Lymphoma, Teratoma; Penile - Squamous cell carcinoma; Prostate – Adenocarcinoma; Renal - Acinar adenocarcinoma, Renal cell carcinoma, Sarcoma; Bladder - Urothelial carcinoma, Transitional cell carcinoma, Lymphoma; Adrenal – Pheochromocytoma; Bone - Osteosarcoma, Ewings sarcoma, Plasmacytoma; Soft tissue - Leiomyosarcoma, Synovial sarcoma, Dermatofibrosarcoma; Skin and Adnexal tumors - Squamous Cell Carcinoma, Basal Cell Carcinoma, or Melanoma. The cancer may be, but is not limited to, Fibroadenoma, Leiomyoma, Cysts, Benign thyroid nodule, Capillary hemangioma, Polyps, Pleomorphic adenoma, BPH (Benign prostatic hyperplasia), Lipoma, Leukoplakia, Fibroadenosis, Hemangioma, Seborrheic keratosis, Adrenal adenoma, Angiomyolipoma, or Follicular adenoma.

[0229] In another operation 220, the method 200 may comprise performing centrifugation (e.g., density gradient centrifugation) on the whole blood sample to obtain a cell pellet.

[0230] The centrifugation may comprise use of a solution configured for the processing of whole blood by density gradient centrifugation. For example, a Ficoll hypaque solution can be used in the density gradient centrifugation. The operation 220 may comprise performing the centrifugation at a g-force of at least about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, or more g. The operation 220 may

comprise performing the centrifugation at a g-force of at most about 5,000, 4,500, 4,000, 3,500, 3,000, 2,500, 2,000, 1,500, 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100, or fewer g. The operation 220 may comprise performing the centrifugation at a g-force in a range as defined by any two of the preceding values. The centrifugation may be for at least about 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, or more minutes. The centrifugation may be for at most about 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, or fewer minutes. The centrifugation may be for a time in a range as defined by any two of the preceding values. For example, the centrifugation may be for a time period of from about 5 minutes to about 20 minutes. The density gradient centrifugation may comprise a differential centrifugation, a rate-zonal centrifugation, an isopycnic centrifugation, or the like, or any combination thereof. The density gradient centrifugation may be repeated one or more times.

[0231] In another operation 230, the method 200 may comprise assaying the extracted nucleic acids to generate at least one of a genetic profile, a gene expression (transcript or transcriptomic) profile, and an epigenetic profile of the subject.

[0232] In some cases, operation 230 can comprise extracting nucleic acids from at least a portion of the at least one of the cell pellet, or any combination thereof. Examples of nucleic acid extraction techniques include, but are not limited to, organic extraction (e.g., use of one or more organic solvents to extract the nucleic acid molecules), solid phase extraction (e.g., use of a binding substrate to bind the nucleic acid, thereby separating it from a sample), chelex extraction (e.g., use of a chelex resin), cetyltrimethylammonium (CTAB) extraction, or the like, or any combination thereof.

[0233] The nucleic acids may be a molecule that comprises one or more nucleotide bases in one or more polynucleotides. The polynucleotides can be, for example, nucleic acid molecules such as deoxyribonucleic acid (DNA) or ribonucleic acid (RNA), including variants or derivatives thereof (e.g., single stranded DNA). Operation 230 may comprise amplifying the extracted nucleic acids. Examples of nucleic acid amplification processes include, but are not limited to, polymerase chain reaction (PCR) amplification (e.g., digital PCR, quantitative PCR, real time PCR, etc.), isothermal amplification, or the like.

[0234] Operation 230 may comprise use of one or more primers or probes to selectively enrich the nucleic acids. The selective enrichment may be for, for example, a set of biomarkers specific to cancer (or a particular cancer). The selective enrichment may enhance a signal generated by the nucleic acids by improving the concentration of the nucleic acids as compared to a sample that has not been enriched. The primers or probes may comprise nucleic acid primers or nucleic acid probes. For example, the nucleic acid primers or nucleic acid probes can comprise sequences

complementary to the nucleic acid sequences of the set of specific biomarkers. In this example, the biomarkers can hybridize to the nucleic acid primers or nucleic acid probes, thereby enriching the nucleic acids.

[0235] Operation 230 may comprise assaying at least a second portion of the extracted nucleic acids/ The assaying may provide complementary data to the Raman spectroscopy profile. Examples of assays include, but are not limited to, DNA sequencing, RNA sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, use of a methylation array, digital droplet PCR, digital PCR, ATAC sequencing, quantitative PCR, methylation-specific PCR, or the like, or any combination thereof. In some cases, very small embryonic-like stem cells (VSELs) can be enriched from the cell pellet. Nucleic acids from the VSELs can be extracted and processed as described elsewhere herein.

[0236] In another operation 240, the method 200 may comprise processing (e.g., computer processing) the at least one of the genetic profile, the gene expression (transcript or transcriptomic) profile, and the epigenetic profile of the subject using a trained machine learning algorithm or against a reference.

[0237] Operation 240 may include discontinuing the use of a housekeeping gene. The housekeeping gene, as it is used with operation 240, has been found to vary considerably between different subjects, and more accurate results may be obtained from operation 240 without it. An internal control database may be used as an internal reference in the place of a housekeeping gene.

[0238] In another operation 250, the method 200 may comprise determining the cancer assessment (e.g., detecting the presence, absence, or risk of the cancer) of the subject based at least in part on the processing of operation 240.

[0239] In some cases, operation 250 comprises determining the cancer assessment (e.g., detecting the presence of the cancer) of the subject and/or a tissue or location of origin of the cancer. For example, not only can the presence of the cancer be determined, but the tissue where the cancer originated can also be determined. In some cases, a presence or an absence of a minimal residual disease in the subject can be determined. The minimal residual disease may be a presence of a number of cancer cells remaining in the body of the subject after treatment. The determination of the presence of a minimal residual disease in a subject may improve outcomes for the subject by indicating that additional treatment may be needed to fully remove the cancer of the subject. The trained machine learning algorithm may detect a presence or absence of each of a plurality of different cancer types. For example, a single trained machine learning algorithm can detect a presence or absence of lung cancer, throat cancer, and mouth cancer. In some cases, other conditions such as, for example, a non-cancer disease state, presence of benign lesions,

surgery, conditions resulting in organ inflammation, or the like, or any combination thereof may be determined using the Method 200 profile of the subject.

[0240] In some cases, a treatment (e.g., surgery, chemotherapy, targeted therapy, radiotherapy, immunotherapy, etc.) can be administered to the subject based on the detected presence of the cancer in the subject. For example, the presence of the cancer can lead to prescription and administration of the treatment to address the cancer. In some cases, the determination of the presence of the cancer, as well as other features of the Method 2 profile of the subject, can inform the types and/or dosage of treatment that are administered. For example, the Method 2 profile of the subject can inform the subtype of cancer present and/or the modulation of cancer specific genes and/or pathways in the subject which can, in turn, inform the type of treatment provided to the subject.

[0241] The trained machine learning algorithm may comprise one or more of a deep learning algorithm, a linear regression, a logistic regression, a support vector machine, a neural network, a random forest, a gradient boosted algorithm, or the like, or any combination thereof. The trained machine learning algorithm may be trained using a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer. For example, data from a cohort of subjects can be separated into a first set of samples and a second set of samples, and the machine learning algorithm can be trained on the two sets of samples to generate the trained machine learning algorithm. In some cases, the trained machine learning algorithm can be trained using at least one of a third set of independent training samples obtained or derived from subjects that are cancer treatment naïve, a fourth set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer, and a fifth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer, a sixth set of independent training samples obtained or derived from subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample), a seventh set of independent training samples obtained or derived from subjects currently undergoing treatment, an eighth set of independent training samples obtained or derived from subjects without cancer but who have another pathology (e.g., organ inflammation, etc.), and a ninth set of independent training samples obtained or derived from subjects with the presence of a benign lesion, or the like, or any combination thereof. The inclusion of the additional training sets may improve one or more of the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the trained machine learning algorithm. The reference may be obtained or derived from

at least one of a set of subjects with a presence or elevated susceptibility of cancer, a set of subjects with an absence or non-elevated susceptibility of cancer, a set of subjects that are cancer treatment naïve, a set of subjects that are indeterminate for presence or elevated susceptibility of cancer, and a set of subjects that are indeterminate for absence or non-elevated susceptibility of cancer. Other sets from which the reference may be obtained or derived may be, for example, subjects without cancer but who have another pathology (e.g., organ inflammation, etc.), subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample), subjects currently undergoing treatment, subjects with the presence of a benign lesion, or the like, or any combination thereof.

[0242] The trained machine learning algorithm may comprise one or more of a deep learning algorithm, a linear regression, a logistic regression, a support vector machine, a neural network, a random forest, a gradient boosted algorithm, or the like, or any combination thereof. The trained machine learning algorithm may be trained using a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer. For example, data from a cohort of subjects can be separated into a first set of samples and a second set of samples, and the machine learning algorithm can be trained on the two sets of samples to generate the trained machine learning algorithm. In some cases, the data from a cohort of subjects can be separated into a third set, fourth set, fifth set, sixth set, seventh set, eighth set, ninth set and/or more of samples, and the machine learning algorithm can be trained on two or more sets of samples to generate the trained machine learning algorithm. In some cases, the trained machine learning algorithm can be trained using at least one of a first set of independent training samples obtained or derived from subjects with an absence of cancer and with a non-elevated susceptibility of cancer, a second set of independent training samples obtained or derived from subjects that have not been diagnosed with cancer but who have another pathology (e.g., organ inflammation) and are indeterminate for the presence or elevated susceptibility of cancer, a third set of independent training samples obtained or derived from subjects that are indeterminate for the absence and/or non-elevated susceptibility of cancer, a fourth set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naïve, a fifth set of independent training samples obtained or derived from subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample) and are radiologically confirmed to be cancer-free, a sixth set of independent training samples obtained or derived from subjects currently undergoing treatment, a seventh set of independent training samples obtained or derived from subjects with the presence

of a benign lesion, an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions (e.g., hypertension, diabetes, cardiovascular disease, obesity, subjects who are overweight, or the like), and a ninth set of independent training samples obtained or derived from subjects who consume tobacco in any form (e.g., including smoking tobacco and chewing tobacco), or the like, or any combination thereof. The inclusion of the additional training sets may improve one or more of the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the trained machine learning algorithm. The reference may be obtained or derived from at least one of a set of subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naive, a set of subjects with an absence of cancer and a non-elevated susceptibility of cancer, a set of subjects that are without cancer but who have another pathology (e.g., organ inflammation, etc.) and are indeterminate for presence or elevated susceptibility of cancer, and a set of subjects that are indeterminate for absence and/or non-elevated susceptibility of cancer. Other sets of training samples can come from, for example, subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample) and are radiologically confirmed to be cancer-free, subjects currently undergoing treatment, subjects with the presence of a benign lesion, subjects who have one or more high risk conditions (e.g., hypertension, diabetes, cardiovascular disease, obesity, subjects who are overweight, or the like), subjects who consume tobacco in any form including smoking tobacco and chewing tobacco, or the like, or any combination thereof. The one or more high risk conditions may comprise a subject having hypertension, a subject with diabetes, a subject with a cardiovascular disease, a subject with obesity, subjects who are overweight, or the like.

[0243] The trained machine learning algorithm may detect the presence or the absence of cancer with an accuracy, sensitivity, specificity, positive predictive value, negative predictive value, or any combination thereof of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, or more. The trained machine learning algorithm may detect the presence or the absence of cancer with an area under receiver operating characteristic curve of at least about 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.999, or more.

[0244] In addition to determining a presence or absence of a cancer in a subject, the methods of the present disclosure may be used to evaluate a therapy response of a subject with cancer (e.g., determine an efficacy of the therapy to the subject). For example, using a Raman spectroscopy profile of a subject who has undergone treatment can provide an indication as to the effect the treatment has on the subject by, for example, determining an amount of the cancer in the subject, determining a change in the cancer of the subject as compared to before the treatment,

determining changes in the cancer caused by the treatment, or the like, or any combination thereof.

[0245] FIG. 3 shows a flowchart of a method 300 for detecting a presence, an absence, or a risk of a cancer in a subject. In an operation 310, the method 300 may comprise obtaining a whole blood sample from the subject. The whole blood sample may not be processed prior to the method 300. For example, the whole blood sample may be collected from the subject and subjected to the method 300 without fractionalization or other pre-treatments. In some embodiments, a urine sample or a saliva sample may be obtained and analyzed in place of whole blood.

[0246] The subject may be, for example, a human or non-human mammal. A subject may be afflicted with a cancer or suspected of being afflicted with or having a cancer. The subject may not be suspected of being afflicted with or having the cancer. The subject may be symptomatic (e.g., symptomatic for the cancer). Alternatively, the subject may be asymptomatic (e.g., asymptomatic for the cancer). In some cases, the subject may be treated to alleviate the symptoms of the cancer or cure the subject of the cancer. The subject may have a risk factor for cancer. For example, the subject may have a genetic profile indicating a predisposition to having a cancer. In another example, the subject can have a family history of cancer. Examples of risk factors include, but are not limited to, a clinical history of cancer in the subject (e.g., the subject previously had cancer or pre-cancer), a family history of cancer (e.g., related members of the subject's family had cancer or pre-cancer), environmental exposure (e.g., acute and/or chronic exposure to one or more carcinogens), smoking history (e.g., a history of the subject or a person close to the subject smoking), genetic variation (e.g., a presence of a gene indicative of a predisposition for cancer), or the like, or any combination thereof. In some cases, the subject has been diagnosed with cancer (e.g., prior to the operation of method 300). For example, the subject can be diagnosed with a cancer and the method 300 can confirm the presence of the cancer in the subject. In another example, the subject can be diagnosed with a first cancer and the method 300 can detect a presence of a second metastasized cancer within the subject. The cancer may be, for example, bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer,

testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and/or cancer with primary unknown.

