

WHAT IS CLAIMED IS:

1. A method of using a classifier capable of distinguishing a population of individuals, the method comprising:

a) assaying a plurality of classes of molecules in a biological sample using a plurality of assays, wherein the assaying provides a plurality of sets of measured values representative of the plurality of classes of molecules, wherein the plurality of classes of molecules includes a first class of nucleic acids and a second class of polyamino acids, wherein the first class of nucleic acids is cell-free DNA, and wherein assaying the cell-free DNA includes a methylation assay;

b) identifying a set of features corresponding to properties of each of the plurality of classes of molecules to be input to a machine learning model;

c) preparing a feature vector of feature values from the plurality of sets of measured values, each feature value corresponding to a feature of the set of features and including one or more measured values, wherein the feature vector includes at least one feature value obtained using each set of the plurality of sets of measured values;

d) loading, into a memory of a computer system, the machine learning model comprising the classifier, the machine learning model trained using training vectors obtained from training biological samples, a first subset of the training biological samples identified as having a specified property and a second subset of the training biological samples identified as not having the specified property; and

e) inputting the feature vector into the machine learning model to obtain an output classification of whether the biological sample has the specified property, thereby distinguishing the population of individuals having the specified property.

2. The method of claim 1, wherein the plurality of classes of molecules are selected from a group consisting of nucleic acid, polyamino acids, carbohydrates, or metabolites.

3. The method of claim 1, wherein one or more additional classes of the plurality of classes of molecules are selected from a group consisting of cellular deoxyribonucleic acid (DNA), ~~genomic DNA~~, plasmid DNA, complementary DNA (cDNA), ~~cell-free (e.g., non-encapsulated) DNA (cfDNA), circulating tumor DNA (ctDNA), nucleosomal DNA, chromatosomal DNA,~~ mitochondrial DNA (mtDNA), an artificial nucleic acid analog,

recombinant nucleic acid, plasmids, viral vectors, chromatin and peripheral blood mononuclear cell-derived (PBMC-derived) genomic DNA.

4. The method of claim 1, wherein one or more additional classes of the plurality of classes of molecules are selected from a group consisting of nucleic acids comprising ribonucleic acid (RNA), messenger RNA (mRNA), transfer RNA (tRNA), micro RNA (mitoRNA), ribosomal RNA (rRNA), circulating RNA (cRNA), alternatively spliced mRNAs, small nuclear RNAs (snRNAs), antisense RNA, short hairpin RNA (shRNA), or small interfering RNA (siRNA).

5. The method of claim 1, wherein the ~~plurality of classes~~ second class of polyamino acids ~~molecules is~~are selected from a group consisting of ~~polyamino acid~~, peptide, protein, autoantibody, or a fragment thereof.

6. The method of claim 1, wherein one or more additional classes of the plurality of classes of molecules are selected from a group consisting of sugars, lipids, ~~amino acids~~, fatty acids, phenolic compounds, or alkaloids.

7. The method of claim 1, wherein the second class and one or more additional classes of the plurality of classes of molecules are selected from a group consisting of at least two of: ~~cfDNA molecules~~, cfRNA molecules, circulating proteins, antibodies, and metabolites.

8. The method of claim 1, wherein one or more additional classes of the plurality of classes of molecules are selected from a group consisting of: ~~1) cfDNA, cfRNA, polyamino acid, and small chemical molecules, or 2) cfDNA and cfRNA, and polyamino acids, 3) cfDNA and cfRNA and small chemical molecules, or 4) cfDNA, polyamino acid, and small chemical molecules, or 5) cfRNA, polyamino acid, and small chemical molecules, or 6) cfDNA and cfRNA, or 7) cfDNA and polyamino acid, or 8) cfDNA and small chemical molecules, or 9) cfRNA and polyamino acid, or 10) cfRNA and small chemical molecules, or 11) polyamino acid and small chemical molecules.~~

9. The method of claim 1, wherein the plurality of classes of molecules is cfDNA, proteins, and autoantibodies.

10. The method of claim 1, wherein the plurality of assays ~~can include~~ includes at least two of: whole-genome sequencing (WGS), whole-genome bisulfite sequencing (WGSB), EM-seq sequencing, small-RNA sequencing, quantitative immunoassay, enzyme-linked immunosorbent assay (ELISA), proximity extension assay (PEA), protein microarray, mass spectrometry, low-coverage Whole-Genome Sequencing (lcWGS); selective tagging 5mC sequencing (~~WO2019/051484~~), CNV calling; tumor fraction (TF) estimation; Whole Genome Bisulfite Sequencing; LINE-1 CpG methylation; 56 genes CpG methylation; cf-Protein Immuno-Quant ELISAs, SIMOA₃; and cf-miRNA sequencing, ~~and cell type or cell phenotype mixture proportions derived from any of the above assays.~~

11. The method of claim 10, wherein the whole-genome bisulfite or EM-seq sequencing includes a methylation analysis.

12. The method of claim 1, wherein the classifier is trained and constructed according to one or more of: linear discriminant analysis (LDA); partial least squares (PLS); random forest; k-nearest neighbor (KNN); support vector machine (SVM) with radial basis function kernel (SVMRadial); SVM with linear basis function kernel (SVMLinear); SVM with polynomial basis function kernel (SVMPoly), decision trees, multilayer perceptron, mixture of experts, sparse factor analysis, hierarchical decomposition, and combinations of linear algebra routines and statistics.