[0247] The cancer may comprise one or more of: Breast - Infiltrating Ductal Carcinoma (NOS), Mucinous Adenocarcinoma, Lobular Carcinoma (NOS), DCIS (Ductal Carcinoma In Situ); Cervical - Adenocarcinoma, Squamous cell carcinoma; Uterine - Endometrioid adenocarcinoma, Malignant mesothelioma, Squamous cell carcinoma, Serous carcinoma, Clear cell adenocarcinoma; Ovarian - Serous tumors, Mucinous tumors, Leiomyosarcoma, Krukenberg tumor, Granulosa cell carcinoma, Serous cyst adenocarcinoma, Adenocarcinoma, Sexcord stromal tumor, Serous papillary tumor; Vulvar and Vaginal - Squamous cell carcinoma; Oesophageal - Squamous cell carcinoma, Adenocarcinoma; Stomach - Adenocarcinoma, Mucinous carcinoma, Neuroendocrine tumors, Signet ring carcinoma, Diffused large B cell lymphoma, Desmoplastic round cell tumor; Duodenum – Adenocarcinoma; Ileo-caecal - Neuroendocrine tumor; Colorectal - Adenocarcinoma, Squamous cell carcinoma, Mucinous carcinoma, Signet ring carcinoma; Appendix - Carcinoid tumor, Adenocarcinoma; Liver - Hepatocellular carcinoma, Adenocarcinoma; Gall Bladder - Adenocarcinoma, Papillary Carcinoma; Bile duct (Ampulla of Vater) – Adenocarcinoma; Pancreatic (Periampullary cancer/Head of pancreas) - Leiomyosarcoma, Cholangiocarcinoma; Oral cavity - Adenocarcinoma, Squamous cell carcinoma, Adenoid cystic carcinoma, Mucoepidermoid carcinoma; Tongue - Pleomorphic adenoma, Squamous cell carcinoma, Adenocarcinoma; Maxilla and Mandible - Follicular Ameloblastoma; Salivary gland - Squamous Cell Carcinoma, Acinic cell carcinoma, Adenocarcinoma, Adenoid cystic carcinoma, Squamous cell carcinoma, Mucoepidermoid carcinoma (salivary); Oro-pharyngeal, Nasal Cavity and Nasopharyngeal - Squamous Cell Carcinoma, Adenocarcinoma, Tonsil-Squamous cell carcinoma; Larynx cancer - Vocal cord carcinoma-Squamous cell carcinoma; Thyroid - Medullary Carcinoma, Papillary carcinoma, Follicular adenoma, Follicular carcinoma, Anaplastic Thyroid Carcinoma; Hematological and Lymphoid - Acute Lymphoblastic Leukemia, Acute Myeloid Leukemia, Chronic Lymphocytic Leukemia, Angio-immunoblastic T cell lymphoma, Anaplastic large cell lymphoma, Follicular Lymphoma, Hodgkin's Lymphoma, Non-Hodgkin's Lymphoma, Lymphoblastic lymphoma, Multiple myeloma, Diffuse large B cell lymphoma, T lymphoblastic lymphoma, Squamous cell carcinoma, B Cell Lymphoma, Mantle Cell Lymphoma, Lymphocytic NHL, Follicular NHL, Hairy cell leukemia, Multiple myeloma, Lymphocytic Lymphoma; Lung and Mediastinum - Squamous cell carcinoma, Adenocarcinoma, Small cell carcinoma, Non-small cell carcinoma, Large cell carcinoma, Neuroendocrine tumors, Lymphoma, Adenosquamous carcinoma, Mesothelioma; Testicular - Germ cell tumors (Seminoma), Leydig cell tumor, Lymphoma, Teratoma; Penile - Squamous cell carcinoma; Prostate – Adenocarcinoma; Renal -

Acinar adenocarcinoma, Renal cell carcinoma, Sarcoma; Bladder - Urothelial carcinoma, Transitional cell carcinoma, Lymphoma; Adrenal – Pheochromocytoma; Bone - Osteosarcoma, Ewings sarcoma, Plasmacytoma; Soft tissue - Leiomyosarcoma, Synovial sarcoma, Dermatofibrosarcoma; Skin and Adnexal tumors - Squamous Cell Carcinoma, Basal Cell Carcinoma, or Melanoma. The cancer may be, but is not limited to, Fibroadenoma, Leiomyoma, Cysts, Benign thyroid nodule, Capillary hemangioma, Polyps, Pleomorphic adenoma, BPH (Benign prostatic hyperplasia), Lipoma, Leukoplakia, Fibroadenosis, Hemangioma, Seborrheic keratosis, Adrenal adenoma, Angiomyolipoma, or Follicular adenoma.

[0248] In another operation 320, the method 300 may comprise performing centrifugation (e.g., density gradient centrifugation) on the whole blood sample to obtain a plasma layer and/or a serum layer and/or a supernatant layer and a cell pellet.

[0249] The centrifugation may comprise use of a solution configured for the processing of whole blood by density gradient centrifugation. For example, a Ficoll hypaque solution can be used in the density gradient centrifugation. The operation 320 may comprise performing the centrifugation at a g-force of at least about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 5,000, or more g. The operation 320 may comprise performing the centrifugation at a g-force of at most about 5,000, 4,500, 4,000, 3,500, 3,000, 2,500, 2,000, 1,500, 1,000, 900, 800, 700, 600, 500, 400, 300, 200, 100, or fewer g. The operation 320 may comprise performing the centrifugation at a g-force in a range as defined by any two of the preceding values. The centrifugation may be for at least about 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, or more minutes. The centrifugation may be for at most about 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, or fewer minutes. The centrifugation may be for a time in a range as defined by any two of the preceding values. For example, the centrifugation may be for a time period of from about 5 minutes to about 20 minutes. The density gradient centrifugation may comprise a differential centrifugation, a rate-zonal centrifugation, an isopycnic centrifugation, or the like, or any combination thereof. The density gradient centrifugation may be repeated one or more times.

[0250] In another operation 330, the method 300 may comprise conducting the method 100 upon the sample, operations 130 through 150, to obtain a result.

[0251] In another operation 340, the method 300 may comprise conducting the method 200 upon the sample, operations 230 through 250, to obtain a result.

[0252] In another operation 350, the method 300 may comprise combining the results of method 100 and method 200 to create the Molecular Profile. Operation 350 may comprise processing

(e.g., computer processing) the Molecular Profile of the subject using a trained machine learning algorithm, a Tzar Labs algorithm and/or against a reference.

[0253] In another operation 360, the method 300 may comprise determining the cancer assessment (e.g., detecting the presence, absence, or risk of the cancer) of the subject based at least in part on the processing of operation 350.

[0254] In some cases, operation 360 comprises determining the cancer assessment (e.g., detecting the presence of the cancer) of the subject and/or a tissue or location of origin of the cancer. For example, not only can the presence of the cancer be determined, but the tissue where the cancer originated can also be determined. In some cases, a presence or an absence of a minimal residual disease in the subject can be determined. The minimal residual disease may be a presence of a number of cancer cells remaining in the body of the subject after treatment. The determination of the presence of a minimal residual disease in a subject may improve outcomes for the subject by indicating that additional treatment may be needed to fully remove the cancer of the subject. The trained machine learning algorithm may detect a presence or absence of each of a plurality of different cancer types. For example, a single trained machine learning algorithm can detect a presence or absence of lung cancer, throat cancer, and mouth cancer. In some cases, other conditions such as, for example, a non-cancer disease state, presence of benign lesions, surgery, conditions resulting in organ inflammation, or the like, or any combination thereof may be determined using the Molecular Profile of the subject.

[0255] In some cases, a treatment (e.g., surgery, chemotherapy, targeted therapy, radiotherapy, immunotherapy, etc.) can be administered to the subject based on the detected presence of the cancer in the subject. For example, the presence of the cancer can lead to prescription and administration of the treatment to address the cancer. In some cases, the determination of the presence of the cancer, as well as other features of the Raman spectroscopy profile of the subject, can inform the types and/or dosage of treatment that are administered. For example, the Molecular Profile of the subject can inform the subtype of cancer present and/or the modulation of cancer specific genes and/or pathways in the subject which can, in turn, inform the type of treatment provided to the subject.

[0256] The trained machine learning algorithm may comprise one or more of a deep learning algorithm, a linear regression, a logistic regression, a support vector machine, a neural network, a random forest, a gradient boosted algorithm, or the like, or any combination thereof. The trained machine learning algorithm may be trained using a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer. For example, data from a cohort of subjects can be

separated into a first set of samples and a second set of samples, and the machine learning algorithm can be trained on the two sets of samples to generate the trained machine learning algorithm. In some cases, the trained machine learning algorithm can be trained using at least one of a third set of independent training samples obtained or derived from subjects that are cancer treatment naïve, a fourth set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer, and a fifth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer, a sixth set of independent training samples obtained or derived from subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample), a seventh set of independent training samples obtained or derived from subjects currently undergoing treatment, an eighth set of independent training samples obtained or derived from subjects without cancer but who have another pathology (e.g., organ inflammation, etc.), and a ninth set of independent training samples obtained or derived from subjects with the presence of a benign lesion, or the like, or any combination thereof. The inclusion of the additional training sets may improve one or more of the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the trained machine learning algorithm. The reference may be obtained or derived from at least one of a set of subjects with a presence or elevated susceptibility of cancer, a set of subjects with an absence or non-elevated susceptibility of cancer, a set of subjects that are cancer treatment naïve, a set of subjects that are indeterminate for presence or elevated susceptibility of cancer, and a set of subjects that are indeterminate for absence or non-elevated susceptibility of cancer. Other sets of training samples can come from, for example, subjects without cancer but who have another pathology (e.g., organ inflammation, etc.), subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample), subjects currently undergoing treatment, subjects with the presence of a benign lesion, or the like, or any combination thereof.

[0257] The trained machine learning algorithm may comprise one or more of a deep learning algorithm, a linear regression, a logistic regression, a support vector machine, a neural network, a random forest, a gradient boosted algorithm, or the like, or any combination thereof. The trained machine learning algorithm may be trained using a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer. For example, data from a cohort of subjects can be separated into a first set of samples and a second set of samples, and the machine learning algorithm can be trained on the two sets of samples to generate the trained machine learning

algorithm. In some cases, the data from a cohort of subjects can be separated into a third set, fourth set, fifth set, sixth set, seventh set, eighth set, ninth set and/or more of samples, and the machine learning algorithm can be trained on two or more sets of samples to generate the trained machine learning algorithm. In some cases, the trained machine learning algorithm can be trained using at least one of a first set of independent training samples obtained or derived from subjects with an absence of cancer and with a non-elevated susceptibility of cancer, a second set of independent training samples obtained or derived from subjects that have not been diagnosed with cancer but who have another pathology (e.g., organ inflammation) and are indeterminate for the presence or elevated susceptibility of cancer, a third set of independent training samples obtained or derived from subjects that are indeterminate for the absence and/or non-elevated susceptibility of cancer, a fourth set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naïve, a fifth set of independent training samples obtained or derived from subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample) and are radiologically confirmed to be cancer-free, a sixth set of independent training samples obtained or derived from subjects currently undergoing treatment, a seventh set of independent training samples obtained or derived from subjects with the presence of a benign lesion, an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions, and a ninth set of independent training samples obtained or derived from subjects who consume tobacco in any form (e.g., including smoking tobacco and chewing tobacco), or the like, or any combination thereof. The inclusion of the additional training sets may improve one or more of the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value of the trained machine learning algorithm. The reference may be obtained or derived from at least one of a set of subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naive, a set of subjects with an absence of cancer and a non-elevated susceptibility of cancer, a set of subjects that are without cancer but who have another pathology (e.g., organ inflammation, etc.) and are indeterminate for presence or elevated susceptibility of cancer, and a set of subjects that are indeterminate for absence and/or non-elevated susceptibility of cancer. Other sets of training samples can come from, for example, subjects who have completed treatment for cancer (e.g., at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or more months prior to taking the sample) and are radiologically confirmed to be cancer-free, subjects currently undergoing treatment, subjects with the presence of a benign lesion, subjects who have one or more high risk conditions (e.g., hypertension, diabetes, cardiovascular disease, obesity, subjects who are overweight, or the like), subjects who consume tobacco in any form including smoking tobacco and chewing tobacco, or

the like, or any combination thereof. The one or more high risk conditions may comprise a subject having hypertension, a subject with diabetes, a subject with a cardiovascular disease, a subject with obesity, subjects who are overweight, or the like.

[0258] The trained machine learning algorithm may detect the presence or the absence of cancer with an accuracy, sensitivity, specificity, positive predictive value, negative predictive value, or any combination thereof of at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.9%, or more. The trained machine learning algorithm may detect the presence or the absence of cancer with an area under receiver operating characteristic curve of at least about 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 0.99, 0.999, or more.

[0259] In addition to determining a presence or absence of a cancer in a subject, the methods of the present disclosure may be used to evaluate a therapy response of a subject with cancer (e.g., determine an efficacy of the therapy to the subject). For example, using a Molecular profile of a subject who has undergone treatment can provide an indication as to the effect the treatment has on the subject by, for example, determining an amount of the cancer in the subject, determining a change in the cancer of the subject as compared to before the treatment, determining changes in the cancer caused by the treatment, or the like, or any combination thereof.

[0260] **Computer systems**

[0261] The present disclosure provides computer systems that are programmed to implement methods of the disclosure. **FIG. 4** shows a computer system 401 that is programmed or otherwise configured to process a Raman spectroscopy profile of a subject using a trained machine learning algorithm or against a reference, and determine a cancer assessment (e.g., detect a presence or an absence of cancer) of the subject. The computer system 401 can regulate various aspects of the present disclosure, such as, for example, processing a Raman spectroscopy profile of a subject using a trained machine learning algorithm or against a reference, and determining a cancer assessment (e.g., detect a presence or an absence of cancer) of the subject. The computer system 401 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.

[0262] The computer system 401 includes a central processing unit (CPU, also "processor" and "computer processor" herein) 405, which can be a single core or multi core processor, or a plurality of processors for parallel processing. The computer system 401 also includes memory or memory location 410 (e.g., random-access memory, read-only memory, flash memory), electronic storage unit 415 (e.g., hard disk), communication interface 420 (e.g., network adapter) for communicating with one or more other systems, and peripheral devices 425, such as cache, other memory, data storage and/or electronic display adapters. The memory 410, storage unit

415, interface 420 and peripheral devices 425 are in communication with the CPU 405 through a communication bus (solid lines), such as a motherboard. The storage unit 415 can be a data storage unit (or data repository) for storing data. The computer system 401 can be operatively coupled to a computer network ("network") 430 with the aid of the communication interface 420. The network 430 can be the Internet, an internet and/or extranet, or an intranet and/or extranet that is in communication with the Internet. The network 430 in some cases is a telecommunication and/or data network. The network 430 can include one or more computer servers, which can enable distributed computing, such as cloud computing. The network 430, in some cases with the aid of the computer system 401, can implement a peer-to-peer network, which may enable devices coupled to the computer system 401 to behave as a client or a server.

[0263] The CPU 405 can execute a sequence of machine-readable instructions, which can be embodied in a program or software. The instructions may be stored in a memory location, such as the memory 410. The instructions can be directed to the CPU 405, which can subsequently program or otherwise configure the CPU 405 to implement methods of the present disclosure. Examples of operations performed by the CPU 405 can include fetch, decode, execute, and writeback.

[0264] The CPU 405 can be part of a circuit, such as an integrated circuit. One or more other components of the system 401 can be included in the circuit. In some cases, the circuit is an application specific integrated circuit (ASIC).

[0265] The storage unit 415 can store files, such as drivers, libraries and saved programs. The storage unit 415 can store user data, e.g., user preferences and user programs. The computer system 401 in some cases can include one or more additional data storage units that are external to the computer system 401, such as located on a remote server that is in communication with the computer system 401 through an intranet or the Internet.

[0266] The computer system 401 can communicate with one or more remote computer systems through the network 430. For instance, the computer system 401 can communicate with a remote computer system of a user. 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 401 via the network 430.

[0267] 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 401, such as, for example, on the memory 410 or electronic storage unit 415. 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 405. In some cases, the code can be retrieved from the storage unit 415

and stored on the memory 410 for ready access by the processor 405. In some situations, the electronic storage unit 415 can be precluded, and machine-executable instructions are stored on memory 410.

[0268] The code can be pre-compiled and configured for use with a machine having a processer adapted to execute the code, or can be compiled during runtime. The code can be supplied in a programming language that can be selected to enable the code to execute in a pre-compiled or as-compiled fashion.

[0269] Aspects of the systems and methods provided herein, such as the computer system 401, can be embodied in programming. Various aspects of the technology may be thought of as "products" or "articles of manufacture" typically in the form of machine (or processor) executable code and/or associated data that is carried on or embodied in a type of machine readable medium. Machine-executable code can be stored on an electronic storage unit, such 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 may provide non-transitory storage at any time for the software programming. All or portions of the software may at times be communicated through the Internet or various other telecommunication networks. Such communications, for example, may enable loading of the software from one computer or processor into another, for example, from a management server or host computer into the computer platform of an application server. Thus, another type of media that may bear the software elements includes optical, electrical and electromagnetic waves, such as used across physical interfaces between local devices, through wired and optical landline networks and over various air-links. The physical elements that carry such waves, such as wired or wireless links, optical links or the like, also may be considered as media bearing the software. As used herein, unless restricted to non-transitory, tangible "storage" media, terms such as computer or machine "readable medium" refer to any medium that participates in providing instructions to a processor for execution.

[0270] Hence, a machine readable medium, such as computer-executable code, may take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or

electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution.

[0271] The computer system 401 can include or be in communication with an electronic display 435 that comprises a user interface (UI) 440 for providing, for example, Raman spectroscopy profiles of a subject. Examples of UI's include, without limitation, a graphical user interface (GUI) and web-based user interface.

[0272] 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 405. The algorithm can, for example, process a Raman spectroscopy profile of a subject using a trained machine learning algorithm or against a reference, and determine a cancer assessment (e.g., detect a presence or an absence of cancer) of the subject.

EXAMPLES

[0273] **Example 1: Raman spectroscopy analysis for pan-cancer detection**

[0274] Using methods and system of the present disclosure, an advanced, rapid, time-sensitive Raman Spectroscopy approach was used to differentiate cancer from non-cancer subjects by using a combination of outcomes from biofluids such as DNA, RNA, enriched stem cells etc. derived from blood. These methods and systems may encompass classification of cancer survivors, benign cases and patients in remission in different bins. These methods and systems may extend to organ-specific cancer clusters to predict tissue of origin of blinded samples using exome and transcriptome analysis. Finally, DNA methylation signatures from RS can potentially differentiate pluripotency vs. cancer stem cell initiation features in normal vs. diseased states respectively.

[0275] Cancer is a serious and complex disease that has profound effects on the patients and their loved ones. Cancer refers to the abnormal and uncontrolled growth and proliferation of cells, which may eventually lead to death. The exact cause for this deadly death is largely unknown, however, early detection and intervention can significantly improve the lives of suspected patients.

[0276] As per global estimates, by Pan American Health Organisation (PAHO), there were 10 million cancer-related deaths and 20 million new cases of cancer worldwide, respectively. Over the following two decades, the cancer burden will roughly double, placing additional strain on communities, individuals, and health systems. By 2040, it is anticipated that there will be roughly 30 million additional instances of cancer worldwide, with the largest increases occurring in low- and middle-income nations. Thus, it becomes necessary to detect cancer at the earliest stage.[1]

[0277] Current methods for diagnosing cancers may involve testing for the presence of cancerous tumors, components and byproducts of these tumors. They include tissue biopsy, liquid biopsies, circulating tumor cells (CTCs), cell-free DNA (cfDNA), cell-free RNA (cfRNA), circulating tumor DNA (ctDNA), and extracellular vesicles (EVs).

[0278] Tissue biopsy may be a vital diagnostic procedure for identifying and verifying the existence of cancer is a tumor biopsy. To check for the presence of cancer cells, a small sample of tissue is taken from a suspected location of the body and inspected under a microscope and put through several laboratory tests. Since it's an invasive test, there are risks that follows this procedure that includes infections, bleeding, pain and discomfort [2] as well as spread of cancer cells to adjacent tissues [3].

[0279] Liquid biopsies may prove to be an alternative, less invasive method that includes the detection of extracellular vesicles (EVs), circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), cell-free DNA (cfDNA) and cell-free RNA (cfRNA). [4]

[0280] CTCs may refer to cancer cells that have separated from the original tumor and infiltrated the bloodstream are known as circulating tumor cells. The results of a CTC study can provide details on the tumor's heterogeneity and metastatic potential. [4]

[0281] DNA fragments that are released into the bloodstream by cells, particularly cancer cells, may be referred to as "circulating free DNA" or "cell-free DNA". As they develop and deteriorate, tumors release DNA into the blood. The presence and genetic make-up of the tumor can be determined by examining cfDNA for certain mutations, changes, or abnormal methylation patterns. [4]

[0282] Circulating free RNA or "cell-free RNA" may refer to RNA fragments that are released into the bloodstream by cells, particularly cancer cells, like circulating free DNA. One can find specific gene expression patterns linked to cancer through cfRNA analysis. [4]

[0283] ctDNA may refer to DNA fragments that tumor cells release into the bloodstream. It includes the genetic changes and mutations found in the tumor. ctDNA analysis can help with non-invasive cancer diagnosis, tracking of treatment response, identifying genetic changes that lead to therapy resistance, and monitoring of tumor progression over time. [4]

[0284] Small membrane-bound particles called extracellular vesicles, or exosomes and microvesicles, are released by cells, including cancer cells. They contain lipids, proteins, and nucleic acids (DNA, RNA). Examining the contents of EVs can reveal details about the tumor microenvironment, interactions between the tumor and the host, and possible biomarkers. [4]

[0285] Overall, liquid biopsies may come with separate set of challenges that include problems with standardization, sensitivity and specificity. Most importantly, it is necessary to note that these tests are carried out once the tumor has been formed.

[0286] The potential of Very Small Embryonic-Like stem cells (VSELs) in cancer detection is very beneficial. VSELs are a subpopulation of stem cells that have gained interest since they are thought to possess properties of the embryo and can be found in adult tissues. These cells are thought to dwell in a variety of organs, including bone marrow and peripheral blood [5].

[0287] VSELs may be characterized by the following properties that make them a suitable target for cancer detection.

[0288] Pluripotent Nature: VSELs may have pluripotent properties that allow them to differentiate into diverse cell lineages, much like embryonic stem cells. This property may be attributed to their potential to form cancer cells. Furthermore, activation of VSELs may aid in the development of tumor-initiating cells or cancer stem cells in a variety of malignancies.

[0289] Small size: VSELs may have a small size, which ranges from 3 to 6 micrometers. Their identification and isolation may be made more difficult by this size difference.

[0290] Quiescent Nature: VSELs may be quiescent in tissues until they are activated by trauma or other stimuli. This quiescent condition may support the long-term preservation of their pluripotent capacity. [5-7]

[0291] Raman spectroscopy is an analytical technique that measure vibration energy between molecules to provide chemical and structural information. It is named after Indian physicist C. V. Raman. A linear molecule has 3N-5 vibrations where N is the number of atoms. The total number of possible vibrations is 3N-6. For a vibration to be Raman-active, there may be a change in the polarizability of the molecule. The symmetrical stretching of a triatomic molecule, such as CO2, leads to polarization; thus, the molecule is Raman-active and shows strong Raman scattering. **FIG. 5** shows molecular vibrations modes relevant to Raman scattering.

[0292] Isolating and identifying these VSELs may be suitable for detecting cancer even before the tumor formation has been initiated. Cancer specific signatures can be detected, thus enabling early detection and further necessary intervention.

[0293] Time-gated Raman Spectroscopy (TGRS) may use picosecond pulsed laser excitation source which reduces the background signal to enhance the Raman Spectra. It reduces the sample induced fluorescence and phosphorescence. Conventional Raman may use a continuous wave

laser, which has low signal-to-noise ratio making it difficult to identify the weak signal. Additionally, Time-gate Raman uses Picosecond laser excitation which reduces the thermal emission interference from real signal. TGRS uses a 532-nm laser. Raman scattering at 532nm is 4.7 times more efficient than 785 nm laser and 16 times better than a 1064-nm laser. The scan time at 532 nm is much lesser than that of 785 nm or 1064 nm.

[0294] **FIG. 6** shows Raman scattering in TGRS vs. conventional Raman spectroscopy.

[0295] Fourier Transform Infra-red spectroscopy (FTIR) may be used for cancer detection and analysis. TGRS exhibits many advantages over IR spectroscopy. TGRS uses Raman shift due to scattering of light, while IR spectroscopy (IR) uses absorption of light by vibrating molecules. In TGRS, water or aqueous solution can be used as it does not absorb the light. In contrast, water cannot be used due to its intense absorption in IR spectroscopy. Raman spectroscopy provides information about the covalent character of the molecule, while IR spectroscopy reads the ionic characters of the molecules. Sample preparation in TGRS is simple, whereby a sample can be loaded onto glass slide. In contrast, in IR spectroscopy, sample preparation is elaborative, such that glass cannot be used as substrate, and salts of Na, K, Ag and Ca are used. TGRS can be used for qualitative as well as quantitative analysis, while IR spectroscopy may be used for qualitative analysis with limited quantitative analysis. Spectra in TGRS is simple compared to IR spectroscopy. Inorganic salts can be measured in TGRS compared IR spectroscopy. TGRS is suitable for biological specimen, while in IR spectroscopy water can interfere with analysis of biological system.

[0296] Raman analysis may be used to detect cancer in biological samples. Raman techniques are sensitive enough to measure histological signatures in cancer samples. Raman techniques measure tissue in a native stage, thereby offering advantages over biopsies. For example, a fully automated Raman system for skin cancer has been approved for the clinical detection of skin cancer. A spectral band around 1445 cm^{-1} and 1655 cm^{-1} exhibited the most prominent changes in lung cancer tissue. The specificity and sensitivity for detection of lung cancer is more than 90%. Further, in breast cancer, a sensitivity and specificity of 100% was observed in discriminating between normal and cancerous tissue.

[0297] A biological sample may have lower Raman scattering. Therefore, TGRS may be used to achieve a high signal-to-noise ratio for biological specimens. For example, TGRS and RS comparisons for detecting signals from animal tissue (adipose and muscle) show that TGRS is effective in signal detection compared to RS.

[0298] Exposure to various exogeneous and endogenous stimulus leads to DNA damage. If these DNA damages are not repaired properly, this may lead to genetic alterations in terms of single base substitution, small insertion deletions, genome rearrangements, chromosome copy number

changes. Such sequence alterations are abundant in the cancer genome. Computational methods may be used in next generation sequencing data (NGS) to capture such alterations by employing machine learning methods. However, at an early stage of cancer, detection of these alterations is challenging through a non-invasive technique. Hence, there remains an urgent need for a robust tool that can identify genetic alterations at its initial stage.

[0299] Various cancer types develop specific signature pattern based on 6 classes of single base substitutions (C>A, C>G, C>T, T>A, T>C, and T>G) [8]. Since cancer may initiate at the stem cell level, stem/progenitor cells may be enriched from blood, from which genomic DNA may be extracted. Genomic DNA may be subjected to analysis by time-gated Raman Spectroscopy to identify base substitution specific pattern. Calcium fluoride (CaF2) substrate may be used to measure Raman spectra by using Confocal Raman spectroscopy with 532 laser excitations. Spectra may be recorded at 100x objective. To address inter-sample variability, spectra may be recorded from multiple areas (minimum 10). The best 5 correlated spectra may be taken for analysis. Multiple data points may be selected across each 5 spectra. Data acquisition may be carried out in a wide spectral range from 0 – 2500 cm^{-1}, while analysis may be performed on a biological footprint region of 600 to 1800 cm^{-1}. Raman data may be pre-processed using the Spectragryph software and univariate/multivariate analysis (PCA, PC-LDA, KNN, random forest, etc.) may be done by Unscrambler. Further, cancer signatures may be identified in patients as a separate cluster as compared to non-cancers. Once new samples are obtained, classification of new samples may be carried out by unsupervised and/or supervised machine learning approach to predict samples as cancer vs non-cancer.

[0300] RNA is easily degradable, hence, its half-life is important for conservation purposes. Enriched VSELs from blood may be subjected to RNA isolation for base transition effects. These transitional rearrangements may be monitored to distinguish cancer from non-cancer.

[0301] Raman spectroscopy (Method 100) that has been complemented with data from processes described by Method 200 may improve diagnostic accuracy significantly (Method 300). Method 300 may be more than 20% more accurate than Method 100 or Method 200. Thus, the combinatorial approach can not only improve sensitivity and specificity, but also accuracy. This is because multi-pronged, multi-omics approaches may improve testing features.

[0302] References

[0303] [1] www.paho.org/en/campaigns/world-cancer-day-2023-close-care-gap is incorporated by reference herein in its entirety.