13. The method of claim 1, wherein the specified property is a presence of a clinically-diagnosed disorder.

14. The method of claim 1, wherein the specified property is a cancer selected from the group consisting of colorectal cancer, liver cancer, lung cancer, pancreatic cancer, ~~and breast cancer~~, adenomatous polyps, sessile serrated adenoma (SSA), advanced adenoma, colorectal dysplasia, colorectal adenoma, colon cancer, rectal cancer, colorectal carcinoma, colorectal adenocarcinoma, carcinoid tumors, gastrointestinal carcinoid tumors, gastrointestinal stromal tumors (GISTs), lymphomas, and sarcomas.

15. The method of claim 1, wherein the specified property is a responsiveness to a treatment.

16. A system for training a classifier for performing classifications of plasma biological samples, the system comprising:

a) a receiver module to receive a plurality of training samples, each of the plurality of training samples having a plurality of classes of molecules, wherein each of the plurality of training samples comprises one or more known labels, wherein the plurality of classes of molecules includes a first class of nucleic acids and a second class of polyamino acids, wherein the first class of nucleic acids is cell-free DNA, and wherein assaying the cell-free DNA includes a methylation assay;

b) a feature selection module to identify a set of features corresponding to each of a plurality of different assays that are operable to be input to a machine learning model for each of the plurality of training samples, wherein the set of features corresponds to properties of molecules in the plurality of training samples,

wherein, for each of the plurality of training samples, the system is operable to subject the plurality of classes of molecules in the training sample to the plurality of different assays to obtain sets of measured values, wherein each set of measured values is from one assay applied to a class of molecules in the training sample, wherein a plurality of sets of measured values are obtained for the plurality of training samples, the set of measured values for the first class includes methylation patterns of the cell-free DNA found in the plasma samples;

c) a feature extraction module to analyze, for each of the plurality of training samples, the sets of measured values to obtain a training vector for the training sample, wherein the training vector comprises feature values of a set of features of the corresponding assay, each feature value corresponding to a feature and including one or more measured values, wherein the training vector is formed using at least one feature from at least two of the sets of features corresponding to a first subset of the plurality of different assays;

d) a machine learning module configured to operate on the training vectors using parameters of the machine learning model to obtain output labels for the plurality of training samples;

e) a comparator module to compare the output labels to the known labels of the training samples;

f) a training module to iteratively search for optimal values of the parameters as part of training the machine learning model based on the comparing the output labels to the known labels of the training samples; and

g) an output module to provide the parameters of the machine learning model and the set of features for the machine learning model.

17. The system of claim 16, wherein the machine learning module comprises a classification circuit that is configured as a machine learning classifier selected from a linear discriminant analysis (LDA) classifier, a quadratic discriminant analysis (QDA) classifier, a support vector machine (SVM) classifier, a random forest (RF) classifier, a linear kernel support vector machine classifier, a first or second order polynomial kernel support vector machine classifier, a ridge regression classifier, an elastic net algorithm classifier, a sequential minimal optimization algorithm classifier, a naive Bayes algorithm classifier, and a NMF predictor algorithm classifier.

18. The system of claim 16, wherein the system further comprises means for performing the method of any one of claims 1-15 and 22-32~~the preceding methods~~.

19. A system for classifying subjects based on a multi-analyte analysis in a biological sample composition comprising: (a) a computer-readable medium comprising a classifier operable to classify the subjects based on the multi-analyte analysis according to the method of any one of claims 1-15 and 22-32; and (b) one or more processors for executing instructions stored on the computer-readable medium.

20. A non-transitory computer-readable medium comprising machine-executable code that, upon execution by one or more computer processors, implements the method of any one of claims 1-15 and 22-32~~any of the methods above or elsewhere herein~~.

21. A system comprising one or more computer processors and computer memory coupled thereto, wherein the computer memory comprises machine-executable code that, upon execution by the one or more computer processors, implements the method of any one of claims 1-15 and 22-32~~any of the methods above or elsewhere herein~~.