[0304] [2] Brown, A. P., Wendler, D. S., Camphausen, K. A., Miller, F. G., & Citrin, D. (2008). Performing nondiagnostic research biopsies in irradiated tissue: a review of scientific, clinical, and ethical considerations. Journal of clinical oncology : official journal of the American Society

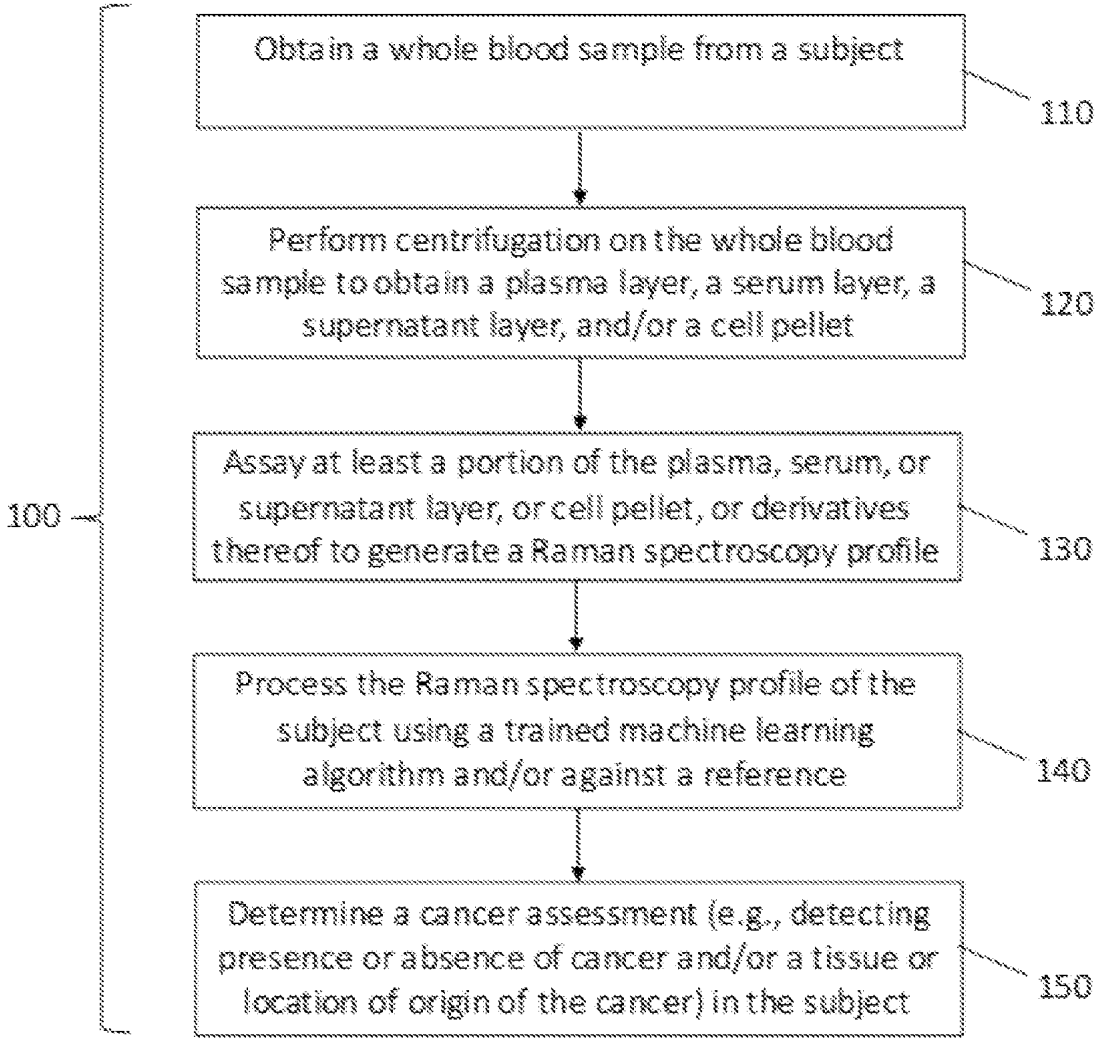

Obtain a whole blood sample from a subject
110
Perform centrifugation on the whole blood sample to obtain a plasma layer, a serum layer, a supernatant layer, and/or a cell pellet
120
Assay at least a portion of the plasma, serum, or supernatant layer, or cell pellet, or derivatives thereof to generate a Raman spectroscopy profile
130
Process the Raman spectroscopy profile of the subject using a trained machine learning algorithm and/or against a reference
140
Determine a cancer assessment (e.g., detecting presence or absence of cancer and/or a tissue or location of origin of the cancer) in the subject
150
100

FIG. 1

biomolecules. Journal of Biophotonics, 2019, 12(12): e201960048. doi: 10.1002/jbio.201960048 is incorporated by reference herein in its entirety.

[0314] [12] Jadhav PA, Pansare K, Hole A, Ingle A, Govekar R, Murali Krishna C. Serum Raman Spectroscopy in experimental carcinogenesis hamster buccal pouch model: exploring early diagnostic capability. Journal of Raman Spectroscopy. 2022, doi: 10.1002/jrs.6450 is incorporated by reference herein in its entirety.

[0315] [13] Hole A, Jadhav P, Pansare K, Noothalapati H, Deshmukh A, Gota V, Chaturvedi P, Krishna CM. Saliva Raman Spectroscopy: Understanding Alterations in Saliva of Tobacco Habitués and Oral Cancer Subjects. Vibrational Spectroscopy, 2022, 103414. doi: 10.1016/j.vibspec.2022.103414 is incorporated by reference herein in its entirety.

[0316] **Example 2: Time-Gated Raman Spectroscopy Analysis**

[0317] Hematological specimen collection can be performed as follows. Peripheral blood samples were aseptically drawn into EDTA vacutainers and subsequently stored at a temperature of 4C. Then, a 3 mL blood sample was withdrawn and diluted with equivalent amount of DPBS. The sample was subjected to Ficoll-Hypaque density gradient centrifugation at 1000g for 15 minutes. The sample stratified into four distinct layers, from which the plasma layer was meticulously extracted and cryopreserved at a temperature of -80C.

[0318] Raman spectral acquisition was performed as follows. Prior to processing, the plasma samples were thawed on ice. The samples were briefly vortexed for 2-3 seconds. A 20 µL plasma sample was loaded onto a Calcium Fluoride (CaF2) substrate. The samples were allowed to semi-dry over a period of 5 minutes. The prepared sample-loaded substrate was placed onto the microscope's stage. Samples were focused using 10X and 100X objective. Spectra was acquired using defined parameters.

[0319] Raman spectral analysis was performed, including spectra pre-processing, multivariate data analysis, and an Tzar Labs (TL) algorithm.

[0320] Spectra pre-processing was performed by using Spectragryph software for pre-processing of raw data. Parameters encompassing spectral interpolation, spectral smoothing, and baseline correction were configured. Post-processing treatment, the data was saved in SPC (Spectrum) format.

[0321] Multivariate data analysis was performed by using Unscrambler software. The samples were subjected to unsupervised Principal Component Analysis (PCA) followed by supervised Principal Component Based Linear Discriminant Analysis (PC-LDA) tests.

[0322] The TL algorithm was performed including a 5 x 6 analysis approach and fold change calculations based on 2 analyses, the first based on the intensity of a set of 16 features, and the second based on the intensity of a set of 8 features.

[0323] In particular, the method involves plasma collection following differential centrifugation at 1000g, instead of 400g. As shown in **FIG. 7**, the spectra shows a comparison between plasma sample obtained following centrifugation at 400g and 1000g.

[0324] Noticeable distinctions are evident when comparing the plasma samples obtained through a centrifugation protocol at 400g and an alternative approach at 1000g. Particularly noteworthy is the heightened intensity of the spectral range between 1000 cm^{-1} and 1400 cm^{-1} observed in the plasma derived from the 1000g centrifugation, surpassing that of the 400g protocol. Furthermore, the plasma acquired at 1000g exhibits additional distinctive features at (1098 cm^{-1}, 1177 cm^{-1}, 1269 cm^{-1}, 1319 cm^{-1}, 1342 cm^{-1}, 1411 cm^{-1}, 1510 cm^{-1} and 1523cm^{-1}) enhancing its overall characterization.

[0325] Further, plasma is derived using a Ficoll-Hypaque density gradient technique, which is distinct from other serum-based methods.

[0326] Further, a calcium fluoride substrate is employed for plasma sample loading, enhancing spectral quality and reducing background.

[0327] Further, the utilization of the Pico Raman Timegate instrument imparts unique analytical capabilities, including fluorescence suppression.

[0328] Raman scattering may be practically immediate, and it may be scattered only during the time period when laser pulses hit the sample. This is not the case with the other light emitting phenomena. For example, fluorescence is emitted with a delay, and it lasts for a long time after a laser pulse (as shown in **FIG. 8**, in which Raman scattering occurs almost instantly whereas fluorescence emission is spread over a time range determined by its lifetime). Time-gated technology enables measurement during Raman scattering and shutting down collection of light for the rest of the time. This means that with PicoRaman, it is possible to collect signal from the moments when Raman scattering is stronger than any other emitting phenomena.

[0329] During time-gated measurements the detector is only active for extremely short periods of time. A pulsed laser which is used in time-resolved Raman has a very high instantaneous pulse energy and the instantaneous Raman scattering "burst" intensities can be quite high compared with conventional Raman scattering. These features lead to collection of high amounts of Raman scattering in short time periods e.g., the collected Raman scattering to background emission ratio is high. Because of this, time-gated technology enables measurements in illuminated environments (no need for dark sample enclosures) and measurements of thermally emitting hot samples.

[0330] Further, a 5 x 6 analysis approach is introduced for distinguishing between cancerous and non-cancerous samples. The systematic approach employed to analyze the spectral data using 5 x 6 approach.

[0331] A comparison of a screening sample is performed using 7 different groups, namely Control, Cancer Treatment Naïve, Control Indeterminant, Cancer Indeterminant, Cancer on Treatment, Cancer Survivor, and Benign. A database is generated for each group. First, the screening sample is labeled as control, and it is tested against the above mentioned 7 groups - Control, Cancer Treatment Naïve, Control Indeterminant, Cancer Indeterminant, Cancer on Treatment, Cancer Survivor and Benign groups. Then the pre-processed spectra are subjected to multivariate analysis, including PCA and LDA. The groups into which the screening sample classified are noted.

[0332] The above operation is repeated by labelling the screening sample as Cancer and testing it against the above-mentioned 7 groups - Control, Cancer Treatment Naïve, Control Indeterminant, Cancer Indeterminant, Cancer on Treatment, Cancer Survivor and Benign groups. Then the pre-processed spectra are subjected to multivariate analysis, including PCA and LDA. The groups into which the screening sample classified are noted.

[0333] The above operations may produce 6 possible results. The group in which the screening sample classified 5 times correctly is then considered as the appropriate classification (5 out of 6).

[0334] Fold change calculations may be performed based on one or more analyses, the first analysing the intensity of 16 features for distinguishing cancerous and non-cancerous samples, and the second analysing the intensity of 8 features for distinguishing cancerous and non-cancerous samples. The systematic approach employed to analyze the spectral data identified 16 characteristic features outlined as follows.

[0335] A comparison of control and cancer spectra is performed. The initial operation involved a comprehensive comparison of the spectra obtained from control and cancer samples. Within this comparative analysis, distinct upshifts and downshifts in spectral features were consistently observed across all cancer samples.

[0336] The uniformity of shifts was evaluated. The observed spectral shifts were found to be consistent among the various cancer samples, indicating a shared underlying phenomenon contributing to these shifts.

[0337] The prevalence of certain features was analyzed. Among the observed shifts, specific spectral features emerged as being particularly prevalent across the spectrum of cancer samples. These prevalent features exhibited similar patterns of shift across the samples.

[0338] Normalization of intensity was performed. In order to quantitatively assess the variations in intensity, a normalization procedure was applied to the spectral data.

[0339] Calculation of intensity was performed. Post normalization, the intensity values of the spectral peaks corresponding to the identified features were calculated for both the cancer and control samples.

[0340] Averaging intensity was performed. The average intensity values for each specific wavenumber (cm^{-1}) were determined for both the cancer and control samples.

[0341] Fold change calculations may be performed. Fold change can be determined using the ratio of the average intensity of the cancer sample to the average intensity of the control sample for each specific wavenumber.

[0342] As shown in Table 1, the following 16 features were identified to be commonly altered in cancers, singularly or in combination.

[0343] Table 1

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1083- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp
1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1158- Beta-carotene
1518-Beta carotene

[0344] **FIG. 9** shows a representative Raman spectra of control and cancer plasma samples.

[0345] A serum analysis protocol was performed. Blood collection was performed, including collecting 3 ml of blood in vacutainer tube. The collected blood in the vacutainer was allowed to clot by leaving it undisturbed at room temperature for 1 hour. The clot-formed vacutainer was placed in a centrifuge and spun at 1000g for 10 minutes at room temperature and 9 acceleration

and 0 deceleration. Serum collection was performed after centrifugation, by carefully collecting the separated serum using a sterile pipette and storing at -80C.

[0346] FIG. 10 shows a comparison of plasma and serum Raman spectra.

[0347] Noticeable distinctions are evident when comparing the plasma and serum samples obtained through centrifugation at 1000g. Particularly noteworthy is the heightened intensity of the spectral range observed in the serum, surpassing that of the plasma with noticeable distinctive characteristic feature at 676 cm^{-1}, 697 cm^{-1}, 714 cm^{-1}, 1069 cm^{-1}, and 1585 cm^{-1}.

[0348] An intensity calculation method was performed as follows. This can be performed irrespective of what method is used. The 16 unique features were identified by comparing Control and Cancer Treatment Naïve spectra. These included 622 cm^{-1}, 645 cm^{-1}, 758 cm^{-1}, 880 cm^{-1}, 943 cm^{-1}, 1006 cm^{-1}, 1034 cm^{-1}, 1087 cm^{-1}, 1159 cm^{-1}, 1344 cm^{-1}, 1406 cm^{-1}, 1453 cm^{-1}, 1518 cm^{-1}, 1558 cm^{-1}, 1659 cm^{-1}, 1736 cm^{-1}. The above identified 16 features (wavenumbers) were used for intensity calculation. An internal database comprised 6 different categories: Control samples (n=129), Control Indeterminate (n=83), Cancer Indeterminate (n=11), Cancer Treatment Naïve (n=79), Cancer On Treatment (n=106), and Cancer Survivors (n=21).

[0349] Average and standard deviations were generated for each category and multiplied by 10000.

[0350] A range was created by adding and subtracting the standard deviation from the average for each category.

[0351] Control samples (Table 2) may have HrC scores of 0-2. These may be obtained from subjects without cancer, and who have never had cancer (note that cancer survivors may not be used as controls).

[0352] Table 2. Control samples:

Control Average	Standard deviation	Control Average *10000	Standard deviation *10000	Control Range	
0.00104	0.000129	10.35	1.29	9.06	11.65
0.00153	0.000181	15.27	1.81	13.46	17.08
0.00205	0.000185	20.49	1.85	18.64	22.35
0.00111	0.000238	11.05	2.38	8.67	13.43
0.00151	0.000278	15.11	2.78	12.33	17.89
0.00535	0.000981	53.54	9.81	43.72	63.35
0.00098	0.000216	9.83	2.16	7.67	11.99
0.00078	0.000142	7.84	1.42	6.42	9.25
0.00188	0.000729	18.78	7.29	11.50	26.07
0.00208	0.000486	20.84	4.86	15.98	25.70
0.00049	0.000199	4.87	1.99	2.88	6.86
0.00434	0.000491	43.45	4.91	38.53	48.36

0.00114	0.000555	11.40	5.55	5.85	16.94
0.00038	0.000139	3.81	1.39	2.42	5.20
0.00401	0.000426	40.15	4.26	35.88	44.41
0.00027	0.000066	2.71	0.66	2.05	3.36

[0353] Control indeterminate samples (Table 3) may be indeterminate for cancer, and may have an HrC score of 2-6. They may be largely determined as groups by their HrC score, PET, and/or histopath report. Control indeterminate may also describe subjects without cancer but who have some other pathology apart from cancer.

[0354] Table 3. Control Indeterminate (CI) samples:

CI Average	Standard deviation	CI average *10000	Standard deviation *10000	**CI Range**	
0.000984	0.000095	9.84	0.95	8.89	10.79
0.001454	0.000119	14.54	1.19	13.35	15.73
0.002173	0.000184	21.73	1.84	19.89	23.56
0.001152	0.000234	11.52	2.34	9.18	13.86
0.001640	0.000244	16.40	2.44	13.95	18.84
0.005756	0.000442	57.56	4.42	53.13	61.98
0.001109	0.000150	11.09	1.50	9.59	12.59
0.000912	0.000105	9.12	1.05	8.07	10.18
0.001160	0.000281	11.60	2.81	8.79	14.41
0.002408	0.000275	24.08	2.75	21.33	26.82
0.000415	0.000088	4.15	0.88	3.27	5.02
0.004707	0.000251	47.07	2.51	44.57	49.58
0.000483	0.000231	4.83	2.31	2.52	7.14
0.000548	0.000131	5.48	1.31	4.17	6.79
0.004069	0.000221	40.69	2.21	38.49	42.90
0.000213	0.000041	2.13	0.41	1.72	2.54

[0355] Cancer indeterminate samples are misclassified in Raman but are PET/histopath confirmed. They are cancer treatment naïve samples and may have an HrC score of 6-10.

[0356] Cancer samples may be obtained from a subject with cancer who is also treatment naïve, and may have an HrC score of 10 or more.

[0357] Table 4. Cancer Treatment Naïve Indeterminate (CTNI) samples:

CTN IN Average	Standard deviation	CTN IN average *10000	Standard deviation *10000	CTNI Range	
0.00098	0.00019	9.80	1.87	7.93	11.67
0.00150	0.00025	15.03	2.51	12.52	17.54
0.00216	0.00032	21.55	3.24	18.31	24.79
0.00108	0.00028	10.79	2.78	8.01	13.57
0.00144	0.00033	14.44	3.29	11.15	17.73
0.00500	0.00108	50.01	10.78	39.22	60.79
0.00089	0.00026	8.91	2.60	6.30	11.51
0.00080	0.00014	8.00	1.35	6.65	9.35
0.00138	0.00030	13.82	2.96	10.86	16.78
0.00200	0.00055	20.04	5.47	14.57	25.51
0.00057	0.00022	5.72	2.19	3.54	7.91
0.00422	0.00046	42.22	4.56	37.66	46.78
0.00087	0.00039	8.69	3.92	4.78	12.61
0.00052	0.00041	5.20	4.14	1.05	9.34
0.00388	0.00054	38.77	5.37	33.40	44.14
0.00026	0.00005	2.62	0.48	2.14	3.10

[0358] Table 5. Cancer Treatment Naïve (CTN) samples:

CTN Average	CTN standard deviation	CTN average *10000	CTN standard deviation *10000	CTN Range	
0.00096	0.00009	9.575201	0.861254	8.71	10.44
0.00143	0.00011	14.33539	1.132392	13.2	15.47
0.00223	0.00020	22.2733	2.011855	20.26	24.29
0.00118	0.00023	11.81562	2.347598	9.47	14.16
0.00149	0.00027	14.88812	2.656031	12.23	17.54
0.00562	0.00052	56.21781	5.200737	51.02	61.42
0.00115	0.00016	11.50449	1.554829	9.95	13.06
0.00099	0.00010	9.94019	1.033676	8.91	10.97
0.00095	0.00023	9.540205	2.312563	7.23	11.85
0.00235	0.00026	23.45817	2.622963	20.84	26.08
0.00042	0.00010	4.182009	0.998962	3.18	5.18
0.00463	0.00022	46.25243	2.1677	44.08	48.42
0.00036	0.00020	3.575872	2.019647	1.56	5.6
0.00062	0.00015	6.162456	1.502312	4.66	7.66
0.00410	0.00026	40.98878	2.582615	38.41	43.57
0.00024	0.00005	2.360053	0.469366	1.89	2.83

[0359] Cancer on treatment samples (Table 6) may be obtained from subjects who are undergoing treatment or are within 6 months of their last treatment, and may have variable HrC scores (e.g., in the 6-10 range).

[0360] Table 6. Cancer On Treatment (COT) samples:

COT Average	COT standard deviation	COT Average *10000	COT standard deviation *10000	COT Range	
0.000954	7.45E-05	9.541767	0.745237	8.8	10.29
0.00141	0.000103	14.10264	1.025099	13.08	15.13
0.002225	0.000188	22.25269	1.879525	20.37	24.13
0.001184	0.000231	11.83978	2.307279	9.53	14.15
0.00158	0.000235	15.79803	2.35209	13.45	18.15
0.005694	0.000432	56.94144	4.316252	52.63	61.26
0.001115	0.000149	11.15106	1.49022	9.66	12.64
0.000955	0.000101	9.552773	1.006002	8.55	10.56
0.001081	0.000292	10.80846	2.922724	7.89	13.73
0.002376	0.000235	23.76278	2.345393	21.42	26.11
0.000408	8.13E-05	4.077328	0.812544	3.26	4.89
0.004705	0.000218	47.05087	2.182159	44.87	49.23
0.00043	0.000218	4.296656	2.177946	2.12	6.47
0.000596	0.000136	5.961388	1.355816	4.61	7.32
0.00405	0.000219	40.49954	2.190079	38.31	42.69
0.000218	3.03E-05	2.17929	0.303184	1.88	2.48

[0361] Cancer survivor samples (Table 7) may be obtained from a subject who had cancer, and may have a variable HrC score that does not return to the 0-2 range (e.g., which may be in the 2-6 range). These may include subjects who have completed cancer treatment more than 6 months prior, and are PET negative.

[0362] Table 7. Cancer Survivor (CS) samples:

CS Average	CS Standard deviation	CS Average *10000	CS Standard deviation *10000	CS Range	
0.000984	9.8E-05	9.84	0.98	8.86	10.82
0.001488	0.000165	14.88	1.65	13.22	16.53
0.002154	0.000214	21.54	2.14	19.40	23.68
0.001348	0.000272	13.48	2.72	10.77	16.20
0.001444	0.000262	14.44	2.62	11.83	17.06
0.005203	0.00112	52.03	11.20	40.83	63.23
0.000997	0.000285	9.97	2.85	7.12	12.82
0.000886	0.000126	8.86	1.26	7.61	10.12

0.001369	0.00039	13.69	3.90	9.79	17.59
0.002103	0.000482	21.03	4.82	16.21	25.85
0.000536	0.000248	5.36	2.48	2.89	7.84
0.00446	0.000463	44.60	4.63	39.98	49.23
0.000753	0.000544	7.53	5.44	2.09	12.97
0.000534	0.000208	5.34	2.08	3.26	7.42
0.003733	0.000489	37.33	4.89	32.44	42.22
0.000206	3.84E-05	2.06	0.38	1.67	2.44

[0363] As shown in Table 8, test samples were compared against each of the following generated categories.

[0364] Table 8.

Wavenumber (-nm)	Control Range		CI Range		CTNI Range		CTN Range		COT Range		CS Range	
622.0056	9.06	11.65	8.89	10.79	7.93	11.67	8.71	10.44	8.8	10.29	8.86	10.82
645.8758	13.46	17.08	13.35	15.73	12.52	17.54	13.2	15.47	13.08	15.13	13.22	16.53
758.0658	18.64	22.35	19.89	23.56	18.31	24.79	20.26	24.29	20.37	24.13	19.4	23.68
880.9973	8.67	13.43	9.18	13.86	8.01	13.57	9.47	14.16	9.53	14.15	10.77	16.2
943.0598	12.33	17.89	13.95	18.84	11.15	17.73	12.23	17.54	13.45	18.15	11.83	17.06
1006.316	43.72	63.35	53.13	61.98	39.22	60.79	51.02	61.42	52.63	61.26	40.83	63.23
1034.96	7.67	11.99	9.59	12.59	6.3	11.51	9.95	13.06	9.66	12.64	7.12	12.82
1087.475	6.42	9.25	8.07	10.18	6.65	9.35	8.91	10.97	8.55	10.56	7.61	10.12
1159.085	11.5	26.07	8.79	14.41	10.86	16.78	7.23	11.85	7.89	13.73	9.79	17.59
1344.079	15.98	25.7	21.33	26.82	14.57	25.51	20.84	26.08	21.42	26.11	16.21	25.85
1406.142	2.88	6.86	3.27	5.02	3.54	7.91	3.18	5.18	3.26	4.89	2.89	7.84
1453.882	38.53	48.36	44.57	49.58	37.66	46.78	44.08	48.42	44.87	49.23	39.98	49.23
1518.332	5.85	16.94	2.52	7.14	4.78	12.61	1.56	5.6	2.12	6.47	2.09	12.97
1558.911	2.42	5.2	4.17	6.79	1.05	9.34	4.66	7.66	4.61	7.32	3.26	7.42
1659.166	35.88	44.41	38.49	42.9	33.4	44.14	38.41	43.57	38.31	42.69	32.44	42.22
1736.744	2.05	3.36	1.72	2.54	2.14	3.1	1.89	2.83	1.88	2.48	1.67	2.44

[0365] The in-range and out-of-range features of the test sample were compared, when compared with each group. The category was selected with the highest number of in-range features for the test samples. This is the group which matches most with our test sample. This identified group was used to classify the test samples based on their 16 features intensity values.

[0366] **Example 3: Molecular Profile Analysis (Raman complemented by qPCR)**

[0367] Using methods and systems of the present disclosure, Raman spectroscopy analysis (Method 100) is performed on a sample (e.g., plasma, serum, supernatant, cell pellet or nucleic acid of the sample) of a subject, and an HrC score is determined. If the HrC score determined at this stage is 0-2, an absence of cancer is determined for the test results.

[0368] If the Raman results are over an HrC score of 2, then a risk assessment is performed, and a second procedure (Method 200) is conducted on the pellet of the sample. There are several possible methods that can be used here.

[0369] A SYBR green qPCR method may be used for expression analysis of one or more genes in the sample. Alternatively, a TaqMan qPCR method may be used for expression analysis of one or more genes in the sample. Alternatively, a ddPCR method (e.g., using 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 genes) may be used for expression analysis of one or more genes in the sample. The ddPCR method may comprise expression analysis of one or more housekeeping genes in the sample, or may be performed without expression analysis of one or more housekeeping genes in the sample. Alternatively, a dPCR method (e.g., using 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 genes) may be used for expression analysis of one or more genes in the sample. The dPCR method may comprise expression analysis of one or more housekeeping genes in the sample, or may be performed without expression analysis of one or more housekeeping genes in the sample.

[0370] Next, the results of the Raman and PCR analyses are combined to create a Molecular Profile of the subject and from this determine an HrC score for the subject. For example, an HrC score of 0-2 may indicate an absence of cancer in the subject. As another example, an HrC score of 2-6 may indicate cancer low risk/organ inflammation. Note this category contains both subjects with very early cancer and those who have an inflamed organ for another reason, such as they have recently had heart surgery. As another example, an HrC score of 6-10 may indicate moderate risk of cancer. Very few conditions other than early cancer can cause this score. As another example, an HrC score of 10 or more may indicate a high risk of cancer or a presence of cancer in the subject. Note that this interpretation relates to a subject who is cancer naïve. Cancers survivors, those on treatment, and benign lesions may have HrC results that are interpreted differently.

[0371] **Example 4: Molecular Profile Analysis (Raman complemented by sequencing)**

[0372] Using methods and systems of the current disclosure, as per Method 100, Raman spectroscopy analysis is performed on a sample (e.g. plasma, serum, supernatant or cell pellet of the sample) of a subject, and a cancer assessment is determined. Using methods and systems of the current disclosure, as per Method 200, extracted nucleic acids from the cell pellet are used to generate transcriptomic,genetic and/or epigenetic profiles.

[0373] Then Supervised Machine Learning is performed in order to classify the subjects using a three pronged approach. An objective is to determine whether the subject has cancer or not, or has the potential risk of developing cancer.

[0374] This approach performs better than relying on one method, leading to a better combination of accuracy, specificity, and sensitivity.

[0375] Three different sources are utilized for biological data for each sample, including the following-

[0376] 1. Transcriptome – The Read Counts of 59,578 Genes are used in order to calculate their FPKM values. (fragments per kilobase of transcript per million fragments mapped).