22. A method of detecting a presence of a cancer in an individual comprising:

a) assaying a plurality of classes of molecules in a biological sample obtained from the individual wherein the assaying provides a plurality of sets of measured values representative of the plurality of classes of molecules, wherein the plurality of classes of molecules include cell-free nucleic acids and polyamino acids, wherein a first assay is applied to the cell-free nucleic acids to obtain a first set of measured values, and wherein a second assay is applied to the polyamino acids to obtain a second set of measured values, the plurality of sets of measured values including the first set of measured values from the first assay and the second set of measured values from the second assay;

b) identifying a set of features corresponding to properties of each of the plurality of classes of molecules to be input to a machine learning model;

c) preparing a feature vector of feature values from each of the plurality of sets of measured values, each feature value corresponding to a feature of the set of features and including one or more measured values, wherein the feature vector includes at least one feature value obtained using each set of the plurality of sets of measured values; and

d) loading, into a memory of a computer system, the machine learning model that is trained using training vectors obtained from training biological samples, a first subset of the training biological samples identified from individuals with cancer and a second subset of the training biological samples identified from individuals not having cancer; and

e) inputting the feature vector into the machine learning model to obtain an output classification of whether the biological sample is associated with the cancer, thereby detecting the presence of the cancer in the individual.

23. The method of claim 22 wherein the output classification includes a detection value that indicates ~~the~~ a presence of the cancer in the individual.

24. The method of claim 22, wherein the machine learning model further outputs another classification that provides a probability of the biological sample not having cancer.

25. The method of claim 22, wherein the cancer is colorectal cancer, liver cancer, lung cancer, pancreatic cancer or breast cancer.

26. A method of determining a prognosis of an individual with cancer comprising:

a) assaying a plurality of classes of molecules in a biological sample wherein the assaying provides a plurality of sets of measured values representative of the plurality of classes of molecules, wherein the plurality of classes of molecules includes a first class of nucleic acids and a second class of polyamino acids, wherein the first class of nucleic acids is cell-free DNA, and wherein assaying the cell-free DNA includes a methylation assay;

b) identifying a set of features corresponding to properties of the plurality of classes of molecules to be input to a machine learning model;

c) preparing a feature vector of feature values from each of the plurality of sets of measured values, each feature value corresponding to a feature of the set of features and including one or more measured values, wherein the feature vector includes at least one feature value obtained using each set of the plurality of sets of measured values;

d) loading, into memory of a computer system, the machine learning model that is trained using training vectors obtained from training biological samples, a first subset of the training biological samples identified from individuals with good cancer prognosis and a second subset of the training biological samples identified from individuals not having a good cancer prognosis; and

e) inputting the feature vector into the machine learning model to obtain an output classification of whether the biological sample is associated with the good cancer prognosis, thereby determining the prognosis of the individual with cancer.

27. The method of claim 26, wherein the cancer can be selected from colorectal cancer, liver cancer, lung cancer, pancreatic cancer or breast cancer.

28. A method of determining responsiveness of an individual to a cancer treatment comprising:

a) assaying a plurality of classes of molecules in a biological sample wherein the assaying provides a plurality of sets of measured values representative of the plurality of classes of molecules, wherein the plurality of classes of molecules include cell-free nucleic acids and polyamino acids, wherein a first assay is applied to the cell-free nucleic acids to obtain a first set of measured values, and wherein a second assay is applied to the polyamino acids to obtain a second set of measured values, the plurality of sets of

measured values including the first set of measured values from the first assay and the second set of measured values from the second assay;

b) identifying a set of features corresponding to properties of each of the plurality of classes of molecules to be input to a machine learning model;₅

c) preparing a feature vector of feature values from each of the plurality of sets of measured values, each feature value corresponding to a feature of the set of features₁ and each feature value including one or more measured values from the plurality of sets of measured values, wherein the feature vector includes at least one feature value obtained using each set of the plurality of sets of measured values;₅

d) loading, into memory of a computer system, the machine learning model that is trained using training vectors obtained from training biological samples, a first subset of the training biological samples identified from individuals responding to the cancer treatment and a second subset of the training biological samples identified from individuals not responding to the cancer treatment;₅ and

e) inputting the feature vector into the machine learning model to obtain an output classification of whether the biological sample is associated with treatment response thereby determining the responsiveness to the cancer treatment.

29. The method of claim 28, wherein the cancer treatment is selected from alkylating agents, plant alkaloids, antitumor antibiotics, antimetabolites, topoisomerase inhibitors, retinoids, checkpoint inhibitor therapy, or VEGF inhibitors.

30. The method of claim 28, wherein the output classification includes a detection value that indicates ~~the a~~ presence of cancer in the individual.

31. The method of any one of claims 1-15 and 22-30, wherein the measured values include methylation patterns of the cell-free DNA found in the biological sample.

32. The method of any one of claims 1-15 and 22-31, wherein the biological sample is a plasma sample.