[0377] 2. Exome – Three approaches are used, all which comprise using data from 19,395 genes.

a. ALL –All the data available per gene are used, including base counts per genes, number of single base changes, insertions and deletions, mutation rate, etc. (5,81,844 Features)

b. SBS –The single base changes are used. (2,32,740 Features)

c. INDEL- The insertions and deletions are used. (38,790 Features)

The Final Exome prediction comprises performing a best of three Vote of these methods.

[0378] 3. Spectroscopy – Using plasma, 1006 spectral features and their intensities are obtained.

[0379] After obtaining these three forms of data, they are used to train a suite of machine learning models using leave one out methodology and each of them is evaluated based on accuracy, specificity and sensitivity, to determine which is the best for each input. The models include the following –

[0380] • Support Vector Classifiers (SVCs) operate in high-dimensional space, constructing an optimal hyperplane that maximizes the margin between distinct classes. This focuses on margin and sparsity, utilizing only support vectors near the hyperplane, leads to robust and accurate classification, even for complex, non-linear data.

[0381] • K-Nearest Neighbors (KNN) algorithms classify data points by identifying the K closest data points from a labelled training set. Their simplicity and non-parametric nature make them effective for diverse datasets, while the choice of K allows for flexibility in balancing training complexity and classification accuracy.

[0382] • Decision trees (DC) are powerful supervised learning algorithms that model data by recursively partitioning features and creating a tree-like structure. Each internal node represents a feature-based decision, and leaf nodes signify class labels. This intuitive and interpretable approach excels at handling diverse data types and visualizing decision-making processes.

[0383] • Random forests (RF) leverage an ensemble of decision trees, each trained on a subset of features and data, culminating in a robust, majority-vote based classification that excels in handling complex datasets and preventing overfitting.

[0384] • Gaussian Naive Bayes (GB) classifiers assume data within each class follows a Gaussian distribution, and employ Bayes' theorem to estimate posterior class probabilities based on observed features. This efficient and interpretable approach excels in low-dimensional, linearly separable data, but may face challenges with complex, non-linear relationships.

[0385] • Linear Discriminant Analysis (LDA) projects data onto a lower-dimensional subspace that maximizes class separation, enabling efficient and interpretable discrimination between distinct groups.

[0386] • Quadratic Discriminant Analysis (QDA) allows for more flexible decision boundaries that can effectively model non-linear relationships between features and class labels vs LDA.

[0387] • Gaussian Process Classifiers (GPCs) model the relationship between data points as a Gaussian process, essentially capturing the "smoothness" of the underlying data distribution. This probabilistic approach allows for flexible, non-parametric decision boundaries, effectively classifying even complex, non-linear data.

[0388] • Multi-Layer Perceptrons (MLP/NN) leverage interconnected layers of artificial neurons to model complex non-linear relationships between input features and output predictions. This versatile architecture enables them to tackle diverse tasks, from image recognition to text classification, by iteratively refining predictions through hidden layers.

[0389] Accordingly, an optimal model was used each for Transcriptome and Spectroscopy. Each approach in Exome has its own optimal model. For each sample in Exome, each approach was allowed one vote. This means that the ALL, SBS and INDEL approaches have their own vote for each sample. Then the majority vote of these three was used to determine the final Exome Result.

[0390] Similarly, Transcriptome, Final Exome and Spectroscopy were allowed to each have one vote, where the majority decides the final guess for the sample. This was used to vote to determine the final model accuracy.

[0391] Outputs include a prediction for any test samples received. Using the optimal models, one classification is obtained for Transcriptome and Spectroscopy, and three from Exome.

[0392] The same voting process is used to classify the sample as either Cancer, Non-Cancer, or Indeterminate (or a three-way tie).

[0393] **Example 5: Combined feature analysis**

[0394] This combined feature analysis is used as a part of a Raman spectroscopy analysis, such as what is described in Example 2: Time-Gated Raman Spectroscopy Analysis. Using methods and systems of the current disclosure, the following combined 8 unique features were identified by comparing Control and Cancer Treatment Naïve spectra. These included 1052, 1067, 1107,

1130, 1176, and 1210 cm^{-1} (Table 9). These 8 features were used in combination for intensity calculation and to discriminate cancer and non-cancer samples.

[0395] Table 9.

Wavenumber (cm-1)	Features
1052	C-N Stretch, Phospholipid ↑
1067	C-N Stretch, Phospholipid, C-C Stretch ↑
1083	DNA/RNA symmetric stretching because of parallel dichroism ↑
1107	C-N Stretch, C-C stretch ↑
1130	Protein and Lipid: C-N stretch and chain C-C stretch ↑
1158	B- Carotene, Retinal, CH3 rock, C-N Stretch ↓
1176	Tyr/Phe, CH2 wag, CH2 twist ↑
1210	Tyr/Phe ↑

[0396] This comprehensive analysis, involving these 8 features along with 16 individual feature analysis, aids in effectively categorizing samples into various groups based on intensity calculations.

[0397] Example 6: Variation of Molecular Profile Analysis (Raman Complemented by Sequencing)

[0398] Using methods and system of the present disclosure a molecular profile analysis is performed, which incorporates variations to an analysis described in Example 4 herein.

[0399] *Incorporation of chromosome data*

[0400] An addition to the exome analysis is based on 7 features using data from 22 chromosomes. The features include single-base substitutions (SBS), double-base substitutions (DBS), single nucleotide polymorphisms (SNP), ATGC numbers, G-quadruplexes, insertion/deletion (In/del) mutations, and all features. 10 models are developed from each feature, and the best model is selected. The 3 best features are selected with the highest accuracy. The final exome prediction is performed following the best of 3 vote method.

[0401] Additional algorithms, including logistic regression and nu-support vector classification, may be used herein.

[0402] Logistic Regression (LogReg): A statistical model is used for binary classification tasks. The statistical model estimates the probability of a binary outcome using a logistic function and assumes a linear relationship between input variables and a log-odds of the outcome.

[0403] NuSVC (Nu-Support Vector Classification): A variant of SVM (Support Vector Machine) that allows for more flexibility by setting the "nu" parameter, which controls the number of support vectors and margin violations when the standard SVM may be too rigid, offering a balance between margin size and misclassification tolerance.

[0404] Case Study

[0405] *Subject details:* A blood sample was received from a 54-year-old male participant in a registered study. The participant's complete medical history, including his clinical and family history of cancer, was provided. The participant was a diagnosed with a case of glioblastoma.

[0406] *Test performed:*

[0407] Validation of cancer prediction based on RNA and DNA features using various algorithms of AI/ML. Raman Spectroscopy, Whole transcriptome, or a combination thereof, was used, which determines and correlates with the patient's current history indicating cancer is present or absent. Whole exome was also used, which determines and correlates with the patient's current history indicating cancer is present or absent.

[0408] *Methods:* Artificial intelligence and machine learning algorithms were employed to extract characteristics from both cancer and non-cancer subjects. A model was created based on the extracted features from both the groups (e.g., cancer and non-cancer subjects). These features were analyzed in both transcriptome and exome data. Transcriptome analysis focused on: All sequenced genes, Cancer-related pathway genes (PIC), Signaling pathway genes (SIP), and Organ panel genes. Exome analysis included: Single-base substitutions (SBS), Double-base substitutions (DBS), Single nucleotide polymorphisms (SNP), ATGC numbers, G-quadruplexes, Insertion/deletion (In/del) mutations, and All features. Three best features are taken (with highest accuracy of the model) following VOTE method and exome prediction is performed.

[0409] *Whole transcriptome and exome analysis:* Integrated transcriptome and exome AI/ML analyses accurately identified and categorized patients with cancer, aligning with their current disease status. For this analysis best of three method was used to develop the resultant vote from 3 categories transcriptome, exome and Raman Spectroscopy analysis.

[0410] **Example 7: Intensity Calculation Method**

[0411] Using methods and system of the present disclosure, an example of an intensity calculation method is performed using the following operations.

[0412] Operation 1. Perform 16 feature analysis: Example 2 herein described 16 unique features that had been identified by comparing Control and Cancer Treatment Naïve spectra. The 16 features included 622, 645, 758, 880, 943, 1006, 1034, 1087, 1159, 1344, 1406, 1453, 1518, 1558, 1659, and 1736 cm^{-1}. The identified 16 features (wavenumbers) may be used for an intensity calculation.

[0413] Operation 2. Compare to internal database: Example 2 herein described 6 different categories used in the internal database, which were: Control samples (n=129), Control Indeterminate (n=83), Cancer Indeterminate (n=11), Cancer Treatment Naïve (n=79), Cancer On Treatment (n=106), and Cancer Survivors (n=21). In addition to these 6 categories, in some

embodiments, an additional 3 categories may be included. The additional 3 categories may include Benign (n=15), High risk (n=263) and Tobacco consumers (n=64).

[0414] Operation 3. Average and Standard Deviations may be generated for each category and multiplied by 10,000.

[0415] Operation 4. A range may be created by adding and subtracting the standard deviation from the average for each category. Table 10 illustrates benign average, standard deviation, and benign range. Table 11 illustrates high risk average, standard deviation, and high risk range. Table 12 illustrates tobacco consumer average, standard deviation, and tobacco consumer range.

[0416] Table 10. Benign:

Benign Average	Standard deviation	Benign Average *10000	Standard deviation *10000	**Benign Range**	
0.000926	0.000125	9.26	1.25	8.01	10.52
0.001422	0.000113	14.22	1.13	13.10	15.35
0.002300	0.000359	23.00	3.59	19.40	26.59
0.001165	0.000245	11.65	2.45	9.20	14.09
0.001566	0.000269	15.66	2.69	12.97	18.36
0.006087	0.000427	60.87	4.27	56.60	65.14
0.001212	0.000115	12.12	1.15	10.97	13.27
0.000947	0.000080	9.47	0.80	8.68	10.27
0.001129	0.000307	11.29	3.07	8.23	14.36
0.002514	0.000208	25.14	2.08	23.05	27.22
0.000405	0.000067	4.05	0.67	3.38	4.72
0.004789	0.000246	47.89	2.46	45.43	50.35
0.000381	0.000203	3.81	2.03	1.78	5.84
0.000624	0.000188	6.24	1.88	4.35	8.12
0.004068	0.000181	40.68	1.81	38.87	42.48
0.000196	0.000052	1.96	0.52	1.44	2.48

[0417] Table 11. High Risk:

High risk Average	Standard deviation	High risk average *10000	Standard deviation *10000	**High risk Range**	
0.000936593	7.70885E-05	9.37	0.77	8.60	10.14
0.001393087	9.39065E-05	13.93	0.94	12.99	14.87
0.002138595	0.000128966	21.39	1.29	20.10	22.68
0.001104165	0.000167808	11.04	1.68	9.36	12.72
0.001742435	0.000185206	17.42	1.85	15.57	19.28
0.006245197	0.000397744	62.45	3.98	58.47	66.43
0.001188441	0.000103056	11.88	1.03	10.85	12.91
0.000830864	0.000128315	8.31	1.28	7.03	9.59
0.001924027	0.000608156	19.24	6.08	13.16	25.32
0.002567625	0.000159476	25.68	1.59	24.08	27.27
0.000398881	6.15424E-05	3.99	0.62	3.37	4.60
0.004696903	0.000152491	46.97	1.52	45.44	48.49
0.000822115	0.000367027	8.22	3.67	4.55	11.89
0.000439808	0.000107953	4.40	1.08	3.32	5.48
0.00407284	0.000144788	40.73	1.45	39.28	42.18
0.000214038	2.88391E-05	2.14	0.29	1.85	2.43

[0418] Table 12. Tobacco consumers (smoking and chewable tobacco):

Tobacco consumers Average	Standard deviation	Tobacco consumers average *10000	Standard deviation *10000	**Tobacco consumers Range**	
0.000963163	6.76651E-05	9.63	0.68	8.95	10.31
0.001380675	0.00010331	13.81	1.03	12.77	14.84
0.002085035	0.000165477	20.85	1.65	19.20	22.51
0.001078631	0.000171869	10.79	1.72	9.07	12.50
0.001732967	0.000192987	17.33	1.93	15.40	19.26
0.006066214	0.000440098	60.66	4.40	56.26	65.06
0.001126788	0.000127617	11.27	1.28	9.99	12.54
0.000818576	0.000153214	8.19	1.53	6.65	9.72
0.001849922	0.000689161	18.50	6.89	11.61	25.39
0.002518773	0.000180882	25.19	1.81	23.38	27.00
0.000408336	8.34086E-05	4.08	0.83	3.25	4.92
0.004693564	0.000199707	46.94	2.00	44.94	48.93
0.000815103	0.000414275	8.15	4.14	4.01	12.29
0.00044024	0.000120599	4.40	1.21	3.20	5.61
0.004130591	0.000163396	41.31	1.63	39.67	42.94
0.000227376	4.17905E-05	2.27	0.42	1.86	2.69

[0419] Operation 5. Compare the test sample against each of the following generated categories shown in Table 13.

[0420] Table 13.

Wavenumber (cm-1)	Control Range		CI Range		CTNI Range		CTN Range		COT Range		CS Range		Benign Range		High risk Range		Tobacco consumers Range	
822.0056	9.38	11.65	8.69	10.79	7.93	11.67	8.71	16.44	8.8	10.28	8.88	10.82	8.41	10.52	8.6	10.14	8.96	10.31
846.9756	13.48	17.08	12.76	16.73	12.02	17.64	13.2	16.47	13.09	15.13	13.22	16.53	13.1	16.26	12.98	14.87	12.77	14.84
858.0658	19.84	22.35	18.89	23.55	16.31	24.78	20.28	24.29	20.37	24.11	19.4	23.88	19.4	26.53	20.1	22.68	19.2	22.51
880.9673	8.67	13.43	9.18	12.86	9.01	13.57	9.47	14.16	9.53	14.15	10.77	15.2	9.2	14.09	9.38	12.72	9.67	12.5
940.6898	12.33	17.89	13.95	18.84	13.15	17.73	12.23	17.54	13.45	18.15	11.63	17.05	12.97	18.19	15.57	19.26	15.4	19.28
1008.319	43.72	53.35	53.13	61.99	39.22	60.79	51.92	61.42	52.63	61.26	40.83	53.23	55.6	65.14	58.47	69.43	58.28	65.08
1034.96	7.67	11.98	9.59	12.59	8.3	11.51	9.95	13.05	9.68	12.84	7.12	12.62	10.87	13.27	11.65	12.91	9.99	12.54
1067.475	8.48	9.25	9.07	10.15	6.66	8.35	9.01	10.97	8.65	10.68	7.91	10.12	8.68	10.27	7.03	9.59	8.65	9.72
1156.085	11.5	26.07	8.79	14.41	11.85	16.78	7.23	13.85	7.86	13.73	8.79	17.59	8.23	14.36	13.16	28.32	11.61	25.36
1344.079	15.98	25.7	21.33	28.62	14.67	25.51	20.84	28.08	21.42	28.11	16.21	25.85	23.05	27.22	24.08	27.27	23.58	27
1408.182	2.88	6.86	3.07	6.02	3.54	7.91	3.18	5.18	3.28	4.89	2.88	7.84	3.18	4.73	3.37	6.8	3.25	4.92
1453.882	38.53	48.35	44.87	48.58	37.68	46.78	44.68	48.42	44.87	48.23	39.68	48.22	45.43	50.35	46.44	48.46	44.84	48.83
1618.332	5.85	16.84	2.52	7.14	6.76	12.61	1.59	6.5	2.12	6.47	2.08	12.97	1.76	5.84	4.86	11.69	4.61	12.26
1658.911	2.42	5.2	4.17	6.79	1.65	3.34	4.66	7.56	4.61	7.32	3.28	7.42	4.35	8.12	3.32	5.48	3.2	5.61
1669.186	39.68	44.61	36.49	42.9	33.4	44.14	38.61	43.57	38.31	42.09	39.44	42.22	38.97	42.48	36.28	42.16	38.87	42.84
1736.744	2.05	3.36	3.22	3.64	2.14	3.1	1.86	2.83	1.68	2.48	1.87	2.44	1.44	2.48	1.85	2.43	1.86	2.69

[0421] Operation 6. Identify the "in range" and "not in range" features of the test sample when compared with each group.

[0422] Operation 7. Select the category with the highest number of "in range" features for the test samples. This may be the group which matches the most with the test sample.

[0423] Operation 8. Use the above identified group in Operation 7 to classify the test samples based on their 16 features intensity values.

[0424] Operation 9. A similar analysis (e.g., Operations 1-8) may also be performed using 1006 features.

[0425] Example 8: Raman Spectroscopy

[0426] A Raman spectroscopy analysis was described in Example 2 herein. Example 2 was based on a 5 x 6 approach, and used 6 different groups: Control, Control Indeterminant, Cancer Treatment Naïve Indeterminant, Cancer Treatment Naïve, Cancer on Treatment, and Cancer Survivor. In some embodiments, the analysis in Example 2 herein may include the use of 3 additional groups. The 3 additional groups may include Benign, High Risk, and Tobacco consumers, and may be based on an 8 x 9 approach. Moreover, the intensity calculations may be based on a 1006 feature analysis, in addition to the previous analyses herein involving 8 features and 16 features respectively.

[0427] Case Study

[0428] Subject details: A blood sample was obtained from a 51-year-old female participant in a registered study. The participant's complete medical history, including her clinical and family history of cancer, was provided. The participant had been diagnosed with breast cancer.

[0429] Raman Spectroscopy Test: A spectra was generated from a plasma sample of a breast cancer treatment naïve subject (test sample), which had been subjected to pre-processing techniques. The pre-processing techniques included interpolation, spectral smoothing, baseline correction, and normalization. A multivariate data analysis was performed using principal component analysis (PCA) and PC-LDA (linear discriminant analysis) approaches. The trained machine algorithm for an 8 x 9 approach was used, comprising of 9 independent groups. The 9 independent groups included Control, Control Indeterminant, Cancer Treatment Naïve Indeterminant, Cancer Treatment Naïve, Cancer on Treatment, Cancer Survivor, Benign, High risk, and Tobacco consumers. A model was created based on the extracted features of, in this instance, 2 groups, which was the minimum number of groups for this analysis. The test sample was compared against the spectral features of the 2 groups. The test sample was compared against each group and classified in the appropriate group using an 8 x 9 approach. In this case, the sample was classified in the Cancer Treatment Naïve group. Furthermore, the intensity calculation was performed using an 8-feature approach, a 16-feature approach, and a 1006-feature approach, and the sample was classified in cancer.

[0430] Validation of cancer prediction based on spectral features generated from plasma sample: The classification was based on the 8 x 9 approach, and the intensity calculations, appropriately classified the sample in the Cancer Treatment Naïve group. The findings correlated with the patient's current status that indicated that cancer was present.

[0431] Methods

[0432] An internal database was used for generating the trained machine algorithm for the 8 x 9 approach, which comprised of 9 independent groups namely: Control, Control Indeterminant, Cancer Treatment Naïve Indeterminant, Cancer Treatment Naïve, Cancer on Treatment, Cancer Survivor, Benign, High risk, and Tobacco consumers. Each group comprises of characteristic spectral features which are unique to the group. A model was created based on the extracted features of a minimum of 2 groups, and the test sample was compared against these spectral features. The test sample was compared against each group to eventually classify it in the appropriate group using the 8 x 9 approach. Moreover, using an internal database, the intensity values across the 8, 16 and 1006 feature sets were generated for all the 9 independent groups. Further, the intensity value of the test sample was compared against these groups, across the mentioned features, for appropriately classifying the test sample in its respective group.

[0433] Example 9: Inter Instrument Normalization in Raman Spectroscopy

[0434] Spectral variation between instruments may need to be addressed to ensure uniformity and repeatability of data. An in-house normalization method was developed to resolve inter-instrument spectral variation. This normalization method may circumvent the need to generate an

internal database across 9 different cohorts (e.g., Control, Control Indeterminant, Cancer Treatment Naïve Indeterminant, Cancer Treatment Naïve, Cancer on Treatment, Cancer Survivor, Benign, High Risk and Tobacco consumers), on every instrument. The internal database generated on Instrument 1 can be used to assess data generated across different instruments, namely, Instrument 2, 3, 4, etc.

[0435] Case Study

[0436] *Subject details:* A blood sample was obtained from an 80-year-old female subject in a registered study. The subject's complete medical history, including their clinical and family history of cancer, was provided. The subject had been diagnosed with esophageal cancer.

[0437] *Raman Spectroscopy Test:* The blood sample collected was subjected to density gradient centrifugation and plasma was extracted. The spectra were generated from the plasma of a cancer treatment naïve, moderately differentiated squamous cell carcinoma of esophagus (test sample), which had been subjected to pre-processing techniques such as interpolation, spectral smoothing, baseline correction, and normalization.

[0438] The average control spectra generated from Instrument 1 was compared to the average control spectra generated from Instrument 2. As the intensity was higher in Instrument 2, the normalized values of 1,006 features of Instrument 1 were deducted from Instrument 2 values. The difference in values across 1,006 features was used as reference for further analysis. As mentioned herein, the test sample spectra was subjected to pre-processing steps (e.g., interpolation, spectral smoothing, baseline correction, and normalization). The reference value generated across 1,006 features (e.g., the difference between two instruments) was deducted from the normalized values of the test spectra. The spectra was again subjected to baseline correction. Further, multivariate data analysis was performed using principal component analysis (PCA) and PC-LDA approaches. The trained machine algorithm for an 8 x 9 approach was used, comprising of 9 independent groups namely, Control, Control Indeterminant, Cancer Treatment Naïve Indeterminant, Cancer Treatment Naïve, Cancer on Treatment, Cancer Survivor, Benign, High risk, and Tobacco consumers. A model was created based on the extracted features of, in this instance, 2 groups, which was the minimum number of groups possible for this analysis. The test sample was compared against the spectral features of the 2 groups. The test sample was thus compared against each group and classified in the appropriate group using the 8 x 9 approach. In this case, the sample was classified in the Cancer Treatment Naïve group. Furthermore, the intensity calculation was performed using the 8 features approach, the 16 features approach, and the 1,006 features approach, and the sample was classified in cancer.

[0439] *Validation of cancer prediction based on spectral features generated from plasma sample:* The classification based on database generated from Instrument 1 using the 8 x 9

approach, and the intensity calculations, appropriately classified the sample in the Cancer Treatment Naïve group. The findings correlated with the subject's current status that indicated that cancer was present.

[0440] *Methods:* The internal database was used for generating the trained machine algorithm for the 8 x 9 approach, which comprised of 9 independent groups, namely, Control, Control Indeterminant, Cancer Treatment Naïve Indeterminant, Cancer Treatment Naïve, Cancer on Treatment, Cancer Survivor, Benign, High risk, and Tobacco consumers. Each group comprises of characteristic spectral features which are unique to the group. A model was created based on the extracted features of a minimum 2 groups, and the test sample was compared against these spectral features. The test sample post inter-instrument normalization was thus compared against each group to eventually classify it in the appropriate group using the 8 x 9 approach. Moreover, using the internal database, the intensity values across the 8 feature set, 16 feature set, and 1,006 feature set were generated for all the 9 independent groups. Further, the intensity value of the test sample was compared against these groups, across the mentioned features, for appropriately classifying the test sample in respective group.

[0441] While preferred embodiments of the present invention 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. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon a variety of conditions and variables. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is therefore contemplated that the invention shall also cover any such alternatives, modifications, variations, or equivalents. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

CLAIMS

WHAT IS CLAIMED IS:

1. A method for determining a cancer assessment of a subject, comprising:

(a) obtaining a whole blood sample from the subject;

(b) performing centrifugation on the whole blood sample to obtain a plasma layer, a serum layer, a supernatant layer, or a cell pellet;

(c) assaying at least a portion of the plasma layer, the serum layer, the supernatant layer, or the cell pellet, or derivatives thereof, to generate a Raman spectroscopy profile of the whole blood sample of the subject;

(d) processing the Raman spectroscopy profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and

(e) determining the cancer assessment of the subject, based at least in part on the processing in (d).

2. The method of claim 1, wherein the subject is asymptomatic for cancer.

3. The method of claim 1, wherein the subject has a risk factor for cancer.

4. The method of claim 3, wherein the risk factor for cancer comprises clinical history of cancer, family history of cancer, environmental exposure, smoking history, or genetic variation.

5. The method of claim 1, wherein the subject has been diagnosed with cancer.

6. The method of claim 1, wherein the subject is a cancer treatment naïve subject, a cancer survivor subject, a cancer subject receiving treatment, or a benign subject.

7. The method of claim 1, wherein the centrifugation is density gradient centrifugation, and wherein the density gradient centrifugation comprises use of Ficoll hypaque solution.

8. The method of claim 1, wherein (b) comprises obtaining any two of the plasma layer, the serum layer, the supernatant layer, or the cell pellet.

9. The method of claim 1, wherein (b) comprises obtaining any three of the plasma layer, the serum layer, the supernatant layer, or the cell pellet.

10. The method of claim 1, wherein (b) comprises obtaining the plasma layer, the serum layer, the supernatant layer, and the cell pellet.

11. The method of claim 1, wherein (b) further comprises performing the centrifugation at about 1000g, for a time period of between about 5 minutes to about 20 minutes.

12. The method of claim 1, wherein (c) further comprises extracting nucleic acids from at least a portion of the plasma layer, the serum layer, the supernatant layer, and the cell pellet, and assaying the extracted nucleic acids to generate the Raman spectroscopy profile of the whole blood sample of the subject.

13. The method of claim 12, wherein the nucleic acids comprise deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or a combination thereof.

14. The method of claim 12, wherein the nucleic acids are extracted from the plasma layer, the serum layer, the supernatant layer, the cell pellet, or a combination thereof.

15. The method of claim 12, further comprising amplifying the extracted nucleic acids.

16. The method of claim 15, wherein the amplifying comprises polymerase chain reaction (PCR).

17. The method of claim 12, further comprising using primers or probes to selectively enrich the nucleic acids for a set of cancer-specific biomarkers.

18. The method of claim 17, wherein the primers or probes are nucleic acid primers or nucleic acid probes.

19. The method of claim 18, wherein the nucleic acid probes have sequence complementarity with nucleic acid sequences of the set of cancer-specific biomarkers.

20. The method of claim 12, further comprising assaying at least a second portion of the extracted nucleic acids by deoxyribonucleic acid (DNA) sequencing, ribonucleic acid (RNA) sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, a methylation array, digital droplet polymerase chain reaction (PCR), digital polymerase chain reaction (PCR), assay for transposase-accessible chromatin (ATAC) sequencing, or methylation-specific polymerase chain reaction (PCR).

21. The method of claim 1, wherein (c) further comprises assaying at least a portion of the plasma layer, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

22. The method of claim 21, wherein (c) further comprises loading the plasma layer, or derivatives thereof, onto a substrate.

23. The method of claim 22, wherein the substrate comprises calcium fluoride.

24. The method of claim 1, wherein (c) further comprises assaying at least a portion of the serum layer, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

25. The method of claim 1, wherein (c) further comprises assaying at least a portion of the cell pellet, or derivatives thereof, to generate the Raman spectroscopy profile of the whole blood sample of the subject.

26. The method of claim 1, wherein (c) further comprises performing time-gated Raman spectroscopy (TGRS), WITec Raman spectroscopy, or Renishaw Raman spectroscopy.

27. The method of claim 1, wherein the Raman spectroscopy profile comprises a spectral range for analysis between 600 cm^{-1} and 1,800 cm^{-1}.

28. The method of claim 1, wherein (c) is performed under light illumination.

29. The method of claim 1, further comprising performing a pre-processing technique on the Raman spectroscopy profile of the whole blood sample of the subject.

30. The method of claim 29, wherein the pre-processing technique comprises spectral interpolation, spectral smoothing, baseline correction, or normalization, or a combination thereof.

31. The method of claim 1, further comprising performing a univariate or multivariate analysis on the Raman spectroscopy profile of the whole blood sample of the subject.

32. The method of claim 1, further comprising performing a dimensionality reduction on the Raman spectroscopy profile of the whole blood sample of the subject.

33. The method of claim 32, wherein the dimensionality reduction comprises principal component analysis (PCA) or linear discriminant analysis (LDA).

34. The method of claim 33, wherein the LDA comprises Principal Component Based Linear Discriminant Analysis (PC-LDA).

35. The method of claim 1, wherein (d) further comprises extracting one or more sets of features, singularly or in combination, from the Raman spectroscopy profile of the whole blood sample of the subject.

36. The method of claim 35, wherein the one or more sets of features comprises an increase or a decrease in a spectral feature selected from the group consisting of:

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1087- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp
1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1160- Beta-carotene

1518-Beta carotene
1052- C-N Stretch, Phospholipid
1067- C-N Stretch, Phospholipid, C-C Stretch
1107- C-N Stretch, C-C stretch
1130- Protein and Lipid: C-N stretch and chain C-C stretch
1176- Tyr/Phe, CH2 wag, CH2 twist
1210- Tyr/Phe

37. The method of claim 36, wherein the one or more sets of features comprises 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 11, at least 12, at least 13, at least 14, at least 15, or at least 16 different increases or decreases in the spectral features selected from the group consisting of:

622.1-Nucleotide conformation/Phenylalanine
644-tyrosine doublet/C-S twisting
757- Guanine/Cytochrome/Protein/Tryptophan, d-ring
881-Disaccharide (cellobiose), (C-O-C) skeletal mode (protein)
942.2-Phenylalanine C-C stretch/ (Amide III helix)
1006-phenylalanine
1034-Protein: Phenylalanine
1087- DNA/RNA symmetric stretching because of parallel dichroism
1343-C-H bend/trp
1407 - ns COO2 (IgG)
1452-C=O stretch and CH2 bending of phospholipids/protein/CH2 deformation amide III
1558-tyr doublet/trp
1659- C=O vibration of proteins (amide I), phospholipid (-C=C-) stretch
1735-Phospholipid: ester carbonyl
1160- Beta-carotene
1518-Beta carotene
1052- C-N Stretch, Phospholipid
1067- C-N Stretch, Phospholipid, C-C Stretch
1107- C-N Stretch, C-C stretch
1130- Protein and Lipid: C-N stretch and chain C-C stretch

1176- Tyr/Phe, CH2 wag, CH2 twist
1210- Tyr/Phe

38. The method of claim 35, wherein the set of features comprises intensity values of spectral peaks.

39. The method of claim 35, wherein the set of features comprises an average intensity value across a plurality of spectral peaks.

40. The method of claim 39, wherein the set of features comprises a change of the average intensity value as compared to a reference average intensity value.

41. The method of claim 35, wherein the set of features comprises features that are considered in combination with each other rather than as singular peaks.

42. The method of claim 1, further comprising determining a presence of cancer in the subject or a tissue or location of origin of the cancer.

43. The method of claim 1, wherein (e) further comprises determining a presence or an absence of minimal residual disease or benign lesion in the subject.

44. The method of claim 1, further comprising administering a treatment to the subject based on a detected presence of cancer in the subject.

45. The method of claim 44, wherein the treatment is selected from the group consisting of surgery, chemotherapy, targeted therapy, radiotherapy, or immunotherapy.

46. The method of claim 1, wherein the trained machine learning algorithm comprises a deep learning algorithm, a linear or logistic regression, a support vector machine (SVM), a neural network, a Random Forest, or a gradient boosted algorithm.

47. The method of claim 1, wherein the trained machine learning algorithm is trained using:

a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer, wherein the subjects are cancer treatment naïve, and

a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer.

48. The method of claim 47, wherein the trained machine learning algorithm is trained using at least one of:

a third set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer;

a fourth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer;

a fifth set of independent training samples obtained or derived from subjects that are cancer survivors;

a sixth set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment;

a seventh set of independent training samples obtained or derived from subjects that benign subjects;

an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions; and

a ninth set of independent training samples obtained or derived from subjects known to consume tobacco.

49. The method of claim 1, wherein the reference is obtained or derived from at least one of:

a set of subjects with a presence or an elevated susceptibility of cancer and are cancer treatment naïve;

a set of subjects with an absence or a non-elevated susceptibility of cancer;

a set of subjects that are cancer treatment naïve;

a set of subjects that are indeterminate for a presence or an elevated susceptibility of cancer;

a set of subjects that are indeterminate for an absence or a non-elevated susceptibility of cancer;

a set of subjects that are cancer survivors;

a set of subjects that are cancer subjects receiving treatment;

a set of subjects that are benign subjects;

a set of subjects that have one or more high risk conditions; and

a set of subjects that are known to consume tobacco.

50. The method of claim 49, wherein the one or more high risk conditions comprise hypertension, diabetes, cardiovascular disease, obesity, subjects that are overweight, or any combination thereof.

51. The method of claim 1, wherein the cancer is selected from the group consisting of: bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and a cancer with primary unknown origin.

52. The method of claim 1, further comprising using the trained machine learning algorithm to detect a presence or an absence of each of a plurality of different cancer types.

53. The method of claim 1, wherein the cancer assessment comprises determining a presence of a cancer, an absence of a cancer, or a likelihood or risk of a cancer.

54. The method of claim 1, wherein the reference is generated based on one or more of: non-cancer subjects, cancer treatment naïve subjects, cancer survivor subjects, cancer subjects receiving treatment, and benign subjects.

55. A system comprising one or more processors and a memory operatively coupled to the one or more processors, wherein the one or more processors are individually or collectively programmed to perform the method of claim 1.

56. A method for determining a cancer assessment of a subject, comprising:

(a) obtaining a whole blood sample from the subject;

(b) performing centrifugation on the whole blood sample to obtain a cell pellet;

(c) extracting nucleic acids from the cell pellet;

(d) assaying the extracted nucleic acids to generate a genetic profile of the whole blood sample of the subject, a transcriptomic profile of the whole blood sample of the subject, an epigenetic profile of the whole blood sample of the subject, or a combination thereof;

(e) processing the genetic profile of the whole blood sample of the subject, the transcriptomic profile of the whole blood sample of the subject, and the epigenetic profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and

(f) determining the cancer assessment of the subject, based at least in part on the processing in (e).

57. The method of claim 56, wherein the subject is asymptomatic for cancer.

58. The method of claim 56, wherein the subject has a risk factor for cancer.

59. The method of claim 58, wherein the risk factor for cancer comprises a clinical history of cancer, family history of cancer, environmental exposure, smoking history, genetic alteration or genetic variation.

60. The method of claim 56, wherein the subject has been diagnosed with cancer.

61. The method of claim 56, wherein the subject is a cancer treatment naïve subject, a cancer survivor subject, a cancer subject receiving treatment, or a benign subject.

62. The method of claim 56, wherein the centrifugation is density gradient centrifugation, and wherein the density gradient centrifugation comprises use of Ficoll hypaque solution.

63. The method of claim 56, wherein the centrifugation is density gradient centrifugation, and wherein (b) further comprises performing the density gradient centrifugation at about 1000g, for a time period of between about 5 minutes to about 20 minutes.

64. The method of claim 56, wherein the nucleic acids comprise deoxyribonucleic acid (DNA), ribonucleic acid (RNA), or a combination thereof.

65. The method of claim 56, further comprising amplifying the extracted nucleic acids.

66. The method of claim 65, wherein the amplifying comprises polymerase chain reaction (PCR).

67. The method of claim 56, further comprising using primers or probes to selectively enrich the nucleic acids for a set of cancer-specific biomarkers.

68. The method of claim 67, wherein the primers or probes are nucleic acid primers or nucleic acid probes.

69. The method of claim 68, wherein the nucleic acid probes have sequence complementarity with nucleic acid sequences of the set of cancer-specific biomarkers.

70. The method of claim 56, further comprising assaying at least a second portion of the extracted nucleic acids by deoxyribonucleic acid (DNA) sequencing, ribonucleic acid (RNA) sequencing, bisulfite sequencing, targeted methylation sequencing, pyrosequencing, enzymatic treatment, a methylation array, digital droplet polymerase chain reaction (PCR), digital polymerase chain reaction (PCR), assay for transposase-accessible chromatin (ATAC) sequencing, quantitative polymerase chain reaction (PCR), or methylation-specific polymerase chain reaction (PCR).

71. The method of claim 56, wherein (c) further comprising extracting the nucleic acids from very small embryonic-like stem cells (VSELs) of the cell pellet.

72. The method of claim 56, further comprising determining a presence of cancer in the subject or a tissue or location of origin of the cancer.

73. The method of claim 56, wherein (f) further comprises determining a presence or an absence of minimal residual disease or benign lesion in the subject.

74. The method of claim 56, further comprising administering a treatment to the subject based on a detected presence of cancer in the subject.

75. The method of claim 74, wherein the treatment is subject to a detected modulation of cancer-specific genes or pathways.

76. The method of claim 74, wherein the treatment is selected from the group consisting of surgery, chemotherapy, targeted therapy, radiotherapy, or immunotherapy.

77. The method of claim 56, wherein the trained machine learning algorithm comprises a deep learning algorithm, a linear or logistic regression, a support vector machine (SVM), a neural network, a Random Forest, or a gradient boosted algorithm.

78. The method of claim 56, wherein the trained machine learning algorithm is trained using:

a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer; and

a second set of independent training samples obtained or derived from subjects with an absence or non-elevated susceptibility of cancer.

79. The method of claim 78, wherein the trained machine learning algorithm is trained using:

a first set of independent training samples obtained or derived from subjects with a presence or elevated susceptibility of cancer and that are cancer treatment naïve; and

a second set of independent training samples obtained or derived from subjects with an absence of cancer and with a non-elevated susceptibility of cancer.

80. The method of claim 79, wherein the trained machine learning algorithm is trained using at least one of:

a third set of independent training samples obtained or derived from subjects that are indeterminate for presence or elevated susceptibility of cancer;

a fourth set of independent training samples obtained or derived from subjects that are indeterminate for absence or non-elevated susceptibility of cancer;

a fifth set of independent training samples obtained or derived from subjects that are cancer survivors;

a sixth set of independent training samples obtained or derived from subjects that are cancer subjects receiving treatment;

a seventh set of independent training samples obtained or derived from subjects that benign subjects;

an eighth set of independent training samples obtained or derived from subjects who have one or more high risk conditions; and

a ninth set of independent training samples obtained or derived from subjects known to consume tobacco.

81. The method of claim 56, wherein the reference is obtained or derived from at least one of:

a set of subjects with a presence or elevated susceptibility of cancer;

a set of subjects with an absence or non-elevated susceptibility of cancer;

a set of subjects that are cancer treatment naïve;

a set of subjects that are indeterminate for presence or elevated susceptibility of cancer;

a set of subjects that are indeterminate for absence or non-elevated susceptibility of cancer;

a set of subjects that are cancer survivors;

a set of subjects that are cancer subjects receiving treatment;

a set of subjects that are benign subjects;

a set of subjects that have one or more high risk conditions; and

a set of subjects known to consume tobacco.

82. The method of claim 81, wherein the one or more high risk conditions comprise hypertension, diabetes, cardiovascular disease, obesity, or subjects that are overweight, or any combination thereof.

83. The method of claim 56, wherein the cancer is selected from the group consisting of: bladder cancer, breast cancer, colorectal cancer, endometrial cancer, kidney cancer, leukemia, liver cancer, lung cancer, lymphoma, melanoma, pancreatic cancer, prostate cancer, thyroid cancer, adnexae cancer, appendix cancer, bone cancer, caecum cancer, duodenum cancer, rectal cancer, anal cancer, gall bladder cancer, gastrointestinal junction cancer, esophageal cancer, larynx cancer, oral cancer, buccal mucosa cancer, oropharynx cancer, nasopharynx cancer, salivary gland cancer, tongue cancer, tonsils cancer, ovarian cancer, cervix, penile cancer, primary peritoneum cancer, prostate cancer, melanoma, soft tissue cancer, stomach cancer, testicular cancer, throat cancer, uterine cancer, vaginal stump cancer, Hodgkin's lymphoma, Non-Hodgkin's lymphoma, multiple myeloma, and a cancer with primary unknown origin.

84. The method of claim 56, further comprising using the trained machine learning algorithm to detect a presence or an absence of each of a plurality of different cancer types.

85. The method of claim 56, wherein the cancer assessment comprises determining a presence of cancer, an absence of a cancer, or a likelihood or risk of a cancer.

86. The method of claim 56, wherein the reference is generated based on one or more of: non-cancer subjects, cancer treatment naïve subjects, cancer survivor subjects, cancer subjects receiving treatment, or benign subjects.

87. A system comprising one or more processors and a memory operatively coupled to the one or more processors, wherein the one or more processors are individually or collectively programmed to perform the method of claim 56.

88. A method for evaluating or monitoring a therapy response of a subject with cancer, comprising:

(a) obtaining a whole blood sample from the subject, subsequent to the subject being administered a therapy;

(b) performing centrifugation on the whole blood sample of the subject to obtain a plasma layer, a serum layer, or a cell pellet;

(c) assaying at least a portion of the plasma layer, the serum layer, or the cell pellet, or derivatives thereof, to generate a Raman spectroscopy profile of the whole blood sample of the subject;

(d) processing the Raman spectroscopy profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and

(e) evaluating the therapy response of the subject, based at least in part on the processing in (d).

89. The method of claim 88, wherein (b) comprises obtaining any two of the plasma layer, the serum layer, or the cell pellet.

90. The method of claim 88, wherein (b) comprises obtaining the plasma layer, the serum layer, and the cell pellet.

91. A method for determining a cancer assessment of a subject, comprising:

(a) obtaining a whole blood sample from the subject;

(b) performing centrifugation on the whole blood sample of the subject to obtain a supernatant or a cell pellet;

(c) assaying at least a portion of the supernatant or the cell pellet, or derivatives thereof, to generate a Raman spectroscopy profile of the whole blood sample of the subject;

(d) processing the Raman spectroscopy profile of the whole blood sample of the subject using a trained machine learning algorithm or against a reference; and

(e) determining the cancer assessment of the subject, based at least in part on the processing in (d).

92. The method of claim 91, wherein (b) comprises obtaining the supernatant and the cell pellet.

93. The method of claim 91, wherein, in (c), the supernatant, or derivatives thereof, is assayed to generate the Raman spectroscopy profile of the whole blood sample of the subject.

94. The method of claim 91, wherein the subject is asymptomatic for cancer.

95. The method of claim 91, wherein the subject has a risk factor for cancer.

96. The method of claim 95, wherein the risk factor for cancer comprises clinical history of cancer, family history of cancer, environmental exposure, smoking history, or genetic variation.

97. The method of claim 91, wherein the subject has been diagnosed with cancer.

98. The method of claim 91, wherein the subject is a cancer treatment naïve subject, a cancer survivor subject, a cancer subject receiving treatment, or a benign subject.

99. The method of claim 91, wherein the supernatant is incubated one or more times prior to the assaying of (c).

100. The method of claim 91, wherein (b) further comprises performing the centrifugation at about 1000g, for a time period of between about 5 minutes to about 20 minutes.